

Cognitive Science (2014) 1–38

Copyright © 2014 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12142

Information Processing and Dynamics in Minimally Cognitive Agents

Randall D. Beer,^{a,b} Paul L. Williams^a

^a*Cognitive Science Program*

^b*School of Informatics and Computing, Indiana University*

Received 19 October 2012; received in revised form 5 December 2013; accepted 16 December 2013

Abstract

There has been considerable debate in the literature about the relative merits of information processing versus dynamical approaches to understanding cognitive processes. In this article, we explore the relationship between these two styles of explanation using a model agent evolved to solve a relational categorization task. Specifically, we separately analyze the operation of this agent using the mathematical tools of information theory and dynamical systems theory. Information-theoretic analysis reveals how task-relevant information flows through the system to be combined into a categorization decision. Dynamical analysis reveals the key geometrical and temporal interrelationships underlying the categorization decision. Finally, we propose a framework for directly relating these two different styles of explanation and discuss the possible implications of our analysis for some of the ongoing debates in cognitive science.

Keywords: Information theory; Dynamical systems theory; Relational categorization; Evolutionary algorithms; Computational model

1. Introduction

The history of cognitive science has been characterized by often spirited disagreements about the proper conceptual foundation for a science of cognition. We focus here on the ongoing debate between information processing and dynamical approaches to cognition. In its classical form, the information processing or computational¹ approach claims that cognitive processes involve the manipulation of quasilinguistic structures very much like a digital computer or Turing machine might manipulate strings of symbols (Fodor, 1975; Newell & Simon, 1976; Pylyshyn, 1984). In contrast, the best-known statement of the

Correspondence should be sent to Randall D. Beer, Cognitive Science Program, 1900 E. 10th St. – 840 Eigenmann, Indiana University, Bloomington, IN 47406. E-mail: rdbeer@indiana.edu

dynamical approach holds that cognitive processes are quantitative and rate-dependent much like a physical system such as the Watt governor (van Gelder, 1998; van Gelder & Port, 1995).

It would be very convenient if the alternatives were as clear-cut as these two statements imply. Unfortunately, the reality is much more complicated. Individual proponents and critics interpret the claims and assumptions of these two approaches in a variety of ways, often taking very different stands on a host of fundamental issues: What is information? What does it mean to process it? What is a computation? What is a dynamical system? Are cognitive processes discrete or continuous? Are they sequential or temporal? Are they local or distributed? Do they involve internal representations and, if so, what kind? Other related disputes also sometimes get pulled into the discussion, such as the extent to which the nature of brain organization and neural activity must be taken into account and the roles that feedback through the body and environment might play in cognitive processes. The many positions one can take on this tangle of interrelated issues make it difficult to resolve the disagreements between information processing and dynamical approaches by conceptual argumentation alone.

In this article, we take a somewhat different perspective on these debates. As Dale and Spivey (2005, p. 318) have argued, “A case may be made for mathematization of scientific domains as a course toward resolving theoretical disputes, clarifying conceptual confusions, and making potential decisions concerning the greater validity of one verbalized scientific description over another.” Thus, we focus our attention on the mathematical theories that underlie the information processing and dynamical approaches, namely information theory (IT) and dynamical systems theory (DST), respectively. As mathematical theories, IT and DST can be applied to any system that takes the proper form to meet their defining requirements; they intrinsically make no scientific claim as to “what’s really going on.” Instead, they are best viewed as distinct mathematical lenses through which we can examine the operation of a system of interest. Our concern here will be the different sorts of explanations that each lens reveals, and, perhaps most important, the relationship between these two explanations when both lenses are applied to the same cognitive system.

To explore these issues, we utilize the methodology of minimally cognitive behavior, in which evolutionary algorithms are used to evolve model agents that can exhibit some cognitively interesting behavior, and these evolved agents are then subjected to a mathematical analysis to understand the mechanisms underlying their behavior (Beer, 1996, 2003; Harvey, Di Paolo, Wood, Quinn, & Tuci, 2005). Specifically, we compare and contrast information processing and dynamical explanations of an evolved model agent engaged in a relational categorization task. Although such agents are not intended here as realistic models of human competencies, they serve as excellent vehicles for exploring theoretical landscapes and developing and testing new concepts and mathematical tools (Beer & Williams, 2009). A classical information processing perspective typically makes strong assumptions about the nature and architecture of the processes of feature extraction, memory, comparison, decision-making, and action that

might be presumed to underlie relational categorization. In contrast, a significant advantage of the evolutionary approach is that it helps to minimize the impact of our a priori preconceptions about how an agent that successfully performs a given task *ought* to work, thus making the analysis of how it actually *does* work an interesting and insightful exercise.

This article is organized as follows. In Section 2, we describe the relational categorization task that we utilize and the model agent that we evolved to solve it. Sections 3 and 4 focus separately on information-theoretic and dynamical analysis, respectively. Each of these sections contains a brief introduction to the relevant mathematical theory and then presents a detailed analysis of the same evolved relational categorization agent using the tools of that theory. Section 5 then examines the relationship between these informational and dynamical explanations and proposes a framework for relating them. Section 6 concludes the article with a discussion of the broader implications of our analysis for some of the ongoing debates in cognitive science.

2. A relational categorization agent

Interest in relations and relational categories has been flourishing within cognitive science (e.g., Gentner & Kurtz, 2005; Markman & Stilwell, 2001). Relational categories are categories determined by common relational structure among category members, rather than intrinsic similarities between category members. For example, same and smaller are instances of relational categories. There is a vast body of research on relational categorization in a wide range of species, including humans (Kurtz & Boukrina, 2004), pigeons (Wills, 1999), rats (Saldanha & Bitterman, 1951), and insects (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001). Moreover, relational categories are of particular importance for cognitive science because they are fundamental to many topics, such as analogy and language. Interestingly, categorization has also been proposed as an argument for why cognitive processes must employ discrete representations (Dietrich & Markman, 2003), a claim we will return to in the discussion.

We have previously described the evolution of model agents solving a relational categorization task involving the smaller-than relation (Williams, Beer, & Gasser, 2008). In this task, an agent with an array of distal sensors can move left or right as objects drop from above (Fig. 1A). We have studied two different versions of this task. In the spatial version, two objects of different size fall simultaneously from different horizontal positions and the agent must catch the smaller one. In the temporal version, the agent is exposed to the two objects in sequence. During the time that a cue object is falling, the agent is free to do whatever it likes. However, when a subsequent probe object falls, the agent must catch the probe if it is smaller than the earlier cue and avoid it otherwise. Because this temporal version of the task is somewhat richer (involving not only relational categorization but also memory), we will focus on it here.

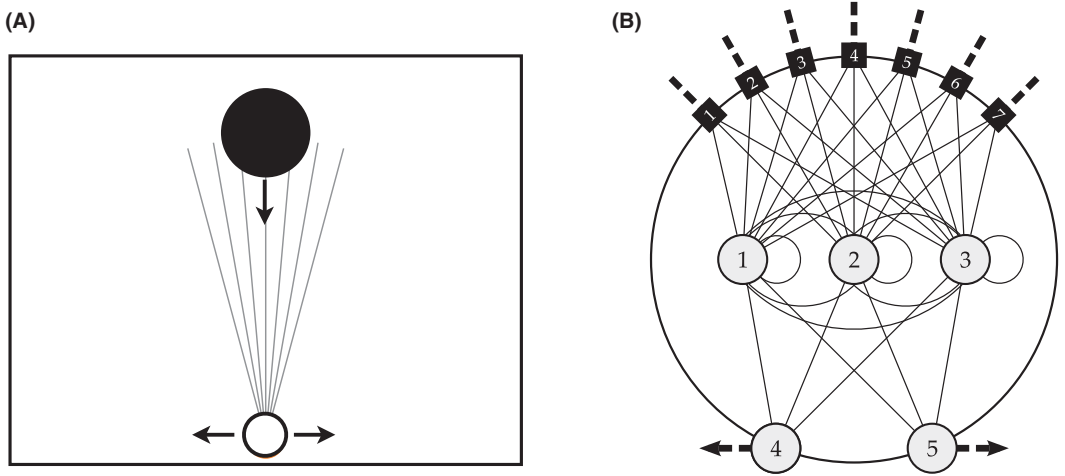


Fig. 1. A simple relational categorization model. (A) An agent with an array of distal sensors moves horizontally as circular objects fall toward it from above. (B) The agent's "nervous system." Seven ray sensors project to three fully interconnected interneurons, which in turn project to two motor neurons.

Our model is specified by the following set of differential equations (Fig. 1B):

$$\begin{aligned}
 \dot{y}_i &= -3 && \text{(Object Distance)} \\
 s_i &= f_i(x, y; z) && i = 1, \dots, 7 \quad \text{(Sensors: S1-S7)} \\
 \tau_i \dot{n}_i &= -n_i + \sum_{j=1}^7 s_j + \sum_{j=1}^3 w_{ji} o_j && i = 1, 2, 3 \quad \text{(Interneurons: N1-N3)} \\
 \tau_i \dot{n}_i &= -n_i + \sum_{j=1}^3 w_{ji} o_j && i = 4, 5 \quad \text{(Motor Neurons: ML, MR)} \\
 \dot{x} &= 5(o_4 - o_5) && \text{(Agent Position)}
 \end{aligned} \tag{1}$$

where y is the vertical distance between the object and the agent, s_i is the output of the i th sensor, n_i is state of the i th neuron, and x is the horizontal separation between the agent and the object. Circular objects, centered horizontally with a size (radius) z in the range [20,50], fall at a constant velocity of 3 from an initial vertical distance at which the object's leading point lies at the maximum length of the central ray. The $f_i(x, y; z)$ are functions that give the output of each sensor depending on the agent's horizontal position x , the object's vertical position y , and the size z of the object, with an output of 0 corresponding to a ray at its maximum length (i.e., no intersection with the object) and an output of 1 corresponding to a ray at its minimum length (i.e., object is touching the agent). The output $o_i = \sigma(n_i + \theta_i)$ of a neuron with internal state n_i is given by the standard logistic activation function $\sigma(x) = 1/(1 + e^{-x})$. The parameters τ_i , θ_i , w_{ji} specify the time constants, biases, and connection weights, respectively, of the continuous-time recurrent neural network representing the agent's "nervous system," giving a total of 46 free parameters that are set by an evolutionary algorithm. Fitness is evaluated by averaging the horizontal separation between the agent and the object at the end of the probe stage for a range of cue (z_{cue}) and probe (z_{probe}) size combinations, with this distance being

minimized for any trial in which $z_{\text{probe}} < z_{\text{cue}}$ and maximized for any trial in which $z_{\text{probe}} > z_{\text{cue}}$.

Over multiple runs of the evolutionary algorithm, we found that two distinct strategies reliably evolved. In the first strategy, which we call *passive*, the agent sits motionless during the cue stage and only starts to move once the probe stage has begun. In the second strategy, which we call *active*, the agent moves throughout both the cue and probe stages. Because our concern here is with the relationship between information processing and dynamical explanations rather than embodiment, we will mostly focus our analysis on the best passive agent that evolved. However, in Section 3.4, we will briefly analyze an active agent and discuss the possible implications for ongoing debates on the role of the body and environment in cognition.

3. Information-theoretic analysis

3.1. The dynamics of information

Underlying any discussion of information processing must be a theory of information: What is it? How do we measure it? By what rules can it be transformed? The best such theory that we currently have available is Claude Shannon's (Shannon, 1948). Although Shannon's mathematical theory of information was originally developed to address the engineering problem of reliable transmission of messages over noisy communication channels, information theory has grown into a ubiquitous general tool for the analysis of complex systems (Cover & Thomas, 2006). The two central concepts in IT are entropy and mutual information. The entropy $H(X)$ is a measure of our uncertainty about the outcome of a measurement on a random variable X . The mutual information $I(X; Y)$ is a measure of the dependence between two random variables; it quantifies the amount by which a measurement on one of the variables reduces our uncertainty about the other.

In addition to these traditional concepts, we will make use of four extensions in our analysis to track the flow of information throughout the brain-body-environment (Williams & Beer, 2010b). First, Shannon's information measures average across all measurement outcomes. The idea behind *specific information* measures $I(X = x_i; Y)$ is to unroll such averages to quantify the information that one random variable provides about each specific outcome of another, providing a more fine-grained analysis of informational relationships (Butts, 2003; DeWeese & Meister, 1999). Second, Shannon's information measures are inherently atemporal; they do not take into account the fact that the information in a random variable may change over time. In order to address the temporal dimension of information, we replace each random variable X with a random process X_t (a time-indexed sequence of random variables; Lasota & Mackey, 1994) and consider separately the information in X for each possible value of t . Third, it is often necessary to consider not only the information in individual variables but also the information encoded by multiple variables. For example, our new measure $I_{\text{syn}}(X; S_1, S_2)$ gives the *synergistic information* about X provided by the simultaneous knowledge of S_1 and S_2 that is not

available from either source alone. Fourth, we use new measures of *dynamic information* to track how information is gained by individual variables $I_G(X; Y_t)$ and transferred between variables $I_T(X; Y_{t-1} \rightarrow Z_t)$ over time. The latter two extensions are based on a novel multivariate generalization of mutual information recently developed by the authors (Williams & Beer, 2010a). The definitions of all information measures used in this paper are given in the Appendix.

