The problem of $\begin{matrix}\text{multimodal}\\ \text{concurrent}\end{matrix}$ serial order in behavior*

Oren Kolodny
Department of Zoology
Tel Aviv University

Shimon Edelman†
Department of Psychology
Cornell University

July 16, 2015

Abstract

The "problem of serial order in behavior," as formulated and discussed by Lashley (1951), is arguably more pervasive and more profound both than originally stated and than currently appreciated. We spell out two complementary aspects of what we term the generalized problem of behavior: (i) multimodality, stemming from the disparate nature of the sensorimotor variables and processes that underlie behavior, and (ii) concurrency, which reflects the parallel unfolding in time of these processes and of their asynchronous interactions. We illustrate these on a number of examples, with a special focus on language, briefly survey the computational approaches to multimodal concurrency, offer some hypotheses regarding the manner in which brains address it, and discuss some of the broader implications of these as yet unresolved issues for cognitive science.

1 Background and motivation

What does it take for an animal species to survive and flourish in the world? Intuitively, embodied and situated behaving agents that are capable of sensing and acting — a broad category, which includes all animals from yeast to cephalopods, insects, and vertebrates, and even some plant species — must balance the dynamic flow of events arising from their own endogenous motivational and cognitive processes, cues derived from sensory data, and decisions that shape and control the agent's ongoing covert comportment and overt behavior.

In psychology and in neuroscience, behavior is too often implicitly assumed to be reducible to a succession of stimulus/response bouts, a notion that has a counterpart in machine learning and artificial intelligence, where the preoccupation is with the input-output mappings arising from specific problems, as in "object recognition" or "question answering." In a recent review, Edelman (2015b) documented the pervasiveness of the stimulus/response doctrine, noting that its resilience is particularly surprising, given that it had been considered problematic already over a century ago, when John Dewey first offered a critique of

*The title is modified from Lashley (1951).

†The authors' contributions to this paper were multimodal and concurrent; authorship order was determined by a random symmetry-breaking event. OK is presently at the Department of Biology, Stanford University, Stanford, CA 94305, USA. Address correspondence to SE, edelman@cornell.edu

"the reflex arc concept in psychology": "What we have is a circuit, not an arc or broken segment of a circle. [...] The motor response determines the stimulus, just as truly as sensory stimulus determines movement. [...] There is simply a continuously ordered sequence of acts [...]" (Dewey, 1896, p.365).

The view of behavior as dynamically unfolding and serially ordered was championed by Karl Lashley, in a paper delivered at the celebrated Hixon Symposium and published in 1951: "The input is never into a quiescent or static system, but always into a system which is already actively excited and organized" (Lashley, 1951, p.112). In his paper, titled "The Problem of Serial Order in Behavior," Lashley argued that this characterization of behavior is very general:

> Certainly language presents in a most striking form the integrative functions that are characteristic of the cerebral cortex and that reach their highest development in human thought processes. Temporal integration is not found exclusively in language; the coordination of leg movements in insects, the song of birds, the control of trotting and pacing in a gaited horse, the rat running the maze, the architect designing a house, and the carpenter sawing a board present a problem of sequences of action which cannot be explained in terms of successions of external stimuli.

Lashley's insights into the serial nature of behavior have since been thoroughly corroborated (for a review, see, e.g., Rosenbaum, Cohen, Jax, Weiss, and van der Wel, 2007) and incorporated into mainstream cognitive science (Henson and Burgess, 1997; Burgess and Hitch, 2005).

In this paper, we argue that even this, by now classical, view of behavior is, however, limited in that it leaves out two key aspects of the problem of control that all animals must solve:

- The *structural* or synchronic aspect: how to deal with multiple input and/or output variables, specified at a given instant of time. Even the simplest sensorimotor systems must deal with multiple streams of information (e.g., those that arrive from multiple sensors or are sent to multiple actuators), which, moreover, may differ radically in their statistical and other properties (e.g., as in the case of auditory and visual cues). We call this the problem of *multimodality*.

- The *temporal* or diachronic aspect: how to deal with multiple streams of information as they unfold over time. The problem of multimodality is exacerbated by the dynamical nature of the processes, both endogenous and exogenous, that affect/comprise behavior. Not only do those processes unfold in parallel: they generally do so at different rates and independently, or asynchronously, with regard to each other. We call this the problem of *concurrency*.

