Journal of Experimental & Theoretical Artificial Intelligence, 2013
Vol. 25, No. 3, 303–322, http://dx.doi.org/10.1080/0952813X.2013.782987

The observer's observer's paradox

Rick Dale* and David W. Vinson

Cognitive and Information Sciences, University of California, Merced, CA, USA

(*Received 10 November 2012; final version received 21 January 2013*)

Perspectives are central to paradox, from perceptual illusions (Bruno, N. (2001). When does action resist visual illusions? *Trends in Cognitive Sciences*, *5*(9), 379–382), to dialetheism (Dietrich, E. (2008). The Bishop and Priest: Toward a point-of-view based epistemology of true contradictions. *Logos Architekton. Journal of Logic and Philosophy of Science*, *2*, 35–58.; Dietrich & Webber, this issue). But why? We argue that apparent inconsistencies among perspectives are driven by the observer's methods, goals, etc., of inquiry: The manner in which the observer is observing. It is well known in various areas of science that observers inherently disrupt the system that is being studied. But even after we have measurement schemes (or theoretical apparatus) for collecting maximally observer-untainted behavioural data, there remains an inextricable observer-centred influence. We consider the case of cognitive science, in which diverse theoretical frameworks on offer continue to be pitted against one another as if one were going to be churned out as true, and the others falsified. A simulation is developed in which the observer's observer's paradox is demonstrated in a highly idealised computational model of a scientist studying a system. We use a model-building method referred to as 'ϵ-machine reconstruction' (Crutchfield, 1994) and provide readers a basic introduction to it. By using this framework to model the observer–measurement process, we explore how different measurement schemes impact resultant theories. The end product is a kind of existence proof, which we liken to more complicated circumstances of cognitive science. We argue that progress in cognitive science – overcoming apparently paradoxical persistence of several theoretical frameworks – may come from rendering the observer as an active aspect of these cognitive theories. The upshot would be one of two possibilities. Either theories are inconsistent because they 'access' divergent levels of organisation in the human cognitive system, or they may be integrated in an as-yet unknown theoretical framework.

Keywords: cognitive science; complexity; dynamical systems; pluralism; ϵ-machine; logistic map

Introduction

Cognitive science faces a paradox in the following sense. Over the past several decades, cognitive scientists have explored a diverse array of theoretical perspectives: symbolic (Newell, 1990), connectionist (Elman et al., 1997; Rumelhart & McCelland, 1986), dynamical (Kelso, 1995; Port & van Gelder, 1995; van Gelder, 1998), Bayesian (Tenenbaum & Griffiths, 2002), etc. These perspectives offer stances from which we can understand many different tasks and behavioural tendencies. This situation is paradoxical in a straightforward sense. A cognitive scientist may recognise the success of several theoretical frameworks, with literatures that

*Corresponding author. Email: rdale@ucmerced.edu

continue to grow; yet these frameworks can seem to offer *contradictory* conceptions of the cognitive system. How can two or more perspectives – often touted as radically different in their conception of the cognitive system – all elucidate the system's operation?

The common solution to this paradox has been to imagine that the situation will resolve itself eventually. This solution, with a long history, considers some theoretical perspectives fundamentally distinct, mutually incompatible and therefore susceptible to the momentum of empirical mitigation that one expects in the sciences (Leahey, 2001; Newell, 1973). Put simply, many are awaiting a paradigm shift, out of which one perspective will be victorious (or perhaps some synergy between theoretical cousins: McClelland et al., 2010; Thelen & Bates, 2003). Another much less explored possibility is that the paradox is unavoidable: It is at least *possible* that two or more apparently incompatible frameworks will withstand this continued empirical mitigation (Dale, 2008). Reasons offered for this unavoidable paradox include that there are diverse goals and methods of inquiry in the cognitive sciences (e.g. Abrahamsen & Bechtel, 2006; Marr, 1982; McCauley & Bechtel, 2001), that the cognitive system, itself, admits of many levels of complex organisation (e.g. Ręczaszek-Leonardi & Scott Kelso, 2008) and that the apparent incompatibility is only superficial, such that we can peel back layers of shared structure among explanations that help us understand their interrelationships (e.g. Dove, 2009; Smolensky, 2012; Smolensky, Goldrick, & Mathis, in press; Smolensky & Legendre, 2005; Weiskopf, 2009). Similar lines of thought have been offered more broadly for theoretical diversity in the sciences (e.g. Cartwright, 1999; Dupré, 1993; Kellert, Longino, & Waters, 2006; Mitchell, 2003).

This possibility, of inevitable theoretical diversity, motivates the current paper. We will argue that this pluralistic outcome derives from the inherent *perspectival* nature of observation itself (Giere, 2006; van Fraassen, 2008). In fact, the paradox goes beyond theory in cognitive science, and into any circumstance in which an observer, with specific goals, intervenes and investigates a complex system. The observer faces the inevitable paradox of observational diversity, because any sufficiently complex observer has an array of methods by which to observe the system he/she wants to study. We dub this the *observer's observer's paradox*, as it is one order above the well-known observer's paradox.

Our goal in the current paper is to add to ongoing discussion in two ways. First, we frame the paradox of incompatible theoretical diversity in terms of the observer's observer's paradox; and second, we offer a 'proof of concept', by developing a simulation of this problem in a computational model. We introduce the ϵ-machine reconstruction process (Crutchfield, 1994; Shalizi & Crutchfield, 2001) and show how it can be used to model the observer–measurement process (as in Crutchfield, 1994). This framework is well suited to simulating the observer's observer's paradox for two reasons. The first is that it is inherently designed to extract computational regularities from an observed process, and these regularities, in the form of a graph structure, can be loosely conceptualised as a 'theory' that an observer builds. The second is that the reconstruction process can be readily modified, to explore different outcomes when the observer's measurement approach is changed. Our version of this model is very simple, utilising the framework 'out of the box', so to speak – but it is instructive in demonstrating that even in the simplest circumstances, an observer faces disparate *theoretical* outcomes depending on how he/she studies a complex system. We conclude that if the observer's observer's paradox holds in the simplest cases, it is likely to hold among the most complex scientific cases, such as the study of human cognitive function.

The observer's paradox: a brief summary

In order to explain the *observer's observer's paradox*, we first summarise the 'observer's paradox', and note its pervasiveness across the sciences. The observer's paradox is the notion that intervention or measurement by an observer can directly impact (or coordinate with) the behaviour of the system being studied. Some of the earliest interpretations of quantum mechanics suggest that all properties of a system cannot be known at any given time, and only through probabilities are we able to adequately conceive of possible states of a system at a given time. Yet, the very act of observing these states will result in measuring only one specific state of that system, when the system (pre-observation) may be best described as multiple probabilistic states simultaneously (e.g. Howard, 2004).

This problem of observation and measurement has produced a series of paradoxes. For example, observation may very well alter the rate at which subatomic particles decay. Both theoretical and empirical investigations indicate that repeated observation can slow the rate of particle decay to a specific limit. As the number of observations reaches infinite, the particle, theoretically, may never decay. This has become known as the 'Quantum Zeno Effect' (Misra & Sudarshan, 1977). In addition to this, extensive discussion on observation and measurement has figured into quantum theory, from Schroedinger's cat, to more recent considerations of the inextricable involvement of the observer in physical systems (see Stapp, 2009). Of course, the area of expertise of the current authors limits review of this issue here; this brief summary simply points to the issue at the 'lowest' levels of organisation, such as subatomic physics.

Scaling up, the observer's paradox pervades cognition itself, holding even within individual cognitive processes such as speech production. A key feature in normal speech production is specific auditory feedback: 'self-observation' at a specific temporal scale. If the timing of auditory feedback from one's own speech production is misaligned, one can induce a stutter or stammer. Some have argued that normal speech production hinges on auditory feedback during language development whereby continued auditory feedback, post-development, aids in maintaining the 'internal model's' feedback system (Perkell et al., 2000). Given the recursive architecture of the speech production system, it seems impossible to avoid an observer's paradox at the level of individual cognitive processes, in which multiple cognitive components coordinate by being mutually responsive to each other's states. Such patterns can be found in other cognitive processes, such as controlled reaching (Miall & Wolpert, 1996), and multi-sensory processes (e.g. visuo-haptic illusion: Bicchi, Scilingo, Ricciardi, & Pietrini, 2008; auditory-visual illusion: Shams, Kamitani, & Shimojo, 2000). At multiple levels of analysis, it seems that an observer's paradox is not simply unavoidable, but even necessary.