As we will be concerned here with how object size information flows throughout our relational categorization agent over time, the first argument to all of our information measures will always be a random variable related to object size, while the remaining arguments will refer to various components of the brain-body-environment system. We therefore normalize all informational quantities by the entropy of the corresponding object size variable to obtain measures that run from 0 to 1. In addition, we consider not just a single instance of our system but an ensemble of such instances obtained by varying the object size. That is, we consider the cue and probe sizes to be uniformly distributed over their allowable ranges, inducing probability distributions on the remaining model components. Thus, each component of our model (z , s_i , n_i , o_i , and x) will be associated with a corresponding random variable (Z , S_i , N_i , O_i , and X). Note that we focus on neuron outputs (O_i) rather than states in this section, both because the boundedness of the outputs makes it easier to compute histograms and because it is only the output of one neuron that can directly affect another.

3.2. Cue stage information processing

An information-theoretic analysis of the cue stage must answer three questions. Where can information about the size of the cue object be found in the brain-body-environment system during this stage? Where is this size information stored at the end of the cue stage for later comparison with the size of the probe object? How does cue size information find its way into this storage location? We address each of these questions in turn.

In order to identify where information about the cue size can be found, we calculate the mutual information $I(Z_{\text{cue}}; E(t))$ between the cue size and each element E of the system as a function of time (Fig. 2). We find that no information about cue size ever appears in the motor neurons or the agent position during the cue stage, which is consistent with the fact that a passive agent does not move during this stage. We also find no cue size information in the central sensor S4 (as the object falls straight down along the midline, this sensor provides information about distance but not size). In contrast, we do find significant size information in the remaining sensors (Fig. 2A). Note that, due to the left/right symmetry of the agent, both sensors in the pair S1/S7 contain the same amount of information about the size of an object along the midline, as do the pairs S2/S6 and S3/S5. Note also that, as the object falls, information becomes available first in the innermost pair S3/S5, followed by S2/S6, and finally the outermost pair S1/S7. We also find significant size information in all three interneurons at various times (Fig. 2B).

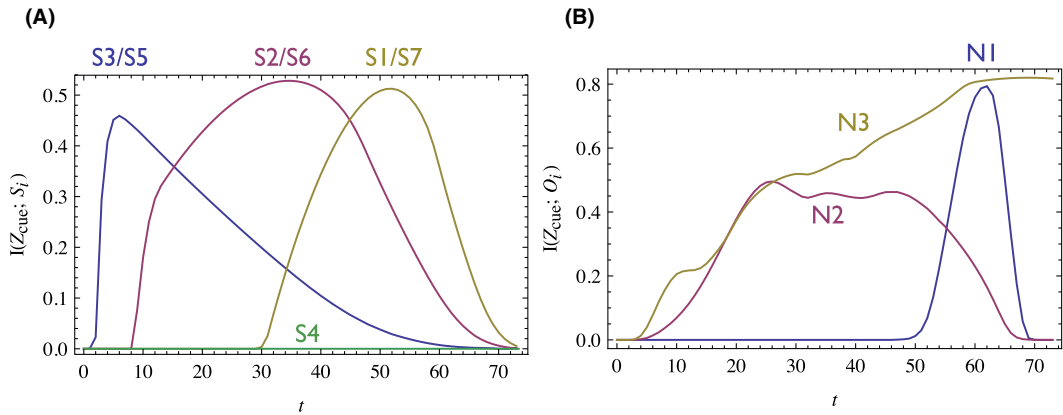
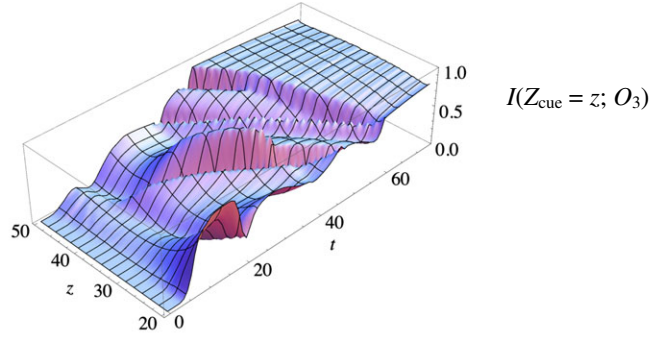


Fig. 2. The dynamics of cue size information during the cue stage. (A) Cue size dynamics in the sensors. Note that the cue size information drops to 0 at the end of the cue stage when the cue is removed. (B) Cue size dynamics in the interneurons. Note that cue size information is maintained only in N3 after the cue is removed.

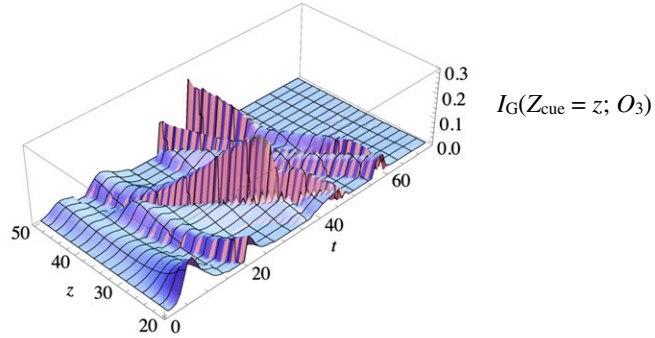
Next, to determine where cue size information is stored for the probe stage, we need only examine where it resides at the end of the cue stage. From Fig. 2A, we see that all sensory information is lost at this point. This is because, as the sensors maintain no state, any size information they contain disappears as soon as the object does. Thus, any sensory information must be transferred to the interneurons before this point. However, from Fig. 2B we see that, despite containing significant size information earlier in the trial, interneurons N1 and N2 also lose this information shortly before the cue stage ends. Therefore, as only N3 retains information beyond this point, we must conclude that it is the location in which the cue size is stored in this agent.

Finally, we must characterize how information about cue size comes to reside in N3. This is best done using a more detailed picture of the cue size information in N3, which can be obtained by specific information $I(Z_{\text{cue}} = z_i; O_3(t))$ to unroll this information across individual size values (Fig. 3A). Calculating the information gain $I_G(Z_{\text{cue}} = z_i; O_3(t))$ reveals that N3 gains cue size information in several distinct waves whose timing and amplitude differ for different cue sizes, with information about larger cue sizes often gained before information about smaller cue sizes (Fig. 3B). Now we are ready to answer our question: Where does this information come from? The only sources of information that N3 has direct access to are the sensors and the other interneurons. Calculating the information transfer $I_T(Z_{\text{cue}} = z_i; E(t-1) \rightarrow O_3(t))$ from each of these elements E to N3, we discover that cue size information from S3/S5 is primarily responsible for the first two waves of information gain in N3, information from S2/S6 is primarily responsible for the third wave, and S1/S7 is primarily responsible for the fourth (Fig. 3C). Although we also find that a small amount of cue size information is transferred from N2 into N3 (results not shown), the majority of cue size information in N3 comes from the sensors.

(A)



(B)



(C)

$$I_T(Z_{\text{cue}} = z; S_{3/5} \rightarrow O_3)$$

$$I_T(Z_{\text{cue}} = z; S_{2/6} \rightarrow O_3)$$

$$I_T(Z_{\text{cue}} = z; S_{1/7} \rightarrow O_3)$$

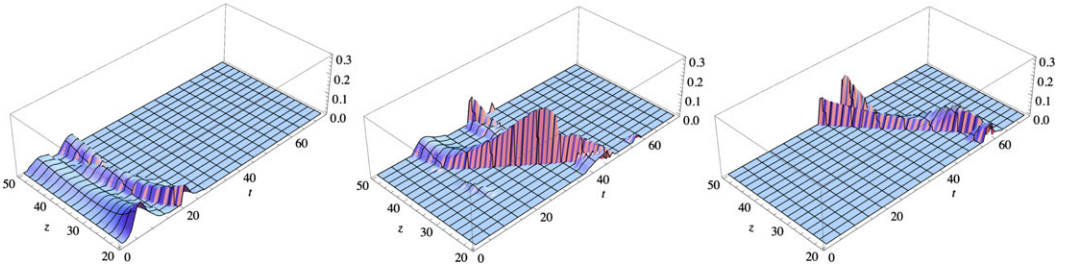


Fig. 3. Specific cue size information flow in N3. (A) Unrolling cue size information in N3 (N3 curve in Fig. 2B) over individual cue size values produces a more detailed visualization of the dynamics of cue size information in N3. (B) A visualization of the gain in cue size information in N3 over time. (C) A visualization of the cue size information transferred from each pair of sensors to N3 over time. Note that distinct transfer events from each pair of sensors underlie the ridges in (B).

3.3. Probe stage information processing

During the probe stage, we could examine where information about the sizes of both the cue and probe objects resides. However, what really matters during this stage is not the absolute sizes of the two objects, but their *relative* size, as that is the basis on which

the categorical decision must be made. In order to explore this information, we define the relative object size r (with corresponding random variable R), with $r = 1$ if $z_{\text{probe}} < z_{\text{cue}}$ and $r = 0$ otherwise. Since the agent's decision is ultimately expressed behaviorally as a catch or avoid response, we can then calculate the mutual information $I(R; X(t))$ between relative object size and the agent's position over time (Fig. 4C). From this plot, it is clear that relative size information begins to be reflected in the agent's behavior at $t \approx 94$ (arrow).

Where does the relative object size information in X come from? The only direct source of information for X is the motor neurons. Calculating the mutual information between R and each of the motor neurons clearly shows that relative size information in MR begins to increase at $t \approx 87$ (arrow), prior to the rise in relative size information in X , suggesting that MR is the source of this rise (Fig. 4B). A more detailed information transfer analysis (not shown) confirms this. Note that saying that $I(R; MR)$ rises is not the same as saying that the MR neuron turns on at this point. In fact, MR, which moves the agent to the left, turns *off* at this point, removing its opposition to ML (which is also on at this point) and resulting in a movement to the right. Rather, an increase in $I(R; MR)$ means that a measurement of the value of MR becomes more predictive of R ; that is, it more substantially reduces our uncertainty about R .

Backtracking one step further, we see that there are two potential sources of relative size information for MR among the interneurons: N1 and N3 (Fig. 4A). Simply looking at this plot, we might guess that N1 is the primary source of relative size information in MR because N1's information begins to rise at $t \approx 78$ (arrow), prior to the rise in MR. An analysis of information transfer demonstrates that this is indeed the case, and that the R information in N3 is not transferred to the motor neurons or behavior. However, we do find that *absolute* cue size information is transferred from N3 (where it was stored at the end of the cue stage) to N1 during this time (results not shown), explaining how it is possible for N1 to gain information about *relative* object size. Thus, relative size information builds up in N1, then flows to MR and subsequently flows to X to be reflected in the agent's behavior.

3.4. Active agents

So far, we have focused on how information flows internally through the neurons in a passive agent, which remains motionless until after the decision to catch or avoid the probe object has been made. However, the same approach can be applied to the entire brain-body-environment system of an active agent, which moves throughout both the cue and probe stages. We merely need to include the relevant body and environment states in our analysis. In this section, we will briefly examine two key ways in which the information flow in an active agent differs from that in a passive one.

As for a passive agent, we can examine the cue size information available in the interneurons of an active agent over time, $I(Z_{\text{cue}}; O_i(t))$. Once again, we find that all three interneurons carry cue size information at some point during the cue stage (Fig. 5A). However, unlike in the passive agent, all of the active agent's interneurons lose their cue

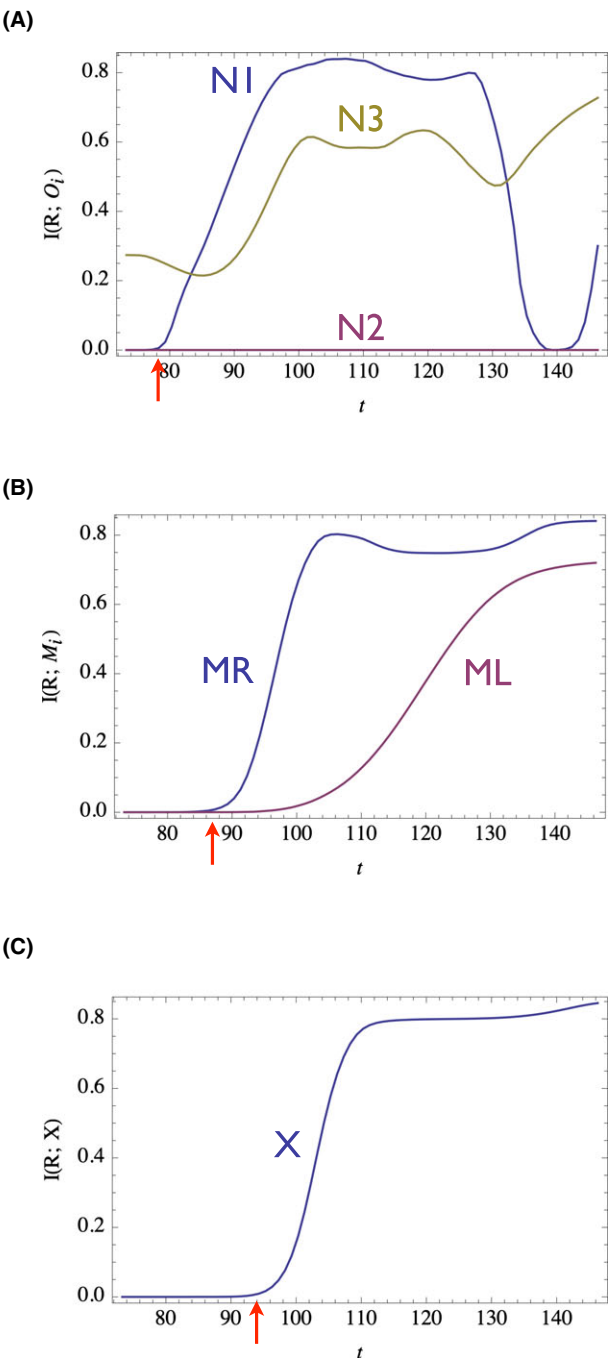


Fig. 4. The flow of relative size information from N1 (A) to MR (B) and finally to X (C) during the probe stage.