Together, multimodality and concurrency form what may be called the generalized problem of behavior.[1]

Multimodality receives much attention in the cognitive sciences, where it drives research into cross-modal sensory or sensorimotor integration (e.g., Kersten and Yuille, 2003; Doubell, Skaliora, Baron, and King, 2003; Angelaki, Gu, and DeAngelis, 2009; Fetsch, DeAngelis, and Angelaki, 2013; Chabrol, Arenz,

[1] It may be useful to note that the two aspects of the problem of behavior, structural and temporal, correspond to the two aspects of the so-called credit assignment problem, first formulated by Minsky (1961, p.432). The credit assignment problem came to be regarded as a core concept in artificial intelligence, and, more recently, in reinforcement learning (Sutton and Barto, 1998; Chater, 2009).

Wiechert, Margrie, and DiGregorio, 2015). Interestingly, Lashley too viewed integration as a central function of the brain. Here, we state and motivate a complementary view, according to which integrating across dimensions and modalities is in many cases not possible without losing potentially important information. This suggests that multimodality cannot be approached exclusively through radical dimensionality reduction or integration: for some tasks, the control problem is irreducibly multivariate.

Concurrency, in contrast to multimodality, is more familiar to the designers of parallel asynchronous systems in computer science and robotics than to behavioral scientists, who are only now becoming aware of the issues it involves. In a recent review of the emerging field of computational ethology, Anderson and Perona (2014) note that it "will require simultaneous representations at multiple timescales," and conclude that "Given these complexities, it is not surprising that a general, computationally sound approach to describing behavior using conventional descriptors has not yet emerged, since it is unlikely to be manageable 'by hand'." One of our goals here is to suggest a requisite computational approach, which, moreover, may be amenable to neural implementation.

The rest of this paper is organized as follows. In section 2, we illustrate the problems of multimodality and concurrency on a case study: that of language. These two problems are then discussed in depth in sections 3 and 4. The ensuing conceptual issues are addressed in a computational framework sketched in section 5. Section 6 offers a glimpse of a possible brain basis for this framework. Finally, section 7 summarizes our thesis and mentions some directions for future exploration.

2 A case study: multimodality and concurrency in language

As a means of communication that has been, and still is, co-evolving with embodied agents (Christiansen and Chater, 2008), language in the wild is essentially multimodal and concurrent (Vigliocco, Perniss, and Vinson, 2014; Hilliard, O'Neal, Plumert, and Wagner, 2015). In its "default" spoken form, it appears that voice dominates over gesture and articulation over prosody — an impression that is strengthened by the possibility of capturing much of the meaning of an utterance by transcribing merely the sequence of words that comprise it (a step that Edelman (2008a, sec. 7.2.1) called "going digital"). The amount and the nature of the information that such transcription leaves out may, however, be quite significant; indeed, in some settings, such as impromptu social interactions, ignoring prosody may leave out most of what is important.[2] Moreover, the existence of sign languages and the documented cases of their spontaneous emergence (e.g., Senghas, Kita, and Özyürek, 2004) suggest that gestural and vocal modalities are equally capable of carrying information.

The main modalities comprising a stream of language — the articulated sequence of phonemes, the prosodic dimensions, and the gestures — are all hierarchically structured; this is a matter of consensus in present-day linguistics (see, e.g., Chomsky, 1957; Hockett, 1960; Fodor, Bever, and Garrett, 1974; Langacker, 1987; Phillips, 2003; Lamb, 2004; Culicover and Jackendoff, 2005; the realization that language and other sequential behaviors must be hierarchically structured was one of the main insights of Lashley's 1951 paper). As an utterance unfolds in time, the concurrent flow of information in the different channels

[2]Consider how easy it is to change the meaning of an utterance such as "Yeah, yeah" by adjusting its prosody, as demonstrated by Sidney Morgenbesser, the John Dewey Professor of Philosophy, late of Columbia University (Shatz, 2014).