While this cognitive form of the observer's paradox is quite different from what readers recognise, a more commonly recognised sense of the paradox has won significant attention in the social sciences. This form of the paradox is a prominent one, having ubiquitous relevance, at virtually every level of human behavioural measurement. A historically prominent articulation of the paradox is found decades ago in sociolinguistics. In a classic study, Labov (1966) showed that when patrons of different department stores were asked to indicate the location of specific objects located on the 'fourth floor', after the experimenter intentionally acted as if he/she had not understood their initial response, the articulation of /r/ sounds in patron responses 'fourth' and 'floor' became more pronounced. Changes in one's pronunciation after social monitoring support the notion that observation of behaviour, at the sociolinguistic level at least, leads to alterations in the behaviour itself. Since Labov's original study, various replications over the past decades have shown the same 'change from above' – alterations to one's own speech due to becoming more self-aware (Becker, 2009). The aim was to capture linguistic information from

individuals who were not under explicit observation, since the mere presence of an observer could have these effects (Labov, 1966). The idea, in some way or another, is that the observer always intervenes when investigating language users in a controlled fashion (sometimes termed the "Hawthorne effect").

This Labovian type of observer's paradox also influences low-level cognitive processing. The extensive work of Kingstone and colleagues has shown how social cues and social presence have a strong effect on micro-behavioural characteristics, such as visual attention (Friesen & Kingstone, 1998; Laidlaw, Foulsham, Kuhn, & Kingstone, 2011; see also Crosby, Monin, & Richardson, 2008). This social observational effect reveals that, when under potential observation, our cognitive system (such as visual attention) may utilise distinctly different strategies to scan the world. So cognition is, in a broad sense, embedded in observation processes, within an individual cognitive agent and across social cognitive agents.

This paradox, that observation is both necessary yet disrupts the system under study, spans the entire scientific spectrum, from natural to social sciences. If our goal is a scientifically explainable 'reality', it must operate under these constraints, tinted always through the lens of observation. In what follows, we focus on how differences in the use of observation as a theoretical and methodological lens of scientific observation inherently alter our access to – and thus our understanding of – a system's underlying structure.

The observer's observer's paradox

The observer is definitive of the scientific enterprise, but it may also be definitive of the cognitive processes of observers themselves. This observer's paradox is not based on contradictory outcomes of a particular mode of thought, thought experiment or some other such conceptual juxtaposition. The paradox instead relates to concern about how the observer is going to intervene on a system, thereby influencing the system, and affecting the regularities that can come out of those measurements. It is a paradox because it is fundamentally problematic, yet unavoidable.

In this section, we wish to extend this problem, into the domain of 'paradox' more commonly construed, namely, bringing about contradictory or counter-intuitive outcomes from comparable situations. Our extension relates to the decisions surrounding how a scientist goes about observing, and measuring, a system. Such decisions are not always made by force of volition; instead, researchers tend to pour effort into an agenda that has some history, and success, and guide future observational strategies (famously: Lakatos, 1978).

Let us consider an example, the debate between 'linear versus non-linear' perspectives on our cognitive system. Recently, there has been some discussion on whether reaction-time experiments offer sufficient means to uncover the organisation of the system (e.g. Holden et al., 2009). There is a long history to the agenda of calculating reaction-time measurements from human beings in particular decision tasks. From 'mental chronometry' and beyond, thousands of publications per year now use this measure. The use of reaction time takes on a variety of forms, but is organised around particular agendas, and so these forms can be quite systematic. In its simplest experimental use, a series of reaction times is collected from subjects in two conditions: A and B. The average response latencies are then compared through null hypothesis test, commonly under particular theoretical suspicions that one condition will be faster than the other.

Such a research agenda makes particular assumptions about the cognitive system. In fact, the measurements themselves may be theory-laden in the strong sense (see Schindler, 2012 for recent discussion of 'theory-ladenness'). The standard agenda assumes, for example, that cognition is a linearly decomposable computational system (especially those still using the

subtractive method; Posner, 2005), and that idealised laboratory performances have external validity to the cognitive process proposed (Banaji & Crowder, 1989).

We do not wish to challenge this agenda, as it is the bread and butter of a wide array of successful sub-fields in the cognitive sciences. However, one may wonder whether there are other circumstances that may be explored that violate the 'linear' assumptions. For example, do the processes supposedly reflected in reaction times of a stable lab performance relate to performance in a highly structured natural environment (Neisser, 1991)? Or, when stable cognitive performance is observed over a longer range than in the common laboratory task, perhaps reaction-time distributions reveal that performance is not so linearly decomposable (Holden et al., 2009; Kello et al., 2010; van Orden, Holden, & Turvey, 2003)?

These concerns do not lead to questioning the result that, for example, Condition A induced faster reaction times than Condition B in a given experiment. Instead, these concerns speak to different issues about the composition of the cognitive system and the capacities that the cognitive system reveals under varying measurement and task variables. A researcher interested in how a human performs under certain task constraints and demand characteristics may be justified in utilising assumptions of linear decomposability. This is because making such measurement assumptions provides vast gains in explanation and prediction, gains that are *not* similarly made by recent challenges to this linear assumption.

These are different measurement contexts, different goals of inquiry, different questions that are 'posed of the system'. In traditional reaction-time experiments that utilise statistical techniques, which assume a linear decomposition of variability, the researcher is typically interested in how a set of task variables (such as stimulus type) lead to facilitation (or hindrance) of responses. The second and non-traditional approach, sometimes termed 'non-linear methods' or 'interaction-dominant dynamics', is motivated by fundamentally different questions, such as how the cognitive system brings about organised performances, and what kind of underlying system would do so (e.g. Dixon, Stephen., Boncoddo, & Anastas, 2010; van Orden et al., 2003). The two seem to produce mutually incompatible perspectives on the cognitive system: linear decomposition versus non-linear interaction-dominant dynamics.

This is where the observer's observer's paradox comes in. In this debate example, linear versus non-linear approaches, researchers take up a specific suite of investigative techniques, and theoretical assumptions, in carrying out their enterprise. In doing so, they still face the traditional observer's paradox; they will remain mindful of how, even after they have chosen their carefully developed measurement methods, they are intervening and influencing the system while measuring it (Holden, Choi, Amazeen, & van Orden, 2011; van Orden, Kello, & Holden, 2010). So we are proposing two processes of coordination here, logically separate and in some sense organised in the form of two steps. First, the scientist is embedded in a tradition of particular theoretical questions and concerns on the one hand, which become 'coordinated' (in the sense of van Fraassen, 2008) with the methods, materials and models on the other hand. The result is that the scientists do not *just* go out with specific goals in mind for understanding the system, but also with a coordinated set of goals and procedures, which, together, dictate the very space of things that can be said about a system. Cartwright (1999) discusses similar issues, by describing how scientific endeavours work to create explanatory boundary conditions, situations that she refers to as 'nomological machines' – law-making measurement schemes.

This is what we mean by the observer's observer's paradox. Our growing scientific endeavour has an incredible capacity to generate novel, diverse, highly heterogeneous coordinations of goals and procedures (Suppes, 1981). By doing so, when studying precisely the same system (such as the human cognitive system) these coordinations can bring about apparent (and possibly real) incompatibilities: theoretical frameworks developed to account for the data

that are collected under those auspices. Our primary aim in the remainder of this paper is to develop an existence proof from computational simulation. As we describe next, there is an elegant computational framework that can explore an idealised 'observer' as a 'theory-builder', known as 'ϵ-machine reconstruction'. By modifying this simulation framework slightly, we can explore the consequence of changing measurement schemes in building such theories. Though ours is a highly idealised application of this framework, it exemplifies in computational form what we have just argued here.[1]

Pattern discovery: ϵ-machines

We sought to develop a very simple demonstration, through computational simulation, of the observer's observer's paradox. Such a demonstration requires a few characteristics. As described above, the paradox relates to how a scientist coordinates research goals with a set of materials and procedures in order to study a system's behaviour. So we need a computational paradigm that naturally involves these key elements.