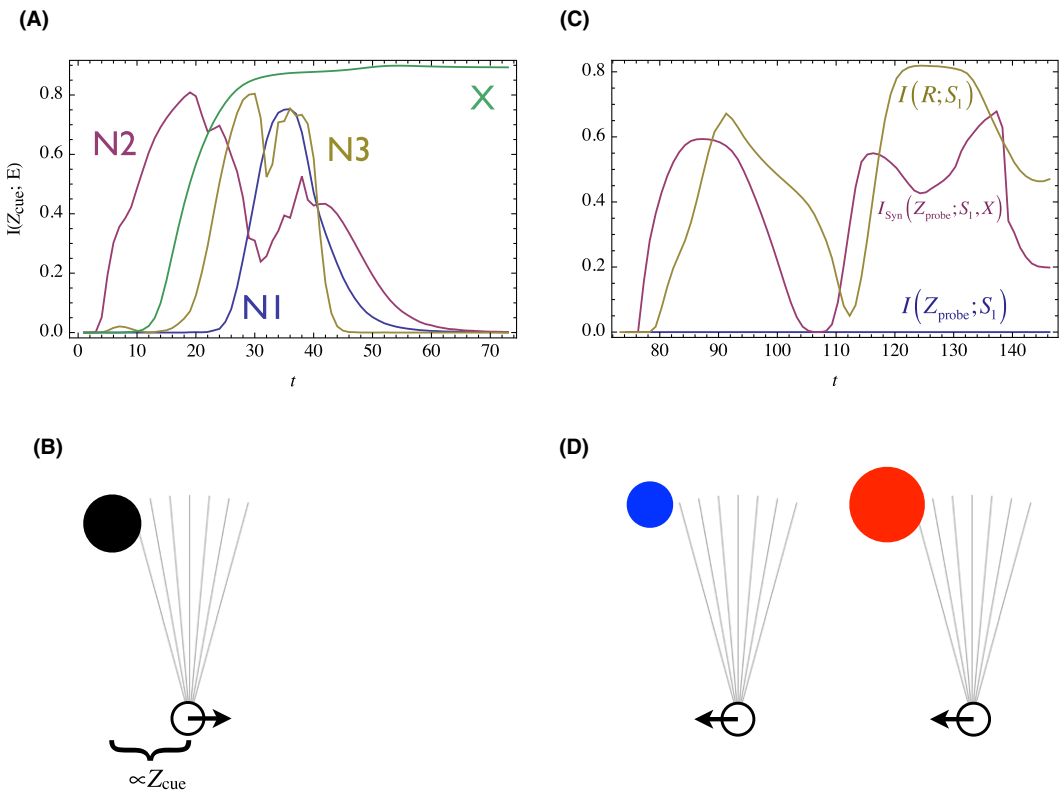


Fig. 5. The flow of size information in an active agent. (A) The flow of cue size information during the cue stage. Note that, although no interneurons retain cue size information by the end of the cue stage, the agent position does. (B) By the end of the cue stage, the agent has moved to the right by a distance proportional to the cue size. (C) The flow of size information during the probe stage. (D) During the probe stage, the agent moves to the left, encountering different probe objects at different times depending upon their size relative to the cue object.

size information before the probe stage begins (compare to Fig. 3B). This raises an obvious problem: Where does the active agent store cue size information for later comparison with the probe object size if not in its nervous system? Examining the remaining brain-body-environment variables, we find that only $I(Z_{\text{cue}}; X(t))$ is nonzero at the end of the cue stage (Fig. 5A). In other words, cue size is stored in the position of the agent's body. Behaviorally, this means that, by the end of the cue stage, the active agent has moved to the right a distance proportional to the size of the cue object (Fig. 5B). This is a simple example of what is sometimes called *information offloading*, in which some aspect of a cognitive process (memory in this case) is offloaded to the environment to simplify processing demands (Hutchins, 1995; Kirsh & Maglio, 1994).

There is also an interesting difference between the passive and active agent during the probe stage. If we examine $I(Z_{\text{probe}}; S_1(t))$, we find, somewhat surprisingly, that there is no information about the probe size available in the individual sensors (Fig. 5C). This is

directly related to the variability in agent position introduced by information offloading. However, probe size information is available synergistically in the combination of sensor output and body position $I_{\text{Syn}}(Z_{\text{probe}}; S_1(t), X(t))$. Even more interesting, significant relative size information is available in this sensor almost immediately after the probe object is encountered. As the agent moves left during the probe stage, larger probe objects are encountered before smaller ones, with the relative scale for “large” and “small” set by the distance the agent moved to the right during the cue stage (Fig. 5D). In essence, the agent structures the body/environment relationship in such a way that it can directly perceive the relative size of the two objects, offloading not only the memory of the first object but the categorization decision itself. Of course, neural activity still plays a central role in both setting up this relationship during the cue stage and in capitalizing on it during the probe stage. Nevertheless, this is a simple example of what has been called *information self-structuring*, in which an agent’s own movements introduce information into its environment that it can subsequently use to guide behavior (Lungarella & Sporns, 2006).

3.5. Summary

In this section, we have demonstrated how a set of novel extensions to IT can be used to track the fined-grained flow of information over time through the brain-body-environment system of a model agent performing a simple relational categorization task. In the passive agent, we showed how waves of cue size information are accumulated and stored in N3 during the cue stage. This information is combined with probe size information to extract relative size information in N1 during the probe stage, which then flows through the right motor neuron to become expressed in the agent’s behavior as a categorization decision. In contrast, the active agent stores cue size information in its body position, arranging to directly extract relative size information from its perception of the probe object. Of course, the details of information flow vary from one evolved agent to the next. For example, a passive agent could store cue size information in any combination of its interneurons and an active agent could store it in any combination of its neurons and body position. Nevertheless, these methods can be used to reveal the information flow architecture of any such agent. Indeed, they can be applied to any system for which the probability distributions of the variables of interest can be estimated.

4. Dynamical analysis

4.1. Dynamical systems theory

Dynamical systems theory (DST) is concerned with characterizing the behavior of systems over time. It grew out of work by Henri Poincaré at the end of the 19th century on the stability of the solar system (Diacu & Holmes, 1996). However, the mathematical concept of a dynamical system was generalized subsequently to include a much wider

class of systems. Indeed, mathematically, a dynamical system is nothing more than a triple consisting of a state space S , a time space T , and an evolution operator ϕ_t (Kuznetsov, 2004). A dynamical system whose dynamical law is fixed is called *autonomous*, whereas one whose dynamical law changes over time is called *nonautonomous*. If ϕ_t also depends on a set of parameters drawn from some parameter space P , we call the resulting quadruple a *family* of dynamical systems. For any system that can be put into the proper form, DST provides a powerful set of concepts and tools for understanding its temporal behavior, including *limit sets* (e.g., equilibrium points, limit cycles, and chaos), *stability* (attractors, repellers, and saddles), *basins of attraction* (the set of points surrounding an attractor that approach it over time), *phase portraits* (a global mathematical picture of all the limit sets, their stabilities, and the ways in which they fit together), and *bifurcations* (changes in the number, type, stability, or interconnection of limit sets as parameters are varied) (Strogatz, 1994).

There are many different ways in which DST can be utilized within cognitive science (e.g., Ashby, 1960; Hotton & Yoshimi, 2011; Kelso, 1995; Schöner, 2008; Spivey, 2007; Warren, 2006). Our approach to dynamical analysis begins by considering an agent and its environment as two coupled dynamical systems: $\dot{\mathbf{x}}_A = \mathcal{A}(\mathbf{x}_A; \mathbf{S}(\mathbf{x}_e))$, $\dot{\mathbf{x}}_e = \mathcal{E}(\mathbf{x}_e; \mathbf{M}(\mathbf{x}_A))$, where the agent state variables $\dot{\mathbf{x}}_A$ and the environment state variables $\dot{\mathbf{x}}_e$ are coupled by sensory functions \mathbf{S} and motor functions \mathbf{M} (Beer, 1995a, 2003). There are two different perspectives we can take on this coupled agent-environment system. On the one hand, we can view it as a single dynamical system $\dot{\mathbf{x}}_U = \mathcal{U}(\mathbf{x}_U)$ and study its autonomous dynamics. On the other hand, we can view it as two separate dynamical systems $\dot{\mathbf{x}}_A = \mathcal{A}(\mathbf{x}_A; \mathbf{S}(t))$ and $\dot{\mathbf{x}}_e = \mathcal{E}(\mathbf{x}_e; \mathbf{M}(t))$ and study their nonautonomous dynamics as their inputs vary over time. We use a quasistatic approach to relate these two perspectives by which we explain the dynamics of the nonautonomous system $\dot{\mathbf{x}}_A = \mathcal{A}(\mathbf{x}_A; \mathbf{S}(t))$ using the bifurcation diagram of the autonomous system $\dot{\mathbf{x}}_A = \mathcal{A}(\mathbf{x}_A; \mathbf{S})$, where the time-varying sensory inputs $\mathbf{S}(t)$ have been replaced by parameters \mathbf{S} . Of course, this decomposition can be iterated, with the agent dynamics being further decomposed into interacting body and nervous system dynamics, and the nervous system dynamics in turn being decomposed into interacting neuronal elements.

Our relational categorization model defines a 2-parameter (z_{cue} and z_{probe}) family of 7-dimensional (n_1, \dots, n_5, x, y) autonomous dynamical systems, with agent states $\mathbf{x}_A = (n_1, \dots, n_5)$, body/environment states $\mathbf{x}_e = (x, y)$, sensory functions $\mathbf{S}(x, y; z) = (f_1(x, y; z), \dots, f_7(x, y; z))$ and motor function $\mathbf{M}(n_4, n_5) = (5(\sigma(n_4 + \theta_4) - \sigma(n_5 + \theta_5)))$. This family of dynamical systems has several special properties. First, as y decreases without bound, the dynamics is divergent (although it is truncated when the falling object reaches the agent). In fact, as y decreases at a constant rate, it is equivalent to time, implying that we can also interpret this model as a 2-parameter family of 6-dimensional nonautonomous dynamical systems. Second, because the sensor ray intersections can change discontinuously as an object moves through the agent's field of view, the dynamics is only piecewise-smooth in x and y , making it a hybrid dynamical system (di Bernardo, Budd, Champneys, & Kowalczyk, 2008). Finally, the dynamics has a compositional structure in which only one of the two parameters is relevant at any given time. Thus, we can first

study the 1-parameter system $\phi_t(\mathbf{0}; z_{\text{cue}})$ from the initial condition $\mathbf{0}$ for $0 \leq t \leq T$ and then study the 1-parameter system $\phi_t(\mathcal{C}_{\text{cue}}; z_{\text{probe}})$ from the set of initial conditions $\mathcal{C}_{\text{cue}} = \phi_T(\mathbf{0}; z_{\text{cue}})$ reached at the end of the cue stage for $T < t \leq 2T$.

4.2. Cue stage dynamics

A dynamical analysis of the cue stage begins with a characterization of the trajectory structure of the entire brain–body–environment system as the cue object size is varied. Thus, as in our informational analysis, we consider not just a single trajectory of the system, but rather an ensemble of trajectories obtained by varying the object size parameter. Since a passive agent does not move during this initial stage, we can ignore the x state variable. We can also ignore the motor neurons at this point, as their only effect on the rest of the system is through x . This reduces the seven differential equations of the system to four and, as y is equivalent to time (because it changes at a constant rate), we can visualize the cue stage dynamics as nonautonomous trajectories in the three-dimensional interneuronal state space (Fig. 6A). The advantage of working in state space rather than

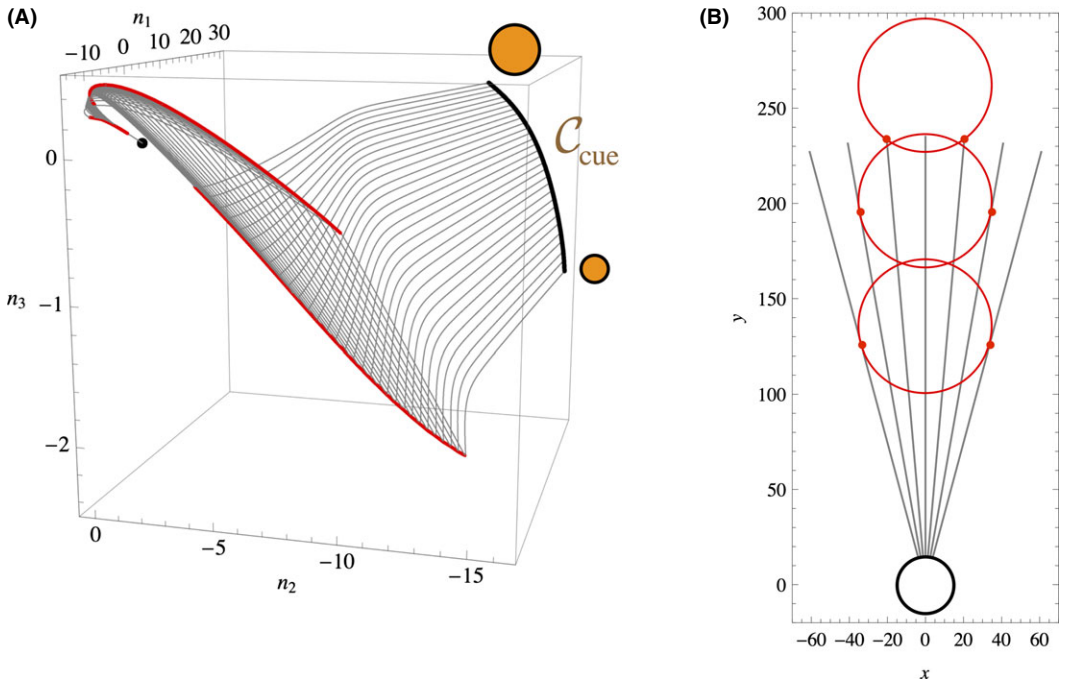


Fig. 6. Cue stage dynamics. (A) Cue trials always begin in the same initial state (black point). As cue objects fall, trajectories spread out according to size until, at the end of the cue stage, they are ordered from smallest to largest along the curve \mathcal{C}_{cue} . Note the three curves along which the trajectory directions change discontinuously (outlined in red). (B) As an object falls toward a stationary agent, it intersects each bilateral pair of rays in sequence from inner to outer (red dots). These points of discontinuity coincide with the discontinuity curves in part A. The times and locations at which these intersections occur depend on the size of the object; only the $z_{\text{cue}} = 35$ case is shown.

output space is that it is usually much easier to understand the structure of trajectories there. Due to the saturation effects of the sigmoid function, sets of trajectories can become very compressed in output space. However, it is important to remember that only the output of a neuron can directly affect other neurons.