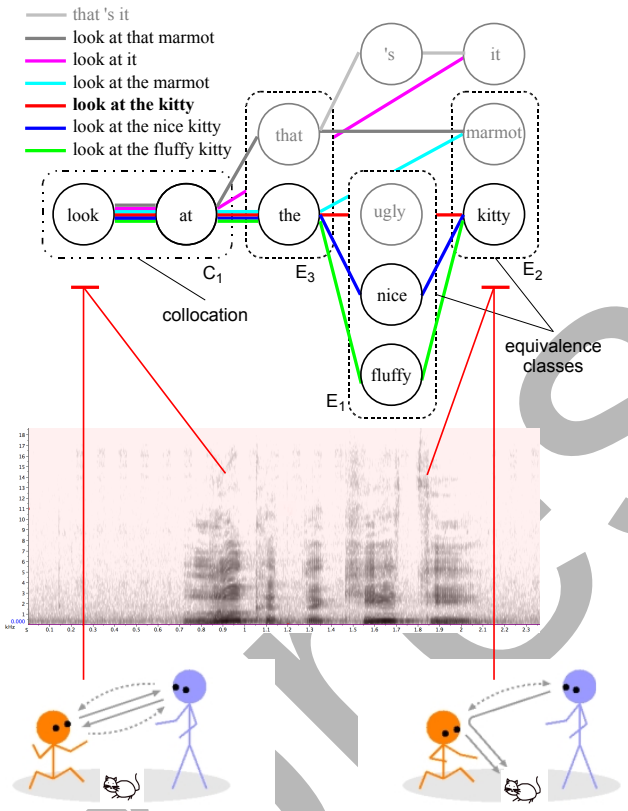

Figure 1: The multimodal, concurrent nature of the "primary linguistic data" (reproduced from Edelman, 2011; cf. Goldstein, Waterfall, Lotem, Halpern, Schwade, Onnis, and Edelman, 2010). *Top:* a graph-like grammar, learned (Solan et al., 2005) from a small corpus of language, consisting of the seven English utterances listed on the left. In building this graph, vertices initially correspond to discrete elements such as phonemes or words, and edges – to transitions between these elements. The graph is then recursively refined by aligning utterances at matching elements and detecting recurring series of elements (collocations); a hierarchical structure emerges when categories (equivalence classes) of vertices are defined, e.g., by grouping together elements that appear in similar contexts. *Middle:* the grammar constructed in this manner leaves out much of the important information in natural language, such as prosody, illustrated here by a spectrogram of one of the utterances ("look at the nice kitty"). Prosodic and other dimensions of language, which in natural discourse appear concurrently with its discrete elements, contain cues that are important in language acquisition and use. *Bottom:* a diagram illustrating some of the social cues, such as timed eye contact and shared attention, which also assist learning (Goldstein et al., 2010; Frank et al., 2013).

is intricately coordinated. During *language acquisition*, this coordination makes learning easier, as, for instance, when shared visual attention between the speaker and the listener helps communicate the referent of a noun (Goldstein, Waterfall, Lotem, Halpern, Schwade, Onnis, and Edelman, 2010; see Figure 1, bottom). Cross-modal coordination also helps the learner grasp and eventually master the hierarchical combinatorial structure of the medium — the structure that makes it possible for language to be both richly expressive (Hockett, 1960) and learnable (Edelman, 2008b). Likewise, at all times during *language use*, multimodal coordination between competent speakers makes it easier for meaning to be shared (Dale, Fusaroli, Duran, and Richardson, 2013).

A number of recent computational modeling efforts achieved some success in learning language in an unsupervised manner exclusively from transcribed speech or written text, unannotated with respect to prosody, gesture, or any other "extralinguistic" cues (van Zaanen, 2000; Adriaans and Vervoort, 2002; Solan et al., 2005; Bod, 2009; Waterfall, Sandbank, Onnis, and Edelman, 2010; Kolodny, Lotem, and Edelman, 2015).[3] In particular, the study described in (Kolodny et al., 2015) aimed not just to attain a grammar for which precision and recall could be measured, but to do so in a biologically inspired architecture and using a realistic, incremental approach to learning, in the hope that the resulting model would replicate a range of psycholinguistic phenomena.

The model of (Kolodny et al., 2015) represents grammar as a directed graph — more precisely, as a higraph, which is a generalization of the familiar graph data structure and which serves as the representational basis of statecharts (Harel, 1988, 2007), more about which in section 5 below). The grammar is initialized as an empty graph. As learning progresses, vertices and edges are added to the graph incrementally, the former representing discrete elements such as phonemes or entire words, and the latter — the observed temporal transitions. At the same time, the graph structure is processed so as to identify and make explicit various types of linguistic structures, such as collocations, substitutability in context, etc. (Figure 2).

This learning process results in a gradual build-up of hierarchical structures that compactly represent the model's experience and that support generalization — that is, the production of novel well-formed utterances — based on similarities among substructures detected during learning. The grammar also supports parsing, or the analysis of new utterances in terms of the existing structures. In addition to being capable of accepting and generating utterances, including some novel ones (an ability that is quantified, respectively, by perplexity or recall and by precision), this model was also shown to replicate certain findings from language acquisition and processing (Kolodny et al., 2015).