A formal framework that intuitively touches on these characteristics has been developed by Crutchfield and colleagues (Crutchfield, 1994; Crutchfield & Young, 1989; Shalizi & Shalizi, 2004; Shalizi, Shalizi, & Crutchfield, 2002). Crutchfield (1994) introduced a graph-induction framework that he originally referred to as *hierarchical ϵ-machine reconstruction*. This computational framework has been applied to many problems, leading to a rapidly growing literature closely affiliated with physics, complexity science, applied mathematics and mathematical statistics. This approach has been termed 'computational mechanics', because it centres on the idea that all systems, even those systems describable through continuous non-linear dynamical systems, have intrinsic computational properties (Crutchfield, 1998). The ϵ-machine can serve as a characterisation of the computational complexity of a system. The machine is, in essence, a hidden Markov process that is induced from a series of measurements. An analogy with the observer/scientist is one that motivated Crutchfield's (1994) early discussions:

> How can an agent detect structure – in particular, computation – in its measurements of the environment? To answer this, let us continue with the restriction of discrete-valued time series; that is, the agent reads off a series of discrete measurements from its sensory apparatus. (p. 25)

The ϵ-machine approach works in the following way. An agent takes measurements from its environment. The agent then scans subsequences of observed measurements of length L or less. *Different* subsequences that lead to *similar* futures define a set of states that are in an important sense 'future equivalent', and are classed together as a set of the so-called 'causal states'. Because such states are approximately equivalent in the way the system evolves from them, the system's behaviour can be described as transitions between causal states. In Crutchfield's (1994) terms, they are 'the set of subsequences that renders the future conditionally independent of the past' (p. 26). Here, 'conditionally independent' means that, given a set of *different* subsequences of length L or less, the future behaviour of the system is highly similar (i.e. the subsequences that define causal states, though different, proceed according to the same future).[2] These sequences of causal states are effectively the ϵ-machine.

For readers unfamiliar with this framework, let us take a look at some highly simplified examples of ϵ-machine construction.[3] As noted, measurements in the ϵ-machine context are taken to be sequences of elements from a finite set of discrete states. For the purpose of demonstration here, let us take dichotomous states, {0, 1}, and imagine that some agent A is extracting sequences of such dichotomous states, which reflect the behaviour of a system at a given time point ({0,1} is often referred to as the measurement's 'alphabet'). Consider these

three time series of measurements:

(i) 1 1 0 1 0 0 1 0 1 1 0 0 1 0 1 1 1 0 1 0 0... random($\{0,1\}$)^n;
(ii) 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0... $(01)^{n/2}$;
(iii) 1 1 0 1 1 1 1 1 1 1 1 0 1 1 0 1 0 1 1 1 0... 1^k0, $k \in \{1,..., 25\}$, $P(k) = 25^{-1}$.

On the right of each time series (i)–(iii) is the specification of its structure. Let us imagine that each time series above continues with its rule for $n = 2000$ observations. Following the 'collection' of these data, the agent then constructs a model, the ϵ-machine, representing its internal structure. Here, we use the Causal–State Splitting Reconstruction (CSSR) algorithm (see Shalizi & Shalizi, 2004) developed for MATLAB by Kelly and colleagues (Kelly, 2011; Kelly, Dillingham, Hudson, & Wiesner, 2012). This algorithm is rooted in the goals of computational mechanics, and so we will continue to refer to the output of the algorithm as an ϵ-machine. However, it is important to note that significant details have been worked out since Crutchfield's original formulation. For example, subsequent papers have demonstrated the framework's potential to serve as a statistically optimal predictor (Shalizi & Moore, 2003; Shalizi et al., 2002), and in some cases the models are simply referred to as 'causal state models' (Kelly et al., 2012).

As noted above, in order to get the agent (simulated 'observer') kick-started in constructing an ϵ-machine of these time series, the agent A must choose a sequence size L over which subsequences will be examined. There are recommendations for this (Kelly, 2011), and consistent with our own goals here, Crutchfield (1994) likens this issue to properties of the agent, such as measurement capabilities or memory. For purposes of demonstration, let us choose a size $L = 4$. In time series (i) there are a large number, 2^4, of possibilities, given it is a random sequence from the alphabet {0,1}. Time series (ii) has just 2 possible sequences of size 4: {(0,1,0,1),(1,0,1,0)}. In time series (iii), there are 8 possible sequences: {(1,1,1,1),(1,1,1,0), (0,1,1,1),(1,0,1,1),(1,1,0,1),(0,1,0,1),(0,1,1,0),(1,0,1,0)}. As noted above, the CSSR algorithm also scans sequences of length less than L, of sizes 1 to $L-1$, as candidate's causal states. Those sequences are easily seen within the longest (L) sequences shown above. These are simply demonstration time series, and are transparent in how they evolve. The idea of the ϵ-machine,

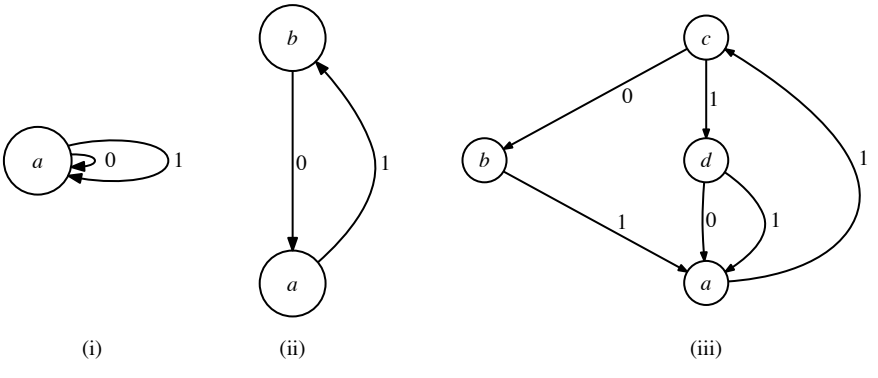

Figure 1. The three ϵ-machines induced from time series (i)–(iii) described in the main text. (i) is highly simple, reflecting an output of 0 or 1 with about equal probability at each time step (i.e. random); (ii) exhibits regular transitions from 0 to 1 and vice versa; (iii) is more complex, and following the causal states, and the corresponding symbol (0/1) output, one can see this ϵ-machine producing sequences of 1 with a single intervening 0. Length and shape of the edges (arrows) are only for presentation purposes.

instead, is to visualise their behaviour as a sequence of causal states. When we run the CSSR algorithm, the ϵ-machines obtained are presented in Figure 1.

The ϵ-machine construction process starts by defining an initial causal state (labelled *a*) and assumes by default that all subsequences of size *L* or less are equally predictable of the future (in other words, they are part of the one, single causal state).[4] The algorithm uses a distribution test over what occurs in the time series after each subsequence, and determines whether the observed future differs significantly from what is predicted from this initial causal state *a*. If so, the CSSR algorithm begins to create new states iteratively, splitting the set of subsequences of size *L* or less into separate, and new, causal states (labelled *b* and beyond). This process stabilises[5] on a hidden Markov structure that characterises the probabilistic transitions of the system in terms of 'hidden' causal states (Kelly et al., 2012). In Figure 1, the edge labels show which observed value {0,1} is output as the system transitions between causal states.[6]

Importantly, each of these causal states is associated with a set of sequences of some length *L* (here, 4) or less. The causal states and their subsequences, for these toy time series, are shown in Table 1. In the simplest ϵ-machine, (i), the sequences of characters are completely random. The CSSR process estimates a single causal state, under which all subsequences of size *L* or less are classed (Table 1, column 1). The second time series induces a second causal state, reflecting whether the string ends with 0 or 1, which will determine the future of this sequence (Table 1, column 2). Finally, time series (iii) induces several more states. When following the edges in Figure 1, (iii) can be seen to reconstruct the rule described in the time series' definition above. Following causal states *a-c-d-a* produces a long string of 1's, instantiating the rule '$1^{[1-25]}0$', which we specified above in time series (iii).