Because the interneurons are always initialized to 0, all trajectories begin at the same state (black point in Fig. 6A). However, over time they evolve along different paths depending on the size of the cue object. By the end of the cue stage, these differences result in the states being spread along a curve \mathcal{C}_{cue} in state space in such a way that they are ordered by object size (thick black curve), with larger objects appearing toward the top of the curve and smaller objects appearing toward the bottom. Thus, the state of the interneurons at the end of the cue stage depends in a systematic way on the size of the cue object.

A key feature of these trajectories is that their time-evolution is not smooth; it changes sharply at three specific times (red curves). Indeed, it is clear from Fig. 6A that these sudden changes play a central role in organizing the layout of the final states along \mathcal{C}_{cue} . What underlies these sharp changes? As an object falls along the midline during the cue stage, it intersects different combinations of sensor rays at different points in time (Fig. 6B). Due to the left/right symmetry, these intersections occur in pairs, beginning with the innermost pair S3/S5, then S2/S6 and finally S1/S7. Every time the object goes from not intersecting a ray to intersecting it, the input from that sensor jumps from zero to some nonzero value. Such jumps introduce discontinuities into the dynamics, as the attractor toward which a trajectory is moving after the jump can be completely different from the attractor toward which it was moving before.

In general, as an object moves through the agent's field of view, the number of ray intersections change at any tangency between the circular object and a ray, producing a network of discontinuity boundaries that divide the state space of the system into cells of smooth dynamics. Using the geometry of circle/line intersection, we can solve for the locations of the three midline discontinuity points mentioned above, and hence the time at which they are encountered during the cue stage, as a function of object size. When we plot the states of the interneurons at these times as a function of cue size, we obtain the red curves shown in Fig. 6A. Thus, the sharp changes in trajectory shape that underlie the ordering of states by cue size are due to the differing times at which each trajectory encounters the three discontinuity points along the midline.

4.3. Probe stage dynamics

The probe stage begins with the interneuron states spread along \mathcal{C}_{cue} (Fig. 7A). Depending on the cue size, probe trajectories begin at some point along this curve and evolve over time into either a catch response (blue) if $z_{\text{probe}} < z_{\text{cue}}$ or an avoid response (red) if $z_{\text{probe}} > z_{\text{cue}}$. Our task in this section is to understand the dynamics underlying the separation of probe stage trajectories into these catch and avoid bundles. Since the two bundles are well separated by $t = 110$, we will focus our attention on this initial time; beyond this point the dynamics reflects merely the playing out of this decision in the agent's movement.

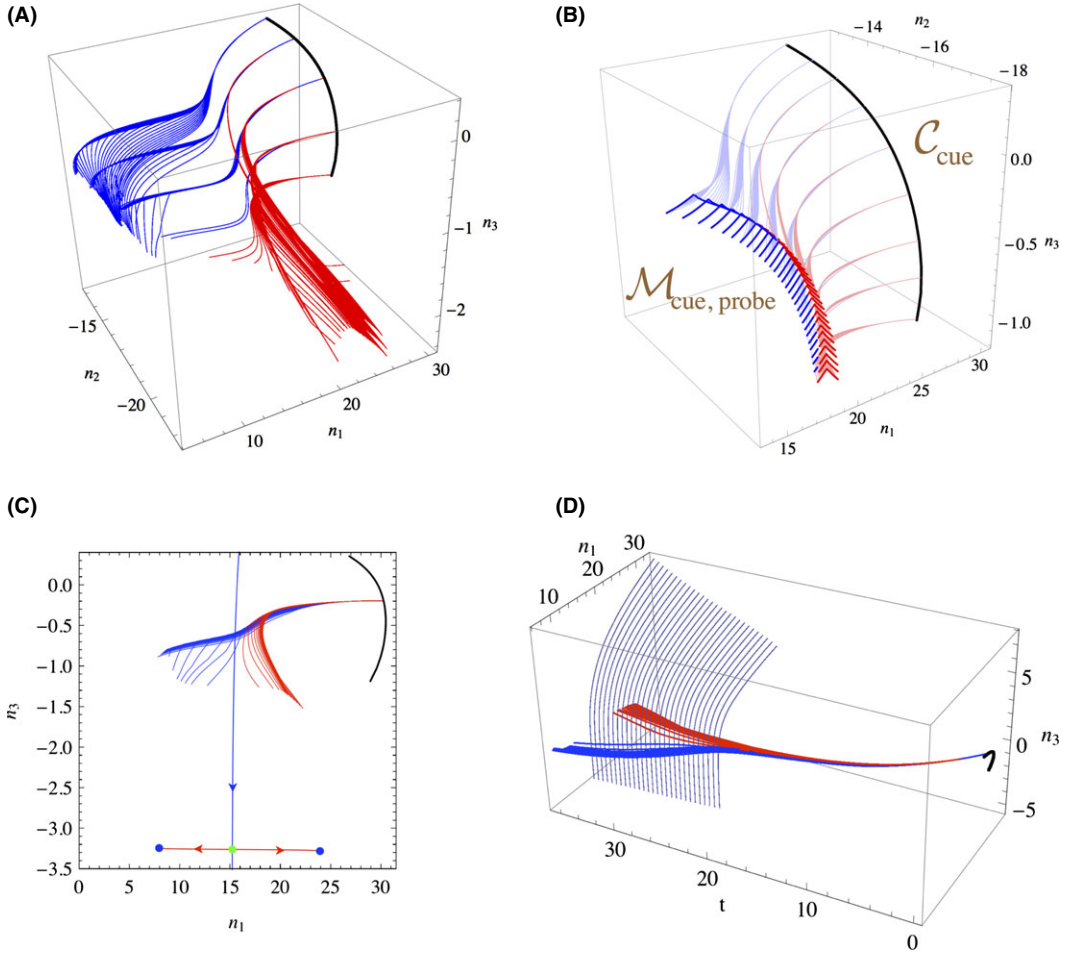


Fig. 7. Probe stage dynamics. (A) Probe trajectories begin at different points along C_{cue} depending on the previous cue size, spread out according to probe size, and then eventually separate into catch (blue) and avoid (red) bundles depending on whether the probe size is smaller or larger than the cue, respectively. (B) During the earliest part of the probe stage, states along the curve C_{cue} spread into a 2D surface $M_{cue,probe}$. Here, the surface is colored according to the eventual fate of the points on it, even though no splitting has yet begun to occur. (C) Splitting occurs when the neural dynamics bifurcates to bistability, with one branch of the stable manifold (blue curve with arrow) of the new saddle point (green) serving as a basin boundary that divides the trajectories into catch and avoid bundles. (D) Another view illustrating how the situation shown in part B develops over time. The sequence of dark blue lines indicates the changing location of the basin boundary. Note that the plots in parts C and D only show the splitting for the initial cue size $z_{cue} = 35$. In fact, a similar splitting effect occurs along the entire length of $M_{cue,probe}$.

During the earliest portion of the probe stage, the trajectories corresponding to different probe sizes from a given initial state along C_{cue} begin to spread out, becoming organized by probe size into wedge-shaped patterns (Fig. 7B). Thus, the initial curve C_{cue} develops into a 2-dimensional surface $M_{cue,probe}$ depending on both object sizes. Here,

points on the surface that will eventually lead to a catch response are colored blue and those eventually leading to an avoid response are colored red.

As this surface develops over time, what causes it to eventually split into distinct catch and avoid subsets along the red/blue seam? Let us focus on the set of trajectories for a single cue size. Because this is a nonautonomous system, its phase portrait depends not only on the parameter z_{probe} but also time. During the initial portion of the probe stage, the underlying autonomous dynamics of the interneurons exhibits only a single stable equilibrium point (whose position and movement over time is different depending on z_{probe} , which is why the trajectories spread out over time). However, as the probe continues to fall, the changing sensory inputs eventually cause the autonomous dynamics of the quasistatic approximation to become bistable, either through a smooth bifurcation or by crossing a discontinuity boundary (Fig. 7C). At this point, there are two stable equilibrium points (blue) and a saddle point (green) whose stable manifold (blue curve with arrow) serves to separate the basins of attraction of the two stable points. Those trajectories that find themselves to the right of the basin boundary will be attracted to the right stable point, whereas those that find themselves to the left of this boundary will be attracted to the left stable point. Thus, an interneuronal bifurcation to bistability sharply splits the surface $\mathcal{M}_{\text{cue,probe}}$ into separate bundles that eventually produce a catch or avoid response.

Although Fig. 7C presents a very static picture of this effect, in actuality both the trajectories and the basin boundary move to the left as time passes. Accordingly, on which side of the boundary a given trajectory ends up will be determined by a “race” between these two movements. Trajectories that win this race will eventually cross the boundary and be attracted to the left stable point, whereas trajectories that lose this race will remain to the right of the boundary and eventually bend back toward the right stable point. This is most clearly seen in a 3-dimensional plot that makes time explicit (Fig. 7D). Thus, we can see how the early probe stage dynamics transforms an initial variation in state along the N3 dimension into variation along the N1 dimension. This initially small N1 variation is then amplified to the scale where it can produce very different actions via splitting by the basin boundary created in a bifurcation.

4.4. Summary

In this section, we have demonstrated how DST can be used to characterize the neuronal dynamics of relational categorization in a passive agent. During the cue stage, differences in the timing of sensor discontinuity crossings spreads the states of the interneurons along a curve according to cue size. During the probe stage, this curve is further spread into a surface organized by both cue and probe size, and then a bifurcation to bistability in the underlying interneuronal dynamics splits this surface into two subsets that result in catch and avoid responses. Of course, the dynamical details will differ from one evolved agent to the next. However, the importance of transient dynamics that our analysis has revealed is quite general and the DST tools we have employed can be applied to a remarkable variety of systems.

5. The relationship between information and dynamics

5.1. The information in dynamics

In the previous sections, we have presented two completely independent explanations of an evolved agent, one using the tools of information theory and the other using the tools of dynamical systems theory. However, the careful reader will have no doubt noticed that there are many potential points of contact between these two analyses. The goal of this section is to make these connections explicit. First, we sketch a framework for linking the terms of an informational explanation to those of a dynamical one. If the focus of Section 3 was on *the dynamics of information*, our emphasis here will be on *the information in dynamics*. Specifically, we will show how to relate the changing geometry of the transient manifolds of a dynamical system to the informational quantities calculated from measurements of subsets of that system's components. We will then utilize this framework to examine how the informational and dynamical explanations of our relational categorization agent fit together.

As a motivating example, consider the Z_{cue} information in N1 during the cue stage (Fig. 2B). In particular, let us focus on the point in time at which $I(Z_{\text{cue}}; O_1(t))$ is maximal, which occurs at $t \approx 61.9$ (dark blue curve in Fig. 8A). As a result of the different trajectories that the dynamics follows during the cue stage for different values of z_{cue} (Fig. 6A), N1 takes on a range of outputs at this time (Fig. 8B). The relationship between z_{cue} and o_1 at $t \approx 61.9$ is made explicit by the curve in Fig. 8C. This functional relationship induces the joint probability distribution $p(Z_{\text{cue}}, O_1(61.9))$ (shading in Fig. 8C) used to compute the $I(Z_{\text{cue}}; O_1(61.9))$ value marked in Fig. 8A. Similar plots can be drawn at any point in time. Thus, Fig. 8C allows us to understand the connection between a dynamical concept (the projection onto the o_1 axis of how the manifold of interneuronal states at a given point in time depends on the parameter z_{cue}) and an information-theoretic concept (the mutual information between Z_{cue} and O_1).

To quantitatively characterize this connection, note that the mutual information $I(Z_{\text{cue}}; O_1)$ can be written as $H(O_1) - H(O_1|Z_{\text{cue}})$ (Cover & Thomas, 2006). The first term, $H(O_1)$ measures the uncertainty in O_1 , that is, the extent to which the outputs of N1 are spread uniformly across the possible values (dark magenta curve in Fig. 8A). Thus, $H(O_1)$ is high at $t \approx 61.9$ because the O_1 values are spread across many of the available bins with roughly equal probability (except for a sharp peak near $o_1 \approx 0.8$ and a much smaller peak near $o_1 \approx 0.1$ due to the flattening of the curve at these values; see Fig. 8C). The second term, $H(O_1|Z_{\text{cue}})$, measures the uncertainty in the dependence of O_1 on Z_{cue} , that is, the extent to which the mapping from cue sizes to N1 outputs fails to be a function (dark yellow curve in Fig. 8A). As this mapping really is a function, the only possible source of uncertainty here is discretization. Thus, as we increase the resolution of our discretization, $H(O_1|Z_{\text{cue}})$ will approach 0 and $I(Z_{\text{cue}}; O_1)$ will approach $H(O_1)$.

There is much more that can be said about this idea of $I(Z_{\text{cue}}; O_1)$ measuring the uniformity of spread in o_1 as z_{cue} is varied, including its relationship with the degeneracy of the

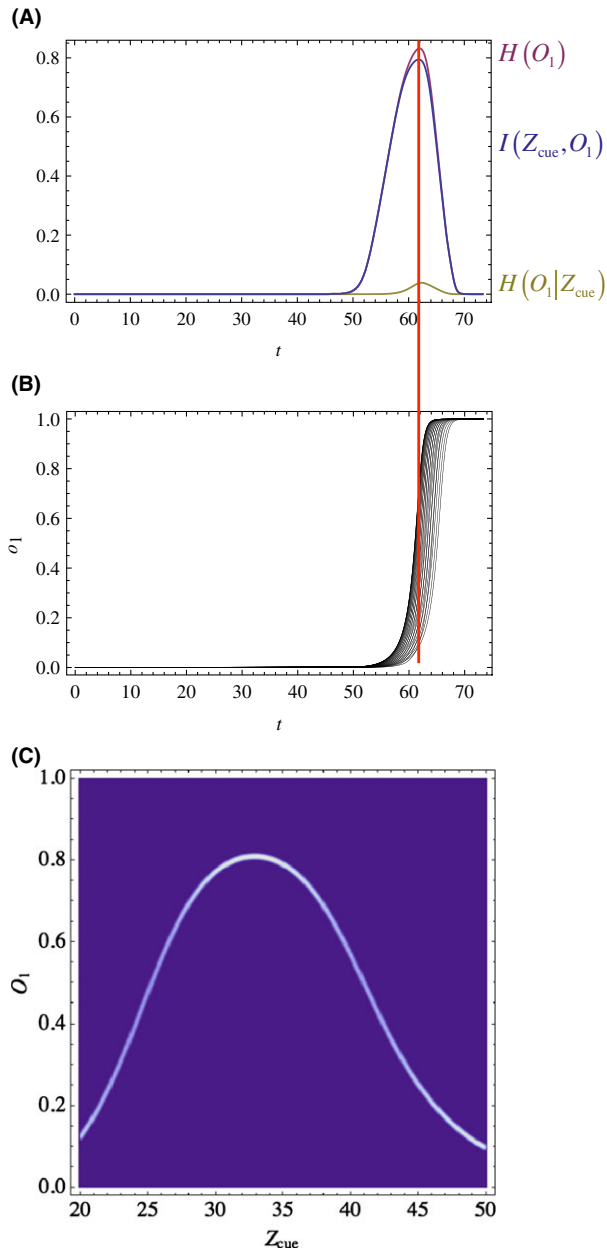


Fig. 8. The relationship between information and dynamics. (A) The cue size information in N1 during the cue stage can be decomposed into the difference between the entropy of N1 and the conditional entropy of N1 given the cue size. (B) Cue size information in N1 peaks at the point at which the outputs of N1 exhibit the highest uniformity of spread with respect to variations in the cue size. (C) The functional relationship between N1 output and cue size at the time of peak information induces a joint distribution between them (brighter areas represent higher densities).

inverse mapping from o_1 to z_{cue} and its generalization to multiple information sources. However, the basic idea is sufficient for our purposes here. For example, note that there is no vertical spread in o_1 with variation in z_{cue} for $t < 50$ and $t > 70$ due to saturation in N1 (Fig. 8B). This explains why $I(Z_{\text{cue}}; O_1)$ is only nonzero for the interval $50 \leq t \leq 70$ (Fig. 8A). We will now use similar reasoning to explain how the major components of our dynamical and informational explanations of the passive relational categorization agent are connected.

5.2. Cue stage

For the cue stage, the key conclusion of our information-theoretic analysis of the passive agent (Section 3.2.) was that waves of cue size information in the sensors are accumulated and stored in the output of neuron N3 at the end of the cue stage (Fig. 4). The key conclusion of our dynamical analysis of this same agent (Section 4.2.) was that the differing timing of sensor discontinuity crossings spreads the states of the interneurons along a curve according to cue size at the end of the cue stage (Fig. 6A). How do these two explanations relate?

Recall that the dynamical analysis was performed in the neuron state space, whereas the information-theoretic analysis was performed in neuron output space. If we translate Fig. 6A into output space, we find that N1 is always saturated on and N2 is always saturated off at the end of the cue stage. Thus, both analyses agree that N3 is the only neuron whose output varies significantly with cue size at the end of the cue stage.

Following the framework sketched in the previous section, we next compare the variation with cue size of N3 output trajectories to the time-dependent mutual information in N3 (Fig. 9A). As in the previous section, high mutual information after $t \approx 60$ corresponds to uniformity of spread in N3 outputs across as many of the available outcomes as possible, whereas low mutual information before $t \approx 5$ corresponds to a lack of spread. At intermediate times, the spread is less uniform, with trajectories confined to a smaller range of outputs and bunching up along the edges of this range. In addition, due to the overlapping of trajectories from distinct cue sizes, the functional relationship between z_{cue} and o_3 is folded over as in the example shown in Fig. 8C.

A more fine-grained picture of the cue stage information/dynamics relationship can be obtained by examining the specific mutual information in N3 (Fig. 9B). For example, we can see that N3 information about smaller cue sizes grows faster at the beginning of the cue stage because their trajectories are the first to spread out from the initial grouping (Arrow 1). In contrast, information about larger cue sizes peaks earlier toward the end of the cue stage because their trajectories are the first to completely separate from the main group so as to unambiguously specify their associated z_{cue} values (Arrow 2).

Thus, the cue size information in N3 can be directly related to the structure of the dependence of N3's trajectories on z_{cue} at any point in time. What about the process by which this information becomes available in N3? On the information-theoretic side, this occurs through periods of information transfer from each bilateral pair of sensors, whereas on the dynamical side, the spreading of trajectories is driven by the crossing of

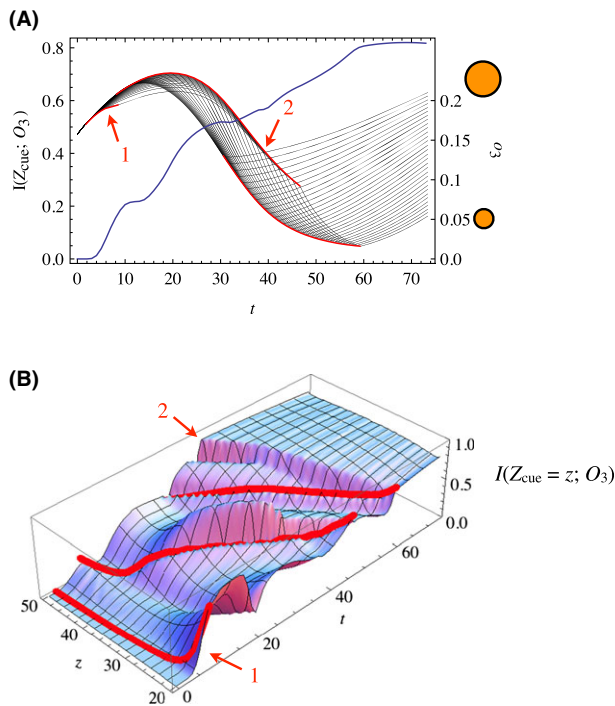


Fig. 9. Cue stage information dynamics. (A) The cue size information in N3 (dark blue curve) superimposed over the variation of N3 output trajectories with cue size (black curves) over time. (B). The specific cue size information in N3. The numbered arrows refer to corresponding events in parts A and B that are explained in the main text. In both plots, red curves highlight the discontinuity boundaries shown in Fig. 6.

discontinuity boundaries in pairs of sensors. It turns out that these two descriptions are also completely consistent. For example, cue size information transfer from S1/S7 to N3 occurs from $t \approx 30$ to $t \approx 60$, with transfer for larger cue sizes preceding transfer for smaller ones (right plot in Fig. 3C). As can be seen in Fig. 9A, the final vertical spreading of trajectories from crossing the S1/S7 discontinuity boundary (bottommost red curve) occurs over exactly the same range of time, with larger cue sizes spreading out before smaller ones. This correspondence can also be seen clearly in Fig. 9B, where the red curves of sensory discontinuity coincide exactly with several sharp increases of specific information in N3 about cue size.

5.3. Probe stage

For the probe stage, the key conclusion of our information-theoretic analysis of the passive agent (Section 3.3.) was that relative size information flows from N1 to MR and then becomes expressed in the agent's actual behavior (Fig. 4). The key conclusion of our dynamical analysis of this same agent (Section 4.3.) was that a bifurcation to bistability in the underlying interneuronal dynamics splits a manifold of trajectories spread out

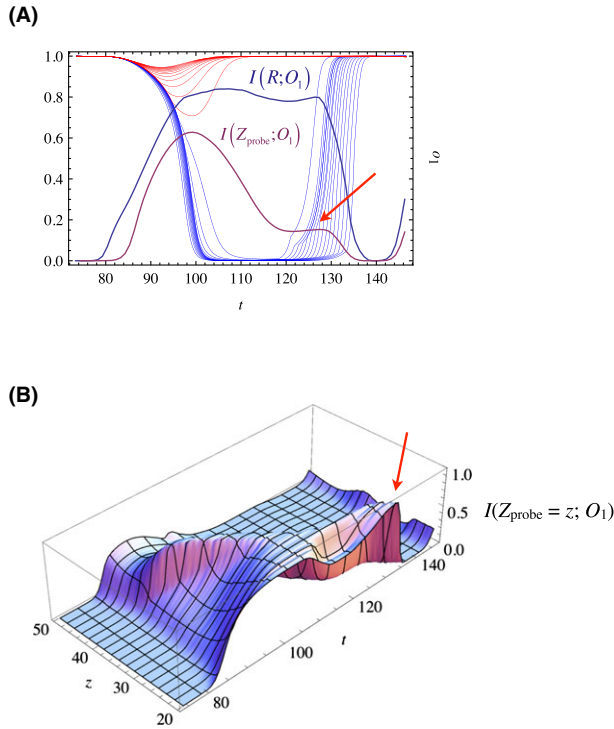


Fig. 10. Probe stage information dynamics. (A) The relative size and probe size information in N1 superimposed on the variation of N1 output trajectories with probe size (with blue trajectories corresponding to catches and red trajectories corresponding to avoidances). (B) The specific probe size information in N1. The arrows refer to corresponding events in parts A and B that are explained in the main text.

by both z_{cue} and z_{probe} into subsets that result in catch and avoid responses (Fig. 7). How do these two explanations relate?

Once again, we answer this question by comparing the structure of a bundle of neuronal trajectories to the information that they carry (Fig. 10A). Here, we can see the sharp divergence between the $z_{\text{probe}} < z_{\text{cue}}$ (blue) and $z_{\text{probe}} > z_{\text{cue}}$ (red) N1 trajectories as they are split by the boundary between the basins of attraction of the two stable equilibrium points that have just bifurcated into existence (Fig. 7C, D). We can also see that relative size information in N1 (dark blue curve) is associated with the extent of vertical separation between the smaller-than and larger-than curves, peaking when the two bundles have completely separated at intermediate times and falling to zero when there is no separation between them at early and late times.

Interestingly, this example also illustrates how the same neuronal element can simultaneously carry information about multiple things. This is a natural consequence of Shannon's definition of information, as measuring the value of one random variable can reduce our uncertainty about many others with which it interacts. During the probe stage, N1 not only carries information about relative object size, but it also carries

information about the absolute size of the probe object (dark magenta curve in Fig. 10A). This is possible because the vertical spreading of N1 trajectories can have a hierarchical structure: They can be broadly separated into smaller-than and larger-than bundles, but the trajectories in the individual bundles can themselves be separated on a finer scale by probe size. Note that probe size information peaks when there is greatest uniformity of spread within the individual red and blue bundles, whereas relative size information peaks where there is greatest uniformity of spread between the two bundles.

Finally, we can once again obtain a more detailed understanding of the information/dynamics relationship by examining specific information plots. For example, a plot of the specific probe size information (Fig. 10B) shows that the flattened portion of $I(Z_{\text{probe}}; O_1)$ in Fig. 10A (arrow) corresponds to information only about smaller objects, not larger ones in Fig. 10B (arrow). This is consistent with the fact that there is significant vertical spread in the $z_{\text{probe}} < z_{\text{cue}}$ (blue) trajectories, but not in the $z_{\text{probe}} > z_{\text{cue}}$ (red) ones at this point (arrow).

6. Discussion

In this article, we have attempted to accomplish three things. First, we have tried to ground the debate between information processing and dynamical approaches by examining its central issues in the context of a concrete model agent engaged in a relational categorization task. Because the model is evolved rather than designed with some a priori mode of operation in mind, it can plausibly admit of either an information processing or dynamical explanation. This allows us to replace competing intuitions about how a cognitive agent ought to work with an empirical investigation of how at least a simple model agent actually does work.

Second, we have tried to move the terms of the debate away from those of confrontation. Instead, we have suggested that information theory and dynamical systems theory should be conceived as merely different mathematical lenses through which we can examine the operation of a cognitive agent. Each lens provides a unique explanatory perspective, distinctive language, and a complementary set of tools for framing and answering certain kinds of questions. Thus, applying the language of IT or DST to a given cognitive system is never “right” or “wrong.” Although specific theories couched in their mathematical terms can certainly be subjected to empirical refutation, the mathematical languages themselves are merely more or less useful to a given purpose. In addition, we have proposed specific extensions to the mathematical tools of IT to better address the flow of information over time through a cognitive system.