This framework has been applied to a variety of domains. For example, one major early application of the framework was to offer a new means of understanding statistical complexity. Crutchfield and Young (1989) applied an early version of ϵ-machine construction to the well-known logistic map. The logistic map is a discrete-time dynamical system, $x(t+1) = rx(t)(1 - x(t))$. The equation has one parameter that determines the behaviour of $x(t)$, r, referred to as its control parameter. Under different values of r, the logistic map showcases a variety of behaviours characteristic of non-linear dynamical systems, including onset to chaos through a so-called period-doubling process that can be seen in Figure 2 across a range of values of r.

Crutchfield (1994) likens this measurement process, over the logistic map, as an exploration of its intrinsic computational capacities. From this perspective, the ϵ-machine construction process can be considered the generation of a 'theory' of the map's behaviour at a given value of r. Because we will also be using the logistic map in this way, we will showcase it in some detail here. Figure 2 shows the different values of $x(t)$ that occur in the logistic map when it is iterated at different values of r.

Despite the simplicity of its specification, this map is of course attractive for the complexity it exhibits within this range. Below the plot we show three arrows, specifying values of r for which 500 observations[7] were extracted as the logistic map was iterated. As described above, the CSSR algorithm, to build the ϵ-machine, requires a discrete alphabet. This would obtain a series of 0's and 1's much like our toy time series, but here from the logistic map. Using the 'generating partition' of 0.5, we obtain the symbol 0 from $x(t)$ if it is below 0.5, and 1 if it is above 0.5. The result is a long string of 500 measurements from alphabet {0,1}. Below the values of $x(t)$ in Figure 2, we show the ϵ-machines constructed at these values of r. With $r = 3.85$, the map has a substantially lowered period (fewer possible $x(t)$ values upon iterations), and the ϵ-machine nevertheless estimates a system that has a similar level of computational complexity to 3.8, with $r = 3.9$ apparently the most complex graph structure among these three.

Table 1. The subsequences of 0's and 1's that are grouped in the same causal state.[a]

(i)	(ii)	(iii)
State a 0,1,00,10,01,11, 000,100,010,110, 001,101,011,111, 000,1000,0100, 1100,0010,1010, 0110,1110,0001, 1001,0101,1101, 0011,1011,0111, 1111	State a 0,10,010,1010 State b 1,01,101,0101	State a 0,1010,0110,1110, 0101,1101,1011, 0111,1111 State b 1 State c 010,110 State d 01,11 State e 101,011,111 State f 10

[a] Note that the causal states in Table 1 may be more numerous than those seen in Figure 1. This is because the CSSR toolbox trims non-recursive network segments in the final graph visualisation. The recursive structure of the ϵ-machine is more likely to characterise the system's long-range behaviour, so this visualisation was used throughout. For the simulations we describe below, we only quantified over the final recursive network per simulation, though sometimes the reconstruction process returns two (and very rarely more than two) recursive structures.

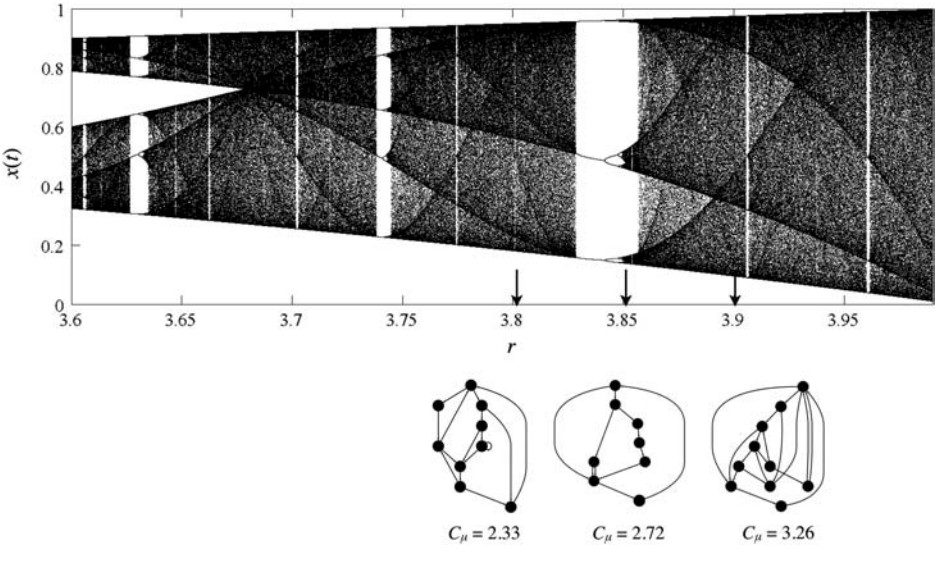

Figure 2. Top panel: The values of $x(t)$ that occur when iterating the logistic map at different values of r. The arrows represent where ϵ-machines were constructed with 500 observations, converting the $x(t)$ time series to symbols from alphabet {0,1}, as specified in the text. Bottom panel: The ϵ-machines induced from those control parameter values. Values below the machines are average statistical complexity scores based on $N = 5$ machines (C_μ, see below for statistical complexity).

This framework has been applied to many domains (see Crutchfield, 2011, for a recent review). It is rooted in characterising the statistical complexity of a system, as inferred from long sequences of measurements. It has allowed Crutchfield (1994) to explore emergentism, Shalizi and Moore (2003) to describe collective macro-states emerging within physical systems, and of course exploration of a variety of physical systems and systems of equations. After this brief introduction to ϵ-machine construction, and showcasing the CSSR algorithm in basic form, we now create a simulation with these aspects as our foundation. Motivated by the centrality given to measurement in van Fraassen (2008), we will refer this simulation as imposing 'measurement models'. It is a very simple variant of the ϵ-machine, but not by changing CSSR's innards, but rather by changing the manner in which the observing agent takes measurements.

Simulating the observer's observer's paradox: measurement models

So far, we have provided a very brief introduction to ϵ-machines (Crutchfield, 1994) by applying the CSSR algorithm (Shalizi & Shalizi, 2004) in MATLAB (Kelly et al., 2012). Here, we will modify the set-up just slightly for the purpose of demonstrating the observer's observer's paradox. As described above, the ϵ-machine has a key 'cognitive' parameter that is chosen (L). When we described the paradox earlier in this paper, we discussed a much wider array of characteristics of the scientist. These included broader characteristics, such as explanatory goals, which relate to more specific aspects, such as the temporal and spatial scales of a measurement (e.g. reaction time). Here, we will implement something akin to this directly, by defining the agent as a function $A(x(t))$ that subjects a raw time series to some transformation or manipulation prior to constructing the ϵ-machine. This $A(x(t))$ function is similar to the 'measurement model' discussed in van Fraassen (2008), in the sense that the agent has some intrinsic goals or interests in understanding a system within some specific contextual constraints:

> ... the representation (the measurement outcome) shows not what the object is like 'in itself' but what it 'looks like in that measurement set-up.' The user of the utilized measurement instrumentation must express the outcome in a judgment of form 'that is how it is *from here'*. And finally, the coin has another side: it is precisely by a process engendering a judgment of that form – that is to say, by a measurement! – that any model becomes usable at all. (p. 86)

In the van Fraassen (2008) view of the scientific agenda, theory and measurement become coordinated, with active involvement of the scientist and his/her instruments, making measurement itself a representational agenda, all while still remaining appropriately inter-subjective.

These goals of the researcher take the form of a model for measurement itself – the agent imposes this model in the context of inquiry. The notion of the observer's observer's paradox is that the selection of this function may radically impact the nature of the theory induced from the data. We will call this manipulation simply the imposition of a 'measurement model', and it is represented diagrammatically in Figure 3.