Third, by applying both mathematical lenses to the same agent, we have tried to open a space for discussion on the relationship between information processing and dynamical styles of explanation. What differing insights do they each offer? What are their relative strengths and weaknesses? Through analysis of our model agent, we have proposed an explicit framework for bridging between the terms of an informational

explanation and those of a dynamical explanation, thus suggesting a specific path for a formal reconciliation between them.

In the remainder of this article, we briefly examine a number of related ongoing controversies within cognitive science in light of the model and analysis presented here. Because the literature on many of these topics is quite large and diverse, it is difficult to do full justice to the many different perspectives on this range of interrelated issues in the space available. Thus, we limit ourselves here to brief comments and suggestions that we hope will stimulate further discussion.

6.1. Information processing

Historically, discussions of information processing in cognition have been grounded in the metaphor, technology, and formalism of digital computation (Fodor, 1975; Newell & Simon, 1976; Pylyshyn, 1984). On this view, cognitive processes involve the rule-governed manipulation of discrete, symbolic structures. However, it has recently been argued that the notions of information processing and computation should be separated, opening a space for broader discussion of the ways in which systems can process information (Piccinini & Scarantino, 2011).

Our analysis has illustrated in some detail at least one way in which cognitive science can move beyond classical notions of information processing to a kind of generalized information processing. It is well known that the standard tools of information theory allow one to quantify the mutual information between, say, the activity of a neuron and some property of a distal stimulus. But we have shown how such tools can be extended to unroll mutual information over time and over individual stimulus values, to track its gain, loss, and transfer across components, and to identify synergistic interactions between them. This has allowed us to provide a very fine-grained analysis of the way information about absolute and relative object size flows throughout our model agent's "nervous system" during the course of its categorization decision. Although the details of this flow would undoubtedly differ for each evolved agent that we analyze, the general terms and nature of the information processing explanation that we have presented would not. This explanation is characterized by a graded, probabilistic, time-varying, and distributed notion of information that can be synergistically spread across multiple components, with the same components carrying information about multiple sources simultaneously. We believe that this formulation of information processing is much better suited to a cognitive science that incorporates situated, embodied and dynamical ideas.

The concept of internal representation plays a central role in the classical information processing perspective on cognition. What role, if any, does this notion play in the operation of our model agent? The fact that there is no universally accepted definition for the concept of representation makes it difficult to provide a definitive answer to this question. As Dietrich and Markman (2003) say: "It is striking that cognitive science does not have a theory of representation. There are many theories out there, but none have the support of anything like the majority of researchers. How could a notion that is so important be so poorly understood?" Nevertheless, we can briefly evaluate the operation of our model

agent with respect to some of the common intuitions drawn from this complicated and contentious literature (Bechtel, 1998; Beer, 2003; Brooks, 1991; Chemero, 2009; Clark & Toribio, 1995; Markman & Dietrich, 2000a; Ramsey, 2007).

At the heart of the concept of an internal representation is the idea of a state internal to an agent that correlates with a distal property. Markman and Deitrich (2000b) call this basic notion a mediating state. Our model agent clearly exhibits such correlations. An additional requirement that is often added before such correlated internal states can be considered representations is that they must be enduring; they must be capable of outlasting, at least under some circumstances, the sensory stimuli that originally produced them. This is the sense in which one talks about representations standing in for objects or properties that are not currently perceptually accessible (Haugeland, 1991). In our model agent, the storage of cue size information beyond the end of the cue stage in the curve C_{cue} is clearly an example of such an enduring state. Finally, representation usually requires that internal states under consideration play an actual causal role in the agent's subsequent behavior; spurious correlations are not considered representational. Although Shannon information is a purely correlational notion, causality can be established in many ways, including by direct dynamical analysis as we have done here or by indirect experimental manipulations such as information lesioning (Keinan, Meilijson, & Ruppin, 2003), perturbation measures (Ay & Polani, 2008), or controllability tests (Mirolli, 2012). Thus, if all one means by a representation is a potentially enduring causally efficacious information-bearing state, then our model agent clearly has them and the information-theoretic tools demonstrated here provide a powerful way of identifying them in a mathematically rigorous manner.

However, the standard criticism of defining representations in this way is that doing so identifies representations everywhere in the physical world, which potentially makes the notion of representation vacuous or, at a minimum, fails to distinguish what is unique about *mental* representations, taken to be a hallmark property of the mind, from other kinds of information-bearing states. For example, when a storm front passes over a given patch of terrain, its states acquire persistent correlations with various properties (temperature, moisture, topography) of that terrain which can profoundly influence the front's subsequent development (as anyone who has ever experienced lake-effect snow has observed firsthand).

The standard response to this criticism is to insist that what distinguishes a genuine representation from a mere correlated internal state (even a causally efficacious one) is that that state must have the purpose or function of carrying the information it does, with the relevant notion of function usually cashed out in terms of learning (Dretske, 1981) or evolution (Millikan, 1984). This teleological notion of an information-carrying function is needed, it is argued, to allow for the possibility of misrepresentation; that is, to allow for the possibility that something can be represented as being so even when it is not. Since a device retains its normative function even when it fails to perform it, a state can retain its information-carrying function even when it fails to convey the information that it is supposed to provide. Because our model was evolved for performing a relational categorization task, such a functional story can be told about its operation. For example, a

necessary component of this task is to store the size of the cue object for later comparison with the probe object, and our analysis revealed that the curve C_{cue} serves this function in the passive agent. Thus, on this view of representation, we can say that a point along C_{cue} represents the cue object as being a particular size. Furthermore, if we perturb the state to a different point along this curve, our agent can be said to misrepresent the size of the cue object, which would entail specific predictions about its behavior upon seeing the probe object.

To what extent can the notion of computation—the intellectual core of the classical information processing approach—be applied to our model agent? Once again, this is a difficult question to answer conclusively given the many ways that computation enters into cognitive science: as a metaphor, a human capacity to be explained, a mathematical theory, a simulation technology, and a theoretical framework (Beer, 1995b). Clearly our agent metaphorically “computes” an answer to the question “Is the second object smaller than the first?”. In addition, we obviously utilize computational technology to simulate it. But does computation play any functional role in the operation of this agent? Because there is no widespread agreement within cognitive science on how to answer such a question (Harnad, 1994; Scheutz, 2002), we briefly explore it within the context of one particular recently proposed taxonomy.

Piccinini and Scarantino (2011) define a notion of generic computation and then argue that digital computation and analog computation are strict, nonoverlapping subsets of this notion, while also leaving room for other possible subsets such as neural computation (Piccinini & Bahar, 2013). On this view, our model agent does not perform digital computation because its operation does not rely on the manipulation of strings of discrete digits. This is not a trivial statement, as it is certainly possible to produce recurrent neural networks whose operation does functionally depend on approximating some digital computation (Casey, 1996; Phattanasri, Chiel, & Beer, 2007). As for analog computation, Piccinini and Scarantino suggest that the usage of this term in the literature is quite vague and the only clear definition is that put forward by Pour-El (1974). However, this seems unnecessarily narrow, as there are many other competing mathematical definitions (for a recent review, see MacLennan, 2013). In addition, this body of mathematical work is focused not on prescribing what is and is not an analog computation, but rather on applying concepts from the formal theory of computation (e.g., computability, decidability and complexity) to continuous dynamical systems. Thus, it is not clear whether or not our model agent would qualify as performing analog computation according to this taxonomy.

Finally, Piccinini and Scarantino (2011) define generic computation as “the processing of vehicles according to rules that are sensitive to certain vehicle properties and, specifically, to differences between different portions of the vehicles” in a medium-independent way. They explicitly state that possible vehicles include continuous variables (as well as neuronal spike trains), which is what makes generic computation a generalization of digital computation. Indeed, they argue that, as Shannon information is a medium-independent notion, any system that processes Shannon information is performing generic computation. Thus, by their definition, our agent is doing so.

6.2. Dynamics

For over 15 years, van Gelder's (1998) dynamical hypothesis has been the center around which debates about the role of dynamical ideas in cognitive science have turned (see also van Gelder, 1995; van Gelder & Port, 1995). This is unfortunate because, although the diversity of dynamical approaches is becoming widely recognized (Zednik, 2011), critical discussion has focused almost exclusively on van Gelder's particular formulation (Bechtel, 1998; Eliasmith, 2001; Grush, 1997). This formulation puts forward both ontological ("cognitive systems are dynamical systems") and epistemological ("cognitive systems are best understood as dynamical systems") claims. From our perspective of treating DST as one of a set of available mathematical lenses, the ontological claim makes no sense, whereas we would qualify the epistemological claim with "under some circumstances" or "for some purposes" (Beer, 1998). In addition, discussion between van Gelder and his critics has mostly emphasized a view of dynamical systems as continuous, low-dimensional and focused on attractors and their bifurcations, although none of these properties are required by dynamical systems theory.

A central issue in the debate between classical information processing and dynamical approaches has been the extent to which cognitive processes should be considered to be discrete or continuous. DST itself is neutral on this issue, supporting any combination of discreteness and continuity in state and in time. Indeed, as many authors have pointed out, symbolic dynamics may provide a mathematical bridge between continuous and discrete perspectives on cognitive systems (Crutchfield, 1998; Dale & Spivey, 2005; Tabor, 2009). From any continuous-state system, these techniques allow one to extract and study a discrete-state system by partitioning the continuous state into a finite number of subsets, although the corresponding discrete system preserves the dynamics of the original continuous one only for very special choices of partition that can be extremely difficult to find (Kennel & Buhl, 2003), and nongenerating partitions can seriously misrepresent the underlying dynamics (Bollt, Stanford, Lai, & Zyczkowski, 2001). Interestingly, discussions of continuity within cognitive science have not fully considered the richness and subtlety of this concept in mathematics. For example, there can be different degrees of continuity to a dynamical system (e.g., discontinuous state, continuous state but discontinuous first derivative, etc.) and these differences can have profound consequences on its dynamics (di Bernardo et al., 2008).

How do these considerations apply to our relational categorization model? Our model is obviously continuous in both state and time.² However, discontinuities enter into its dynamics in at least two ways. First, the nature of the object/ray intersections introduce discontinuities into the dynamics of this model (Fig. 6). Mathematically, this makes our model a piecewise-smooth dynamical system (di Bernardo et al., 2008). As described in Section 4.3, these sensory discontinuities play a key role in spreading out the trajectories of the system along the curve \mathcal{C}_{cue} . Second, although our model's state evolves continuously, its ultimate behavior diverges sharply into either a catch or avoid response. Thus, it provides a nice illustration of the idea that continuously unfolding cognitive processes can lead to categorical decisions *as coarse-grained into discrete choices by an external*

observer (Dale & Spivey, 2005; Spivey, 2007). Furthermore, our model seems to provide evidence against the claim by Dietrich and Markman (2003) that categorical discrimination *requires* fundamentally discrete cognitive processes as the divergence of nearby trajectories can occur smoothly in a continuous system.

Discussions of dynamical approaches to cognition often mistakenly assume that dynamical analysis is limited to low-dimensional systems. This is simply false: Although our ability to completely *visualize* the dynamics of a system depends on it being low-dimensional, the mathematical concepts of DST itself are defined for arbitrary dimension. However, there are at least two ways in which a low-dimensional description of a high-dimensional dynamical system may be appropriate. First, the dynamics of a high-dimensional system can temporarily collapse to a low-dimensional pattern under appropriate constraints (e.g., when the enormous degrees of freedom of the human neuromuscular system reduce to a low-dimensional rhythmic pattern during juggling) or due to saturation effects in its elements (when any of our model neurons saturate on or off, they become constant inputs to the other neurons and effectively drop out of the dynamics temporarily (Chiel, Beer, & Gallagher, 1999)). Second, near a bifurcation, there is a mathematically rigorous sense in which the behavior of all the degrees of freedom in a high-dimensional system become chained to a few so-called collective variables (Haken, 1983; Kelso, 1995). When low-dimensional descriptions are appropriate, it obviously makes good theoretical sense to take advantage of them.

Discussions of dynamical approaches in cognitive science also typically focus on attractors and their bifurcations. There are many very good reasons for this. Attractors are discrete, low-dimensional, robust entities for which the mathematical tools of DST are most highly developed. Attractors can also be a useful concept when an entire brain-body-environment system is engaged in some repetitive task or when a nervous system is receiving constant sensory input. But neither of these situations is typical. In general, the rich internal dynamics of nervous systems are continually being driven by time-varying sensory signals, causing the internal states to be always chasing—but rarely catching—constantly moving and occasionally bifurcating attractors. In the dynamical analysis performed in this paper, we have shown in detail how the interplay between intrinsic neural dynamics and time-varying sensory signals parameterized by features of the external environment induce time-varying manifolds of transient states, with the movement and deformation of these manifolds playing a central role in the dynamical explanation of our model agent. The importance of externally driven dynamics in cognitive science has also been emphasized by Spivey (2007) and, most recently, by Hotton and Yoshimi (2011), and the study of such nonautonomous dynamical systems is an active area of research within DST (Barreira & Valls, 2008; Pötzsche, 2010; Rasmussen, 2007).

6.3. *Information processing and dynamics*

Information processing and dynamical approaches to cognition are often considered to be deeply at odds with one another, offering fundamentally conflicting accounts of cognitive processes. These two perspectives certainly do seem to bring very different intuitions

about cognition to the explanatory table. This presumed incompatibility significantly raises the stakes of any debate between them. If only one side can be right, then the other one must be wrong.

However, when we adopt the more neutral position that IT and DST are merely mathematical lenses, then the relationship between the different explanations that these lenses offer becomes an empirical question. Indeed, we have shown that, far from being in conflict, information-processing and dynamical explanations are deeply connected. *Information is necessarily dynamical*. The Shannon information carried by particular system components about a stimulus feature will in general vary over time, with subsets of components gaining, losing and transferring information to other components. Likewise, *dynamics is necessarily informational*. Interacting with an environment whose characteristics vary in some systematic way can induce a time-dependent manifold of variation in the internal states of an agent. Making measurements of subsets of these states at any point in time corresponds to a projection of this manifold, and Shannon information simply quantifies the extent to which such a projection preserves the original environmental variation. Furthermore, information gain, loss, and transfer can be directly related to the changing geometry of these manifolds and the consequences of these changes for the projections of interest.

We have found that information processing and dynamical explanations of our model agent are both consistent and complementary. They are consistent because they are different ways of talking about the same underlying property, namely the variation in neuronal states with object sizes over time. These two explanatory frameworks also have complementary strengths and weaknesses. For example, although DST can describe the geometry of the N3 projection of the passive agent's trajectories at the end of the cue stage, IT is required to characterize the significance of that structure with respect to the agent's environment and task. Conversely, although IT can describe the rise of relative size information in N1 as an important event in the agent's categorization, DST is required to understand the bifurcation that underlies it.

Although a number of authors have suggested that, suitably generalized, informational and dynamical explanations of cognition need not be in conflict (Crutchfield, 1998; Mitchell, 1998; Rocha & Hordijk, 2005), the work by Crutchfield on computational mechanics (CM) is probably the most extensive examination of this issue. In particular, he and his collaborators have developed a method for producing stochastic automata models of a discretized time series from measurements of a process (Crutchfield & Young, 1989) which can be proven to be minimal and optimally predictive given the discretization (Shalizi & Crutchfield, 2001). The original motivation for this work was to use properties of the induced automata to define measures of complexity and to formally characterize emergence (Crutchfield, 1994). While our work is broadly similar, it is rather different in emphasis and approach. First, whereas CM begins with a discrete time series derived from measurements of a presumed unknown underlying dynamics, we begin with trajectories of continuous states of a known dynamical system. Second, whereas CM seeks to construct a model of the given time series and analyze the properties of this model, we seek to relate the time-varying geometry of manifolds of states of the given

dynamical system to the flow of information throughout the system's components. Finally, whereas the focus of CM is on characterizing the intrinsic information processing implicit in a time series of measurements of an arbitrary process, we are interested in understanding how an agent evolved to perform a specific cognitive task processes extrinsic information about its environment.

Within cognitive science itself, Smolensky has been developing an influential architecture that seeks to reconcile macrolevel discrete symbolic computational descriptions of cognition with microlevel continuous subsymbolic connectionist descriptions (Smolensky, 1986, 2012). In his framework, the macrolevel typically takes the form of transformations of tree-like filler-role structures and the microlevel takes the form of relaxation over the distributed activation of representational microfeatures to an attractor that optimizes a particular measure of consistency and symbolic interpretability. This architecture has been successfully applied to a variety of problems in linguistic processing (Smolensky, Goldrick, & Mathis, 2013; Smolensky & Legendre, 2006). However, although an early paper on Smolensky's work was provocatively entitled "Information processing in dynamical systems," our work differs from Smolensky's in several ways. First, unlike Smolensky, we are proposing no specific symbolic or subsymbolic architecture for cognition. Rather, we seek to use the tools of IT and DST to discover through analysis the information-processing architecture and dynamics underlying any particular cognitive competence observed in a given agent. Second, whereas Smolensky's architecture focuses on attractors, we have emphasized the role of transient dynamics in cognitive processes. Finally, while Smolensky takes an internalist view of cognitive processing, our approach is specifically designed to accommodate a situated and embodied perspective.

More generally, the perspective on information processing and dynamical explanation that emerges from our analysis is broadly supportive of the notion of explanatory pluralism in cognitive science (see Dale, 2008 for a review). We certainly agree that ontological arguments about "what's really going on" in a cognitive system are unlikely to be resolved anytime soon, if ever. However, this does not mean that "everyone is right" and therefore intertheoretical debate should cease. Rather, we wish to emphasize the importance of seeking to articulate in detail the interrelations between different explanatory frameworks as we have tried to do in this article.

One interesting issue about this relationship in our model agent is the extent to which one explanation determines the other. Given a set of properties of interest, the dynamics of our model agent uniquely determines the information that any subset of its components carry about those properties. The converse, however, is false; The informational explanation in general underdetermines the dynamical one in this model. For example, different transient manifolds of states may produce the same informational quantities. Moreover, the dynamics does not by itself determine the properties of informational interest; these are externally imposed by the questions we ask about the system. For example, at any given time the same system component can carry information about absolute cue size, absolute probe size, relative object size, vertical distance to the object, horizontal offset to the object, and so on. Finally, although in our model the dynamical description is

complete (because it is defined at the level of the causal mechanisms in the model), this need not be the case. Dynamical explanations can be pitched at different levels. When both dynamical and informational descriptions abstract over the underlying causal mechanisms, neither may uniquely determine the other.

6.4. *Situatedness and embodiment*

To examine the nature of and relationship between informational and dynamical explanations in the simplest possible context, we have focused on analyzing the internal neuronal processing in passive agents. However, the situated and embodied nature of our relational categorization model and the existence of evolved agents employing an active strategy also make this model an ideal vehicle for exploring the role that ongoing feedback through the body and environment can play in cognitive processes (Clark, 1997; Varela, Thompson & Rosch, 1991). The tools of IT, DST, and the IT/DST framework that we have put forward can also be applied to entire brain-body-environment systems without modification; we merely need to include the relevant body and environment states in our analysis. As demonstrated in Section 3.4 with the tools of IT, such analyses of situated and embodied agents can reveal interestingly different processing strategies from those observed in agents making no essential use of feedback through the body and environment. Further development of this approach could lead to a framework for rigorously analyzing embodied information processing.

6.5. *Methodology*

Our focus in this article has been on the detailed analysis of simple model agents exhibiting a cognitively interesting behavior. Unlike empirically grounded models, the specific inferences that one could hope to make about relational categorization in animals or humans from our model are obviously extremely limited. However, experimental verification is not the purpose of such models. Rather, they are designed to explore in concrete ways fundamental conceptual issues in the foundations of a field (Beer, 2003; Beer & Williams, 2009). For example, only in such “toy” models can we currently hope to perform sufficiently detailed dynamical and information-theoretic analysis of a cognitive agent that we can directly confront questions about their relationship. Toy models are rigorous examples of what Dennett (1984) has called “intuition pumps.” But they are much more than this. As we have seen in this article, they can also be powerful “theory pumps,” leading to the development of new mathematical tools and theoretical frameworks. In addition, they can sometimes serve as “experiment pumps,” suggesting new kinds of experimental questions to be asked even if they cannot directly predict the answers. Thus, although the careful study of toy models of minimally cognitive behavior alone may never tell us exactly how human cognition works, it may play a vital role in discovering such an explanation and building the right conceptual framework and theoretical language for formulating it rigorously.

Acknowledgments

We would like to thank Colin Allen, Anthony Chemero, Andy Clark, and Eduardo Izquierdo for their feedback on an earlier draft of this article. In addition, we thank Rick Dale and the reviewers for their suggestions. This work was supported in part by NSF grants IIS-0916409 and IIC-1216739 to R.D.B. and an NSF IGERT fellowship to P.L.W.

Notes

1. Most authors use the terms “information processing” and “computation” interchangeably, but Piccinini and Scarantino (2011) have argued that these notions should be distinguished. We will use the term “information processing” in this paper and return to a discussion of computation at the end.
2. Of course, our model is simulated on a digital computer and thus the states and times are ultimately discrete. However, in a properly implemented model of a continuous system, this fact plays no *functional* role in its operation.

References

- Amblard, P. O., & Michel, O. J. J. (2009). Measuring information flow in networks of stochastic processes. arXiv preprint 0911.2873.
- Ashby, W. R. (1960). *Design for a brain* (2nd ed.). New York: Wiley.
- Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7, 358–366.
- Ay, N., & Polani, D. (2008). Information flows in causal networks. *Advances in Complex Systems*, 11(1), 17–42.
- Barnett, L., Buckley, C. L., & Bullock, S. (2009). Neural complexity and structural connectivity. *Physical Review E*, 79(5), 051914.
- Barreira, L., & Valls, C. (2008). *Stability of nonautonomous differential equations*. Berlin: Springer.
- Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist’s challenge in cognitive science. *Cognitive Science*, 22, 295–318.
- Beer, R. D. (1995a). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72, 173–215.
- Beer, R. D. (1995b). Computational and dynamical languages for autonomous agents. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 121–147). Cambridge, MA: MIT Press.
- Beer, R. D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Mataric, J. -A. Meyer, J. Pollack & S. Wilson (Eds.), *From animals to animats 4: Proceedings of the fourth international conference on simulation of adaptive behavior* (pp. 421–429). Cambridge, MA: MIT Press.
- Beer, R. D. (1998). Framing the debate between computational and dynamical approaches to cognitive science (A commentary on “The dynamical hypothesis in cognitive science” by Tim van Gelder). *The Behavioral and Brain Sciences*, 21(5), 630.

- Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11(4), 209–243.
- Beer, R. D., & Williams, P. L. (2009). Animals and animats: Why not both iguanas? *Adaptive Behavior*, 17, 296–302.
- di Bernardo, M., Budd, C. J., Champneys, A. R., & Kowalczyk, P. (2008). *Piecewise-smooth dynamical systems: Theory and applications*. London: Springer.
- Bollt, E. M., Stanford, T., Lai, Y.-C., & Zyczkowski, K. (2001). What symbolic dynamics do we get with a misplaced partition? On the validity of the threshold crossings analysis of chaotic time-series. *Physica D: Nonlinear Phenomena*, 154, 259–286.
- Brenner, N., Strong, S. P., Koberle, R., Bialek, W., de Ruyter, R., & van Steveninck, (2000). Synergy in a neural code. *Neural Computation*, 12(7), 1531–1552.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Butts, D. A. (2003). How much information is associated with a particular stimulus? *Network*, 14, 3371–3374.
- Casey, M. (1996). The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction. *Neural Computation*, 8(6), 1135–1178.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.
- Chiel, H. J., Beer, R. D., & Gallagher, J. C. (1999). Evolution and analysis of model CPGs for walking I. Dynamical modules. *Journal of Computational Neuroscience*, 7(2), 99–118.
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Clark, A., & Toribio, J. (1995). Doing without representing? *Synthese*, 101, 401–431.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed). New York: Wiley-Interscience.
- Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics and induction. *Physica D: Nonlinear Phenomena*, 75, 11–54.
- Crutchfield, J. P. (1998). Dynamical embodiments of computation in cognitive processes. (A commentary on “The dynamical hypothesis in cognitive science” by Tim van Gelder). *The Behavioral and Brain Sciences*, 21(5), 635.
- 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 and 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.
- Dennett, D. C. (1984). *Elbow room*. Cambridge, MA: MIT Press.
- DeWeese, M. R., & Meister, M. (1999). How to measure the information gained from one symbol. *Network*, 10, 325–340.
- Diacu, F., & Holmes, P. (1996). *Celestial encounters*. Princeton: Princeton University Press.
- Dietrich, E., & Markman, A. B. (2003). Indiscrete thoughts: Why cognition must use discrete representation. *Mind & Language*, 18(1), 95–119.
- Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Eckhorn, R., & Popel, B. (1974). Rigorous and extended application of information theory to the afferent visual system of the cat. i. basic concepts. *Kybernetik*, 16, 191–200.
- Eckhorn, R., & Popel, B. (1975). Rigorous and extended application of information theory to the afferent visual system of the cat. ii. experimental results. *Biological Cybernetics*, 17(1), 7–17.
- Eliasmith, C. (2001). Attractive and in-discrete: A critique of two putative virtues of the dynamicist theory of mind. *Minds and Machines*, 11, 417–426.
- Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2, 56–78.

- van Gelder, T. (1995). What might cognition be if not computation? *The Journal of Philosophy*, 97(7), 345–381.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *The Behavioral and Brain Sciences*, 21, 615–655.
- van Gelder, T., & Port, R. F. (1995). It's about time: An overview of the dynamical approach to cognition. In R. F. Port, & T. van Gelder (Eds.), *Mind as motion* (pp. 1–43). Cambridge, MA: MIT Press.
- Gentner, D., & Kurtz, K. (2005). Relational categories. In P. Wolf, A. W. -K. , R. L. Goldstone, B. C. Love & A. B. Markman (Eds.), *Categorization inside and outside the laboratory* (pp. 151–175). Washington, DC: APA.
- Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature*, 6831, 930–933.
- Grush, R. (1997). A review of Port and van Gelder's *Mind as motion*. *Philosophical Psychology*, 10(2), 233–242.
- Haken, H. (1983). *Synergetics* (3rd ed.). Berlin: Springer-Verlag.
- Harnad, S., Ed. (1994). What is computation (and is cognition that)? *Special issue of Minds and Machines*, 4 (4).
- Harvey, I., Di Paolo, E., Wood, R., Quinn, M., & Tuci, E. (2005). Evolutionary robotics: A new scientific tool for the study of cognition. *Artificial Life*, 11, 79–98.
- Haugeland, J. (1991). Representational genera. In W. Ramsey, S. P. Stich, & D. E. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 61–89). Hillsdale, NJ: Lawrence Erlbaum.
- Hinrichs, H., Heinze, H. J., & Schoenfeld, M. A. (2006). Causal visual interactions as revealed by an information theoretic measure and fMRI. *NeuroImage*, 31(3), 1051–1060.
- Honey, C. J., Kotter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Science USA*, 104(24), 10240–10245.
- Hotton, S., & Yoshimi, J. (2011). Extending dynamical systems theory to model embodied cognition. *Cognitive Science*, 35(3), 444–479.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Keinan, A., Meilijson, I., & Ruppín, E. (2003). Controlled analysis of neurocontrollers with informational lesioning. *Philosophical Transactions of the Royal Society of London. Series A*, 361, 2123–2144.
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
- Kennel, M. B., & Buhl, M. (2003). Estimating good discrete partitions from observed data: Symbolic false nearest neighbors. *Physical Review Letters*, 91, 084102.
- Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18, 513–549.
- Kjaer, T. W., Hertz, J. A., & Richmond, B. J. (1994). Decoding cortical neuronal signals: Network models, information estimation and spatial tuning. *Journal of Computational Neuroscience*, 1(1), 109–139.
- Kurtz, K., & Boukrina, O. (2004). Learning relational categories by comparison of paired examples. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 756–761). Mahwah, NJ: Lawrence Erlbaum.
- Kuznetsov, Y. A. (2004). *Elements of applied bifurcation theory* (3rd ed.). New York: Springer.
- Lasota, A., & Mackey, M. C. (1994). *Chaos, fractals, and noise: stochastic aspects of dynamics* (2nd ed.). New York: Springer-Verlag.
- Latham, P. E., & Nirenberg, S. (2005). Synergy, redundancy, and independence in population codes, revisited. *Journal of Neuroscience*, 25(21), 5195–5206.
- Liang, H., Ding, M., & Bressler, S. L. (2001). Temporal dynamics of information flow in the cerebral cortex. *Neurocomputing*, 38–40, 1429–1435.