Let us return to the logistic map described above. This will be our 'raw system', over which a measurement model will be applied by some agent A. The model we choose is motivated by very common strands of debate in cognitive science: temporal aggregation (see review in Riley & van Orden, 2005). To generate a derivative time series from the original sequence of measure $x(t)$, we do the following. The agent chooses a window size, w, with which to aggregate symbol sequences from alphabet {0,1}, using the 0.5 partition described above. Within each window of size w, the agent maps the time series back onto the same alphabet {0,1} based on whether 1's are more dominant within that window range than 0. This is not an arbitrary possibility. It is common in analyses of human systems to carry out aggregate tallies of events and behaviours in

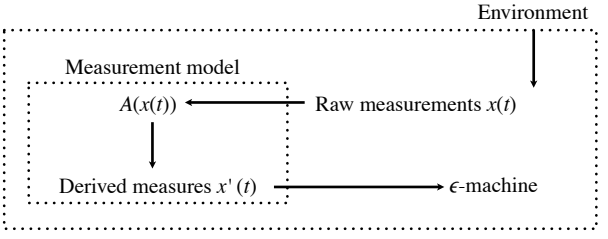

Figure 3. The general form of the modification, in which a model of measurement through the agent, $A(x(t))$, is applied over the system's raw measurements (raw appearances) in a fashion that generates a goal-oriented 'measurement representation', a derived time series, $x'(t)$, which is then used for ϵ-machine construction (drawing some inspiration from Figure 1 in Crutchfield, 1994).

the interest of understanding broader temporal-scale organisation.[8] This specific instantiation of the measurement model is depicted in Figure 4.

Here, we showcase what happens to ϵ-machines when such an agent-centred function transforms the logistic map's output prior to inducing the machine. In order to characterise the influence on the machines, we look to Crutchfield and colleagues' measure of statistical complexity, which is output by the CSSR MATLAB toolbox (Kelly, 2011):

$$C = -\sum_s P(s) \log_2 P(s)$$

Here, C will be referred to as 'statistical complexity', and $P(s)$ is the probability of occurrence of a particular causal state (s) in the graph. This is akin to an information measure, 'the amount of information the process stores in its causal states', (Crutchfield, 2011, p. 20) or alternatively, the number of bits required to reconstruct the original time series of observations using the causal model (Shalizi & Moore, 2003).

We simulated an 'experiment' in which an agent A extracts dichotomous observations ($n = 250$) from the logistic map under a given value of r, then builds a 'theory' in the form of an ϵ-machine (with $L = 4$, and $\alpha = 0.05$).[9] The agent runs $N = 20$ of these models under some value of w, and calculates the mean statistical complexity, C_μ. We used an r range of 3.8–3.9, as illustrated above, using 10 evenly spaced values of r within this range, which include a transition from chaos into a new period-doubling cascade, and back into chaos. A few other details are important here. The measurement model $A_w(x(t))$ was instantiated with the values $w = 1,..., 5$. In addition, we imposed a restriction on reconstruction of the machines. Only observed time series that had at least 10% of the least frequent observation (1 or 0) was used. For example, a high value of w may coarse-grain the system so extensively that the observations extracted 'collapse'

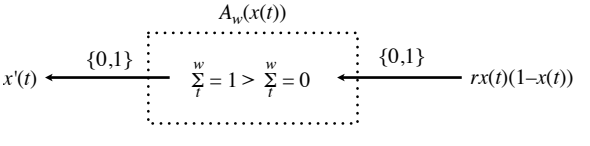

Figure 4. Depiction of the measurement model used in this demonstration. The agent function $A_w(x(t))$ involves taking observations of symbols {0,1} and tallying whether 1 occurs more than 0 within a window range w. If so, the system outputs a 1, if not, a 0. This returns a new sequence of 1's and 0's depending upon the relative density of these events within some coarser time scale. Windows, w, are genuine 'cognitive' sampling constraints: They are non-overlapping during measurement.

into the most frequently observed value. For a given value of r, all experiments started with the same random value. Each logistic run was seeded with 3000 iterations. All time series had the same length (250) which means that, for example, a $w = 5$ experiment required the collection of many more 'raw' values (which the agent 'sees' only in the form of a down-sampled 250-element time series of 1's and 0's).

Figure 5 shows the region of the logistic map explored, and C_μ observed across w and r combinations. What occurs is relatively straightforward, but we wish to provide a gloss over these results that hearken to the observer's observer's paradox. First, when window size is $w = 1$, the ϵ-machines revealed approximately increasing statistical complexity towards $r = 3.9$, which is statistically significant using simple linear regression, $R^2 = 0.52$, $F(1,198) = 215.9$, $p < 0.0001$. Importantly, $w = 1$ theoretically resembles a scientist whose measurement scheme theoretically fits the smallest possible scale of analysis available to him/her. However, as window size increases, the relatively more complex ϵ-machine at $r = 3.9$ rapidly reduces in complexity, and the pattern inverts. The ϵ-machine comes to behave more like the form of that for time series (i) in our toy examples above: it is estimating with some base-rate probability transitions from 0 to 1. This negative relationship is statistically significant, $R^2 = 0.38$, $F(1,98) = 60.15$, $p < 0.0001$. What this result overall suggests is that there is an interaction between the 'real' behaviour of the logistic map and the 'imposed' constraints of the measurement scheme. This can be statistically tested in the form of a centred interaction term, added to a larger regression model (using all w's), and it is indeed significant: $\beta = -0.35$, $t(750) = -7.5$, $p < 0.0001$.

Put simply, the statistical complexity of ϵ-machines of this particular measurement model is *not* proportional to the key parameter of the measurement model, w. Measurement choices matter. In fact, at the same measurement granularity w, a process that looked more complex before appears maximally simple now ($r = 3.9$ at $w = 5$); meanwhile, a relatively simpler process (near $r = 3.85$) retains robust statistical complexity characteristics. A measurement model can *invert* the computational inferences of the logistic map's behaviour over different values of r. Put differently, if two researchers studying this non-linear system were using different measurement schemes, they would come to different conclusions about its computational characteristics in their experiments.

Our description of the observer's observer's paradox was based primarily on an intuitive basis. The regularities that emerge from a scientific agenda are driven by the observer's choices of measurement scheme. The paradox derives, we proposed, from the wide heterogeneous array of goals and motivations and measurements, and thereby a diverse array of regularities that may seem mutually incompatible. Here, we have shown, in a simulation of the measurement choices of a highly simplified 'observer agent', that the same sorts of situations emerge. The logistic map looks different when a process of aggregation is applied in a measurement scheme. Future applications of the above model aim to determine the meaning of changes (or lack thereof) in complexity over variations in measurement models as characterised by $A(x(t))$. For example, the reduced period of the logistic map near $r = 3.85$ may of course be the reason for its preserved statistical complexity over the aggregation parameter, w. So, the discourse regarding measurement model must also factor in, to a great extent, some assumptions about the system itself. Here, we have focused on the measurement model, but clearly more work could be done to strengthen this simple simulation as a demonstration.

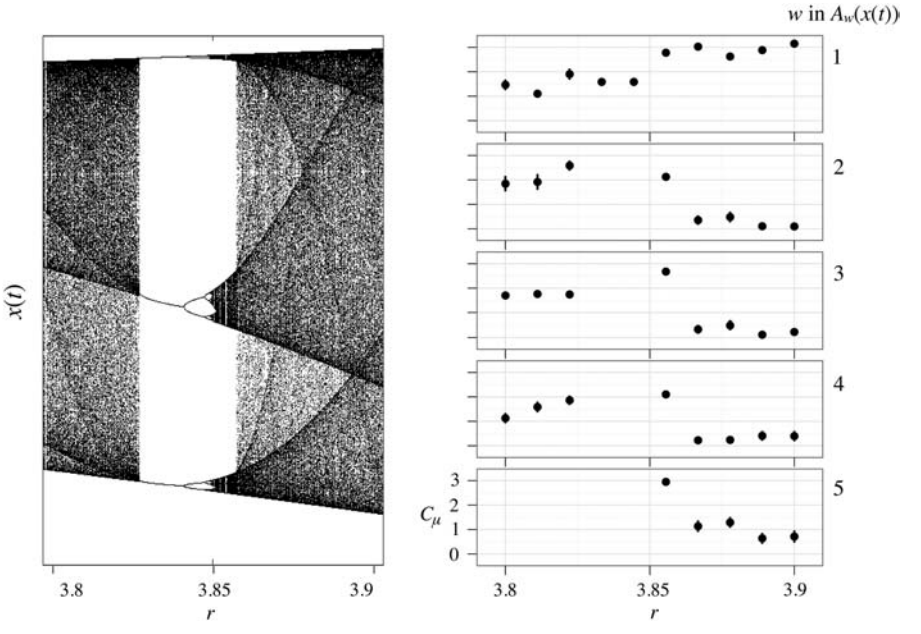

Figure 5. Left side: The region of the logistic map explored with the measurement model $A_w(x(t))$. Five values of r were chosen between 3.8 and 3.9, and 250 observations, {0,1} were collected for ϵ-machine reconstruction ($L = 4$). Right side: As window size (w) increases from 1 to 5, the measurement scheme in $A_w(x(t))$ produces considerably different outcomes in the statistical complexity of ϵ-machines induced from the derived series $x'(t)$. As r increases within this range, in general statistical complexity as measure in the CSSR toolbox increases; however, as the coarseness of sampling increases, the pattern inverts. Lower panels in this plot show no points in the range near 3.8 because these time series stabilised into either many 1's or many 0's given the value of w and a machine was not built for these window sizes (either 1 or 0 came to dominate in the derived series $x'(t)$; see main text). Error bars, difficult to see among some values, are standard errors based on $N = 20$ simulations.

Summary, limitations and conclusion

In summary, we have argued for an additional paradox over and above the well-known 'observer's paradox': Before a researcher even comes to grips with the manner in which their specific measurements disrupt systems, a lot of the *possibilities* for theoretical regularities have been to some extent decided by *coordinating* theoretical goals and measurement models (Giere, 2006; van Fraassen, 2008). We presented a computational model that, while highly idealised, demonstrates that a change in measurement parameters can cause changes in the inferences made about a system. In the realm of cognitive science, such measurement goals and procedures are not trivial, nor are they random or unsystematic. In the model we presented, one might imagine contexts within which larger w parameters reflect a particular domain of exploration that tie deeply into broader questions about system function. The example we provided of reaction-time research suggests that cognitive science does this all the time.

From low-level concerns with perceptual physiology to higher-level explorations of reasoning and decision-making, the human cognitive system is studied from a diverse set of temporal and spatial scales. Along with the heterogeneity of theoretical goals and measurement procedures, the result is paradoxical in the following sense: diverse, perhaps seemingly mutually

incompatible, formalisms or frameworks may coexist across these scales to help us understand aspects of the same system (Shalizi & Moore, 2003).

Our argument is inspired by previous work. As we described above, a similar perspective has been offered by Cartwright (1999). Dennett's 'stances' (1998, 1999) also bear a close relation. In fact, many philosophical perspectives on explanatory pluralism encourage better understanding the epistemological goals of observers as crucial to theoretical mitigation (for reviews see Anderson, 1972; Cartwright, 1999; de Jong, 2002; Dupré, 1993; Giere, 2006; Kellert et al., 2006; Kelso & Engstrøm, 2006; McCauley & Bechtel, 2001; Mitchell, 2003; Putnam, 2004; Suppes, 1981; Tabor, 2002; Weiskopf, 2009). Here, we have shirked a wide variety of issues that are relevant to the discussion, such as the ontological status of the resulting theories and their constructs (e.g. ontological promiscuity: Dupré, 1993; emergence: Atmanspacher, 2007; and beim: Graben, Barrett, & Atmanspacher, 2009), or the diverse patterns of inter-theoretical mapping that may be explored both in current scientific theory and perhaps even computational models (Dale & Spivey, 2005; McCauley & Bechtel, 2001).

There are some obvious issues and limitations of the simulation that we have presented in the previous section. Admittedly, it is a simple demonstration of the issues we have described. However, we believe some of these critiques can be anticipated and responded to readily.

(1) *Sophisticated analysis of the logistic map may allow derivation*. We seem to have assumed that there are no tractable derivations of relative statistical complexity under the varying parameters that may be chosen (e.g. L, or w in the measurement model). This is not true in all cases. For example, the relationship between the number of states and the subsequence length (L) can be derived (states increase as L increases; Crutchfield, 1994). This leads to the immediate suspicion that perhaps, for example, the behaviour of the measurement model, under w, can be derived in some fashion from our understanding of the logistic map. In this vein, someone may say that 'the measurement model cannot be inferred to be theoretically (mathematically) independent of the overall system as we understand it'. While this is beyond the scope of this article, and is anyway beyond the (admitted) limited expertise of the current authors, it raises two potential issues. Let us assume for the sake of discussion that measurement models indeed serve as analogies to procedural schemes in cognitive science (more on this below). First, if it is indeed true that the relationship among statistically disparate measurement contexts may be derived in a unified account of its behaviour, then the upshot is that something similar may happen to theories of cognitive science. This is something we wholeheartedly agree with (see discussion in concluding paragraph below). Second, if it cannot be so derived, then the theories are fundamentally distinct, driven by potentially wide array of behaviours of $A(x(t))$. In either case, we face the observer's observer's paradox. In the former possibility it is a tidy mathematical cleanup, but it would still, presumably, require integrating the behaviour of the observer (the measurement model) to derive it.

A wonderful recent example, far more formally sophisticated than the current presentation, is found in Tabor (2009). Here, by exploring a class of models referred to as 'affine dynamical automata', Tabor demonstrates that a simple dynamical system framework can give rise to the full array of computational behaviour in the Chomsky hierarchy, and beyond – advocating in the end a kind of super-Turing computation that challenges classical conceptions of symbolic computation as being simply compatible with more dynamical approaches (such as neural network models). Tabor's presentation offers proofs of these inter-relationships (for another closely related exploration, see Kolen & Pollack, 1995).

(2) *The measurement-model outcome does not result in fundamentally different theories*. It is true that the example we have generated does not produce fundamentally different theories, really. What it produces are two theories that lie on opposite ends of one characteristic (here, statistical complexity). One could imagine, however, pitting different theory-construction procedures against each other, and exploring their behaviours in different measurement contexts (e.g. Dale & Spivey, 2005). In a sense, this makes the demonstration even stronger: Despite the agent 'working' under the same general theoretical apparatus, ϵ-machines reveal disparate patterns of behaviours under different measurement parameters (e.g. w and r). One would presume that differential theoretical apparatus would only compound the conundrum.

(3) *ϵ-machines look nothing like cognitive theories*. ϵ-machines behave a lot like classical cognitive theories, in fact. If anything, they at least resemble a kind of nascent exploratory situation, where the scientist is 'splitting processes' to capture a complex unfolding stream of data (e.g. see recent methods reviewed in Newell & Dunn, 2008). From the fractionation of working memory (Baddeley, 1998), to the discoveries of linguistic levels of organisation (Fromkin, Rodman, & Hyams, 2010), this is precisely what some cognitive scientists have done. It may be an abstract relation, but we argue that if anything the ϵ-machines work to constrain their theories from the data by proposing new kinds of states and transitions between them. Such a process is the *sine qua non* of classical cognitive science.[10]

(4) *The measurement-model demonstration does not reveal the functional importance of the exemplified measurement scheme, A(x(t))*. This is true. We have only offered a kind of observer's existence proof. But we argue that if it holds in the simplest of circumstances, then heterogeneous procedural assumptions operating over a vastly more complicated complex dynamical system (human cognition) show the same paradoxical behaviour. The purpose of the existence proof is to show the predicted paradox in a well-controlled computational circumstance that can be built on. Assuming, for the sake of modelling, the scientist/agent is a measurement model, one could apply a plethora of variables within the agent itself.