- Lizier, J. T., Heinzle, J., Horstmann, A., Haynes, J. D., & Prokopenko, M. (2011). Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of Computational Neuroscience*, 30(1), 85–107.
- Lungarella, M., & Sporns, O. (2006). Mapping information flow in sensorimotor networks. *PLoS Computational Biology*, 2, 1301–1312.
- MacLennan, B. (2013). Analog computation. In R. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 271–294). New York: Springer.
- Markman, A. B., & Deitrich, E. (2000b). In defense of representation. *Cognitive Psychology*, 40, 138–171.
- Markman, A. B., & Dietrich, E. (2000a). Extending the classical view of representation. *Trends in Cognitive Sciences*, 4, 70–75.
- Markman, A., & Stilwell, C. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, 13(4), 329–358.
- Millikan, R. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.
- Mirolli, M. (2012). Representations in dynamical embodied agents: Re-analyzing a minimally cognitive model agent. *Cognitive Science*, 36, 870–895.
- Mitchell, M. (1998). A complex systems perspective on the “Computation vs. Dynamics” debate in cognitive science. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 710–715). Mahwah, NJ: Lawrence Erlbaum.
- Narayanan, N. S., Kimchi, E. Y., & Laubach, M. (2005). Redundancy and synergy of neuronal ensembles in motor cortex. *Journal of Neuroscience*, 25(17), 4207–4216.
- Newell, A., & Simon, H. A. (1976). Computer science as an empirical enquiry: Symbols and search. *Communications of the ACM*, 19, 113–126.
- Neymotin, S. A., Jacobs, K. M., Fenton, A. A., & Lytton, W. W. (2011). Synaptic information transfer in computer models of neocortical columns. *Journal of Computational Neuroscience*, 30(1), 69–84.
- Panzeri, S., Schultz, S. R., Treves, A., & Rolls, E. T. (1999). Correlations and the encoding of information in the nervous system. *Philosophical Transactions of the Royal Society. B*, 266(1423), 1001–1012.
- Phattanasri, P., Chiel, H. J., & Beer, R. D. (2007). The dynamics of associative learning in evolved model circuits. *Adaptive Behavior*, 15(4), 377–396.
- Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science*, 37(3), 453–488.
- Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics*, 37, 1–38.
- Pötsche, C. (2010). *Geometric theory of discrete nonautonomous dynamical systems*. Berlin: Springer.
- Pour-El, M. B. (1974). Abstract computability and its relation to the general purpose analog computer (some connections between logic, differential equations and analog computers). *Transactions of the American Mathematical Society*, 199, 1–28.
- Puchalla, L., Schneidman, E., Harris, R. A., & Berry, M. J. (2005). Redundancy in the population code of the retina. *Neuron*, 46, 493–504.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Quiroga, R. Q., & Panzeri, S. (2009). Extracting information from neuronal populations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, 10, 173–185.
- Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge, MA: Cambridge University Press.
- Rasmussen, M. (2007). *Attractivity and bifurcation for nonautonomous dynamical systems*. Berlin: Springer.
- Reich, D. S., Mechler, F., & Victor, J. D. (2001). Independent and redundant information in nearby cortical neurons. *Science*, 294, 2566–2568.
- Rocha, L. M., & Hordijk, W. (2005). Material representations: From the genetic code to the evolution of cellular automata. *Artificial Life*, 11, 189–214.

- Saldanha, E., & Bitterman, M. (1951). Relational learning in the rat. *The American Journal of Psychology*, 64(1), 37–53.
- Scheutz, M. (Ed.) (2002). *Computationalism: New directions*. Cambridge, MA: MIT Press.
- Schneidman, E., Bialek, W., & Berry, M. J. (2003). Synergy, redundancy, and independence in population codes. *Journal of Neuroscience*, 23(37), 11539–11553.
- Schöner, G. (2008). Dynamical systems approaches to cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 101–126). Cambridge, UK: Cambridge University Press.
- Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85, 461–464.
- Scott, D. W. (1985). Average shifted histograms: Effective nonparametric density estimators in several dimensions. *The Annals of Statistics*, 13, 1024–1040.
- Shalizi, C. R., & Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104, 819–881.
- Shannon, C. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27(379–423), 623–656.
- Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), *Parallel distributed processing, volume 1: Foundations* (pp. 194–281). Cambridge, MA: MIT Press.
- Smolensky, P. (2012). Symbolic functions from neural computation. *Philosophical Transactions of the Royal Society of London. Series A*, 370, 3543–3569.
- Smolensky, P., Goldrick, M., & Mathis, D. (2013). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete. *Cognitive Science*, doi:10.1111/cogs.12047.
- Smolensky, P., & Legendre, G. (2006). *The harmonic mind*. Cambridge, MA: MIT Press.
- Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.
- Sporns, O. (2011). *Networks of the brain*. Cambridge, MA: MIT Press.
- Strogatz, S. H. (1994). *Nonlinear dynamics and chaos*. Reading, MA: Addison Wesley.
- Szczepanski, J., Arnold, M., Wajnryb, E., Amigó, J. M., & Sanchez-Vives, M. V. (2011). Mutual information and redundancy in spontaneous communication between cortical neurons. *Biological Cybernetics*, 104, 161–174.
- Tabor, W. (2009). A dynamical systems perspective on the relationship between symbolic and non-symbolic computation. *Cognitive Neurodynamics*, 3, 415–427.
- Theunissen, F. E., & Miller, J. P. (1991). Representation of sensory information in the cricket cercal sensory system. ii. information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons. *Journal of Neurophysiology*, 66(5), 1690–1703.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind, Cognitive science and human experience*. Cambridge, MA: MIT Press.
- Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, 113, 358–389.
- Williams, P. L., & Beer, R. D. (2010a). Nonnegative decomposition of multivariate information. *arXiv preprint, 1004*, 2515.
- Williams, P. L., & Beer, R. D. (2010b). Information dynamics of evolved agents. In S. Doncieux, B. Girard, A. Buillot, J. Hallam, J.-A. Meyer, & J.-B. Mouret (Eds.), *From Animals to Animats 11: Proceedings of the 11th International Conference on the Simulation of Adaptive Behavior* (pp. 38–49). Berlin: Springer.
- Williams, P. L., Beer, R. D., & Gasser, M. (2008). An embodied dynamical approach to relational categorization. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 223–228). Austin, TX: Cognitive Science Society.
- Wills, S. (1999). Relational learning in pigeons? *The Quarterly Journal of Experimental Psychology: Section B*, 52(1), 31–52.
- Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science*, 78(2), 238–263.

Appendix

The central concept of information theory is entropy, which is defined for a discrete random variable $X \in \{x_1, \dots, x_N\}$ as

$$H(X) = \sum_i p(x_i) \log \frac{1}{p(x_i)} \quad (2)$$

where $p(x)$ gives the probability of occurrence of x . The mutual information between two discrete random variables X and Y is defined as

$$I(X; Y) = \sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (3)$$

where $p(x, y)$ is the joint probability distribution of X and Y .

We use a measure of specific information to unroll mutual information over measurement outcomes. Such measures have been applied previously only to studies of neural coding (Butts, 2003; DeWeese & Meister, 1999), for example, in the cat visual system (Eckhorn & Popel, 1974, 1975), the cercal system of crickets (Theunissen & Miller, 1991), and the primary visual cortex of monkeys (Kjaer, Hertz, & Richmond, 1994). Here, we make use of the following measure of specific information:

$$I(X = x_i; Y) = \sum_j p(y_j|x_i) \left[\log \left(\frac{1}{p(x_i)} \right) - \log \left(\frac{1}{p(x_i|y_j)} \right) \right] \quad (4)$$

where $j p(x|y)$ is the conditional probability of x given y . Intuitively, specific information measures how much knowing the value of Y reduces the surprise associated with learning that X takes on the specific value x_i .

When analyzing the information in a system, it is necessary to consider not only the information carried by individual variables but also the information that may be encoded redundantly or synergistically by multiple variables. The concepts of synergy and redundancy have been of great recent interest in studies of neural coding, including the use of synergy to detect population coding (Averbeck, Latham, & Pouget, 2006; Latham & Nirenberg, 2005; Panzeri, Schultz, Treves, & Rolls, 1999; Quiroga & Panzeri, 2009; Schneidman, Bialek, & Berry, 2003) and the encoding of information in compound spike events in single neurons (Brenner et al., 2000), and the use of redundancy to characterize the robustness of individual neurons to damage (Narayanan, Kimchi, & Laubach, 2005; Puchalla, Schneidman, Harris, & Berry, 2005; Reich, Mechler, & Victor, 2001; Szczepanski, Arnold, Wajnryb, Amigó, & Sanchez-Vives, 2011). We developed the method of partial information decomposition specifically to address several shortcomings in these earlier measures (Williams & Beer, 2010a). Whereas previous measures often confound synergistic and redundant interactions and have problematic interpretations when more than three variables are involved, measures based on partial information decomposition clearly delineate synergy and redundancy and extend straightforwardly to higher dimensions. Partial information decomposition is based on a measure of information overlap that satisfies certain axioms (Williams & Beer, 2010a). Different choices for this redundancy measure are

possible. In this article, we use a measure of redundancy that we call I_{\min} , which, for two sources S_1, S_2 , defines the redundant information that S_1 and S_2 provide about X as

$$I_{\min}(X; S_1, S_2) = \sum_x p(x) \min\{I(X = x; S_1), I(X = x; S_2)\} \quad (5)$$

or the minimum information that any source provides about each outcome of X , averaged over all possible outcomes. The information provided synergistically by S_1 and S_2 is then calculated using a form of inclusion-exclusion as

$$I_{\text{Syn}}(X; S_1, S_2) = I(X; S_1, S_2) - I(X; S_1) - I(X; S_2) + I_{\min}(X; S_1, S_2) \quad (6)$$

and corresponds to the information provided by the simultaneous knowledge of both S_1 and S_2 that is not available from either source alone.

The standard measures of dynamic information are based on transfer entropy (Amblard & Michel, 2009; Barnett, Buckley, & Bullock, 2009; Schreiber, 2000), which has been widely used in neuroscience to, for example, identify functional brain networks (Friston, 1994; Honey, Kotter, Breakspear, & Sporns, 2007; Sporns, 2011) and explore how they change with task difficulty (Hinrichs, Heinze, & Schoenfeld, 2006; Liang, Ding, & Bressler, 2001; Lizier, Heinze, Horstmann, Haynes, & Prokopenko, 2011; Neymotin, Jacobs, Fenton, & Lytton, 2011). The crucial difference between our measures and those based on transfer entropy is that transfer entropy quantifies the *total* information that one random process transfers to another, whereas our measures are designed to capture only the information that is gained or transferred *about* some other variable of interest. Because measuring dynamic information often requires characterizing the information relationships between more than two variables, we once again apply partial information decomposition. Using this approach, the information gained by Y about X at time t is defined as

$$I_G(X; Y_t) = I(X; Y_t) - I_{\min}(X; Y_{t-1}, Y_t) \quad (7)$$

which can be interpreted as the information that Y contains about X at time t that was not already in Y at time $t - 1$. The information about X transferred from Y to Z at time t is likewise defined as

$$I_T(X; Y_{t-1} \rightarrow Z_t) = I_{\min}(X; Z_t, \{Z_{t-1}, Y_{t-1}\}) - I_{\min}(X; Z_t, Z_{t-1}) \quad (8)$$

where $\{Z_{t-1}, Y_{t-1}\}$ is a vector-valued random variable representing the joint outcome of Z_{t-1} and Y_{t-1} . Information transfer can be interpreted as the information about X shared by Z_t and Y_{t-1} that was not already in Z at time $t-1$.

In our analysis, the time-varying distributions of random variables are estimated over a fine grid of possible values and times using a kernel density estimation technique known as average shifted histograms (Scott, 1985). This discreteness can be interpreted as due to noise that sets a fundamental precision with which we can measure the state of the system. Furthermore, we have verified that the results of our analysis are qualitatively robust over a wide range of grid resolutions (from 20 to 200 bins). Although IT can be applied directly to continuous random variables, there are technical difficulties involved that we do not wish to engage here (Cover & Thomas, 2006).