(5) *This is just the application of one theory (ϵ-machine reconstruction) to a variety of contexts.* It may be said that the simulations have not shown different theories emerging from different w's, but rather differential application of the same theoretical framework of computational mechanics. We have to acknowledge that this gloss over the simulations cannot be addressed robustly with the current results. We would respond that the theoretical differences should be attended to by reference also to w, rather than ϵ-machines and computational mechanics alone. If two researchers worked in disparate measurement contexts (akin to the example we provided above in linear vs. non-linear approaches), then they would make different claims about the system under similar circumstances (e.g. r values in the logistic map; or in a reaction-time laboratory study). The resolution to such a dispute could come from recognising their measurement model and associated goals, and identifying the basis on which the system seems to exhibit different capacities under different models. This leads to an important final concern.

(6) *What is real?* In Shalizi and Moore's (2003) discussion of macro-states and emergence, they identify computational mechanics as an arena in which the objective and subjective qualities of emergent states and levels of organisation can be explored. Their paper is challenging for readers unfamiliar with computational mechanics, but the implications of the work are substantial and have immediate relevance to this final issue: If one allows that disparate theories may coexist under different measurement models, *what is real* about these theories? Some philosophers have discussed this, of course. Dupré (1993) is 'ontologically promiscuous', arguing that nature has many levels of organisation with their own respective categories and exemplars that are worthy of the label 'real'. Cartwright (1999) describes

nomological machines (mentioned above) as revealing *genuinely real* 'capacities' that nature reveals under those circumstances. There are others. In cognitive science, 'symbol-like' properties are acknowledged even by some dynamical systems theorists, but these properties are often denigrated as 'mere approximations' to something at a lower level (which is often a researcher's favoured theory, that is yet an inexplicable explanatory floor not breachable into even lower levels – which it surely is).

Shalizi and Moore (2003) offer a provocative possibility: 'This also gives us a hint as to what constitutes a good set of macrovariables: it should not just be causally complete, but also *more predictively efficient than the microvariables'* (p. 9, emphasis added). They end by imagining a situation in which scientists are faced with the problem of experimenting with and building theories of an inexplicable goo. The scientists go about collecting data, proposing measures and macro-states and building causal models. In complex systems, such as their mystery goo or the human brain, there are many potential 'real' collective variables identifiable by measurement and causal modelling. And objective, quantifiable criteria such as *improved predictive* abilities (along with associated epistemological benefits) of macro-states can facilitate embracing them as 'real' (whatever that really means). As these authors conclude, 'for every question we ask It, Nature has a definite answer; but Nature has no preferred questions' (p. 11). Even here, however, we are stuck with the idea of whether (or which) theoretical constructs, abstracted or induced from data, can be given that delightful membership into ontological endorsement. Maybe the only best means of getting at this at all is through the proposals such as computational mechanics and clear definitions of emergence. Without such tools, we may be confined to ever more debate about ontology, of which some prominent philosophers have begun to tire (e.g. Putnam, 2004).

Maybe what is more amazing about the human cognitive system is not the array of sophisticated capacities it exhibits, but rather the fact that these capacities can tap into so many scales, and brands, of analysis; the human cognitive system is radically multi-perspectival (cf. Giere, 2006).

There are surely many improvements to be made, but we would argue that the general strategy with the simulation is borne out by the results: As measurement systems change, theories *can* change. This leads to paradox, through potentially contradictory frameworks emerging out of those measurement systems. Perhaps this characterisation of a 'paradox' is most evident from the vibrant debate that has taken place in the cognitive sciences. We may sometimes seem irked by the fact that a complex, multi-scale system such as human beings can be given to such diverse theoretical analysis, especially when one of these contradictory analyses is situated close to our domain of interest. Surely there will be battles at the edges of these concerns (McClelland & Patterson, 2002; Pinker & Ullman, 2002). But our argument here is that, at the very least, it may be useful to consider integrating the goals and coordinated measurement systems in understanding how these disparate accounts arise and are sustained.

Notes

1. A comment on the first submission of this paper noted that by advancing a model of this process, we are inadvertently taking up a particular brand of computational thinking (which may, e.g. disagree with some versions of dynamical accounts, some of which argue against computational instantiations). Addressing this issue requires more space, and is perhaps best suited to another paper, as this may raise interesting issues of self-referentiality in this kind of 'meta-modelling'. We would not be opposed, however, to considering how a cognitive scientist's agenda could take the form of a computational process in the traditional sense (after all, cognitive computation and dynamics are complementary; Edelman, 2008).
2. We are deliberately simplifying here for exposition, but there are more details important to this approach. Conditional independence also relates centrally to the Markovian characteristic of the

machine's state transitions: When an agent knows the causal state a system is in, the future is determined without having to consider other past states through which it may have transitioned.

3. We use 'reconstruction' and 'construction' interchangeably, though they have important and distinct connotations in this literature. Most often 'reconstruction' is used for this process, in the tradition of seeking means of 'recovering' or 'rebuilding' aspect of a complex system's behaviour (e.g. phase-space reconstruction; see examples in psychology in Riley & van Orden, 2005). The idea is that the machine that is induced from the data is 'reconstructing' the computational qualities inherent in the original system, as inferred from the series of measurements.
4. In fact, the causal states are labelled using numeric identifiers 0, 1, 2, etc. We have changed these to *a*, *b*,... to avoid ambiguity with the system's measured states (e.g. from the alphabet {0,1}, as we are doing in this example).
5. Here, stabilisation means that as the distribution test (Kolmogorov?Smirnov) is conducted, the CSSR algorithm no longer finds significant differences between predictions of the machine and the observed measurements. This requires the CSSR's only other parameter to be chosen, the significance level used by the distribution test (Kelly, 2011, recommends a low value, and for these toy series we chose the default $p = 0.005$).
6. The algorithm also includes a probability associated with each transition. We omit this here for simplicity.
7. We chose a smaller number of observations so as to facilitate running multiple machines to ensure the consistency of the converged model. In order to ensure the model had settled after random initialisation, we iterated it for 1000 runs before extracting the 500 measurements. We also chose a significance level of 0.05 as recommended by the CSSR toolbox to avoid closed communication classes (see Kelly, 2011). Finally, we chose these values of r so as to be situated near a phase transition in the map's periodic behaviour.
8. Sophisticated discussion of similar measurement scenarios is found in Shalizi and Moore (2003). The authors offer examples from the physical sciences, and describe how the causal-state approach, in the context of particular measurement goals and procedures, may provide a basis for defining emergence, and the relationship between micro- and thermodynamic macro-states of physical systems.
9. As noted above, the selection of these parameters in more realistic circumstances is guided by pre-existing conceptions of the causal model to be reconstructed, the length and nature of the observed time series, and any conceptual issues related to the system under study. $L = 4$ and $n = 250$ would both be considered relatively small. However, we choose these small values for two reasons. First, they reflect the level of complexity of cognitive theories, because as L increases, the number of causal states estimated increases; large values of L would thus induce unrealistic numbers of components in a 'cognitive theory'. The length of 250 observations is comparable to your average 'realistic' cognitive psychology experiment. Obviously whether these values are 'realistic' for the logistic map's 'true' underlying behaviour is not explored here. But we wish to stick to the idea that we are very simply demonstrating a scenario comparable to a human study, in which considerably less is known about a vastly more complex non-linear dynamical system. Besides, our second reason is just that these values vastly facilitate running the CSSR toolbox on a Macbook Air.
10. A distinction worth noting is that our demonstration relies on a kind of 'exogenous' exploration of inter-theoretical contrast. The scientist is arriving at disparate theories by imposing a measurement scheme 'from the outside'. Other approaches to relating, and integrating, disparate theoretical accounts are 'endogenous': They explore how the neural system may give rise to computational patterns from finer-grained 'microstates' (e.g. Smolensky et al., in press). It may be worth comparing these approaches. We would contend that the heterogeneity of goals and measurement schemes may bring about theoretical diversity even when there is no *obvious* endogenous basis for it.

References

Abrahamsen, A., & Bechtel, W. (2006). Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective. In R. Stainton (Ed.), *Contemporary debates in cognitive science*. Oxford: Basil Blackwell.

Atmanspacher, H. (2007). Contextual emergence of mental states from neurodynamics. *Chaos and Complexity Letters*, *2*(2-3), 151–168.

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8*(2), 234–238.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, *44*(9), 1185–1193.

Becker, K. (2009). /r/ and the construction of place identity on New York City's Lower East Side. *Journal of Sociolinguistics*, *13*(5), 634–658.

Bicchi, A., Scilingo, E. P., Ricciardi, E., & Pietrini, P. (2008). Tactile flow explains haptic counterparts of common visual illusions. *Brain Research Bulletin*, *75*(6), 737–741.

Cartwright, N. (1999). *The dappled world*. Cambridge: Cambridge University Press.

Crosby, J. R., Monin, B., & Richardson, D. (2008). Where do we look during potentially offensive behavior? *Psychological Science*, *19*(3), 226–228.

Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics and induction. *Physica D: Nonlinear Phenomena*, *75*(1-3), 11–54.

Crutchfield, J. P. (1998). Dynamical embodiments of computation in cognitive processes. *Behavioral and Brain Sciences*, *21*(5), 635. doi:10.1017/S0140525X98291734

Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Physical Review Letters*, *63*(2), 105–108.

Dale, R. (2008). The possibility of a pluralist cognitive science. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 155–179.

Dale, R., & Spivey, M. J. (2005). From apples and oranges to symbolic dynamics: A framework for conciliating notions of cognitive representation. *Journal of Experimental & Theoretical Artificial Intelligence*, *17*(4), 317–342.

de Jong, H. L. (2002). Levels of explanation in biological psychology. *Philosophical Psychology*, *15*(4), 441–462.

Dennett, D. C. (1998). *The intentional stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1999). Real Patterns. *Mind and Cognition: An Anthology*.

Dixon, J. A., Stephen, D. G., Boncoddo, R., & Anastas, J. (2010). The self-organization of cognitive structure. *Psychology of Learning and Motivation*, *52*, 343–384.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*, 412–431.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge, MA: Harvard University Press.

Edelman, S. (2008). On the nature of minds, or: Truth and consequences. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 181–196.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1997). *Rethinking innateness: A connectionist perspective on development*, *Vol. 10*. Cambridge, MA: MIT Press.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495.

Fromkin, V., Rodman, R., & Hyams, N. (2010). *An introduction to language*. Wadsworth.

Giere, R. N. (2006). *Scientific perspectivism*. Chicago, IL: University of Chicago Press.

Graben, P., Barrett, A., & Atmanspacher, H. (2009). Stability criteria for the contextual emergence of macrostates in neural networks. *Network: Computation in Neural Systems*, *20*(3), 178–196.

Holden, J. G., Choi, I., Amazeen, P. G., & van Orden, G. (2011). Fractal 1/*f* dynamics suggest entanglement of measurement and human performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 935–948.

Howard, D. (2004). Who invented the 'Copenhagen Interpretation'? A study in mythology. *Philosophy of Science*, *71*(5), 669–682.

Kellert, S. H., Longino, H. E., & Waters, C. K. (2006). *Scientific pluralism*. Minneapolis, MN: University of Minnesota Press.

Kello, C. T., Brown, G. D., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Van Orden, G. C. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences*, *14*(5), 223–232.

Kelly, D. (2011). Causal state modeller toolbox help file. Retrieved from http://davekelly377.weebly.com/code.html

Kelly, D., Dillingham, M., Hudson, A., & Wiesner, K. (2012). A new method for inferring hidden Markov models from noisy time sequences. *PloS One*, *7*(1), e29703.

Kelso, J. A. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT Press.

Kelso, J. A., & Engstrøm, D. A. (2006). *The complementary nature*. Cambridge, MA: MIT Press.

Labov, W. (1966). *The social stratification of English in New York city*. Cambridge University Press.

Kolen, J. F., & Pollack, J. B. (1995). The observers' paradox: Apparent computational complexity in physical systems. *Journal of Experimental & Theoretical Artificial Intelligence*, *7*(3), 253–269.

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, *108*(14), 5548–5553.

Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge: Cambridge University Press.

Leahey, T. H. (2001). *A history of psychology*. Upper Saddle River, NJ: Prentice-Hall International.

Marr, D. (1982). *Vision: A computational approach*. San Francisco, CA: Freeman.

McCauley, R. N., & Bechtel, W. (2001). Explanatory pluralism and heuristic identity theory. *Theory & Psychology*, *11*(6), 736–760.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279.

Misra, B., & Sudarshan, E. C. G. (1977). The Zeno's paradox in quantum theory. *Journal of Mathematical Physics*, *18*(4), 756–763.

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge University Press.

Neisser, U. (1991). A case of misplaced nostalgia. *American Psychologist*, *46*(1), 34–36. doi: 10.1037/0003-066X.46.1.34

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York/London: Academic Press.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*(8), 285–290.

Port, R. F., & van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Posner, M. I. (2005). Timing the brain: Mental chronometry as a tool in neuroscience. *PLoS Biology*, *3*(2), e51.

Putnam, H. (2004). *Ethics without ontology*. Harvard University Press.

Ręczaszek-Leonardi, J., & Scott Kelso, J. A. (2008). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New Ideas in Psychology*, *26*(2), 193–207.

Riley, M. A., & van Orden, G. C. (2005). *Tutorials in contemporary nonlinear methods for the behavioral sciences*. National Science Foundation.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, *Vol. 1*, (Foundations).

Schindler, S. (2012). Theory-laden experimentation. *Studies in History and Philosophy of Science*.

Shalizi, C. R., & Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, *104*(3), 817–879.

Shalizi, C. R., & Moore, C. (2003). What is a macrostate? Subjective observations and objective dynamics. *arXiv preprint cond-mat/0303625*.

Shalizi, C. R., & Shalizi, K. L. (2004). Blind construction of optimal nonlinear recursive predictors for discrete sequences. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 504–511.

Shalizi, C. R., Shalizi, K. L., & Crutchfield, J. P. (2002). Pattern discovery in time series, Part I: Theory, algorithm, analysis, and convergence. *Journal of Machine Learning Research*, 02–10.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788.

Smolensky, P. (2012). Symbolic functions from neural computation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *370*(1971), 3543–3569.

Smolensky, P., Goldrick, M., & Mathis, D. (in press). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*.

Smolensky, P., & Legendre, G. (2005). *The harmonic mind*. Cambridge, MA: MIT Press.

Stapp, H. P. (2009). The role of human beings in the quantum universe. *World Futures*, *65*(1), 7–18. doi: 10.1080/02604020802557318

Suppes, P. (1981). The plurality of science. In I. Hacking (Ed.), *PSA 1978*, *Vol. 2*, (pp. 3–16).

Tabor, W. (2002). The value of symbolic computation. *Ecological Psychology*, *14*(1-2), 21–51.

Tabor, W. (2009). A dynamical systems perspective on the relationship between symbolic and non-symbolic computation. *Cognitive Neurodynamics*, *3*(4), 415–427.

Tenenbaum, J. B., & Griffiths, T. L. (2002). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(04), 629–640.

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: are they really different? *Developmental Science*, *6*(4), 378–391.

van Fraassen, B. C. (2008). *Scientific representation: Paradoxes of perspective*. Oxford University Press.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21* (05), 615–628.

van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology. General*, *132*(2), 331–350, doi:10.1037/0096-3445.132.3.331

van Orden, G. C., Kello, C. T., & Holden, J. G. (2010). Situated behavior and the place of measurement in psychological theory. *Ecological Psychology*, *22*(1), 24–43.

Weiskopf, D. A. (2009). The plurality of concepts. *Synthese*, *169*(1), 145–173.