Despite this progress, the learning abilities and the formal linguistic performance of the grammars acquired by this and related approaches still fall far short of the human-set standards. In line with (Goldstein et al., 2010), we conjecture that these shortcomings stem to a large extent from two design choices: first, treating language as a single sequence of tokens that bear no relation to the world or to each other, except as members of the sequence, and, second, from pretending that language learning is completely unsupervised.

In reality, of course, linguistic behavior is a bundle of concurrently unfolding, diverse, multidimensional processes, which, moreover, are socially situated and coordinated. In these respects, language is just like any other complex behavior — as indeed stressed by Lashley in his discussion of the problem of serial order. To understand how language and other behaviors like it are learned and used, one must, therefore, take up

[3] Rather than attempting a review of the vast literature on language acquisition, we refer here only to those few implemented systems that were shown to scale up to realistic corpora and to be capable of unsupervised learning of a generative grammar.

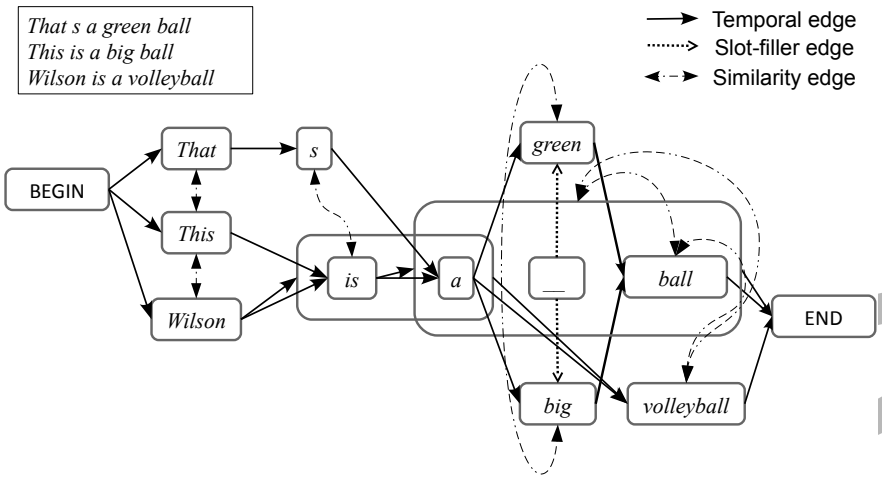

Figure 2: The graph structure used by the U-MILA language acquisition model of (Kolodny et al., 2015) to represent its knowledge of language, or grammar. The example shown here illustrates the patterns derived from the three-sentence corpus listed in the upper left corner. As in Figure 1, one type of pattern is collocation: a sequence of tokens that occurs more often than expected by chance (e.g., "is a"). A collocation may contain a slot (as in "a ___ ball"), which may be filled by any of a class of tokens that are deemed sufficiently similar, hence equivalent, to each other (e.g., "green" and "big"); as in (Solan et al., 2005), a recursive application of this process leads to the emergence of hierarchical structures. Various elements of the approach to language pattern detection described by Kolodny et al. (2015), such as the combination of bottom-up and top-down cues, are applicable to the detection of objects in other modalities (more about which in section 3.3).

the twin problems of multimodality and concurrency, which we do, respectively, in sections 3 and 4.

3 The structural aspect of the problem of behavior: multidimensionality and multimodality

To fully realize how pervasive is the need to deal with multidimensional and multimodal information, consider as an example President Lyndon Johnson's assertion that the then House minority leader (and future President) Gerald Ford was so dumb as to be unable to walk and chew gum at the same time. The first key observation here is the fact that walking, chewing gum, and pretty much any other of the many "simple" behaviors of which even Mr. Ford was presumably capable, even when undertaken separately from the others, is multidimensional.

In particular, mastication involves at least four distinct muscles, which entails that the representation space for the instantaneous state of the gum-chewing motor program is nominally at least four-dimensional. This is because four independent numbers are needed in principle to individuate each such state. The formal concepts on which this observation is based are taken up in section 3.1.

The second key observation is that walking in bipedal vertebrates involves at least two distinct sets of

variables that cannot be mixed: those that control the (dynamic) upright balance of the body and those that control perambulation. Any attempt to "integrate" these variables (e.g., by projecting them onto a common set of dimensions) would result in a pratfall. This issue is further discussed in section 3.2.

3.1 Dimensionality: nominal, effective, and intrinsic

The (multi)dimensional aspect of problems arising in neurobiology and neuroethology stems, on one level, from the simple fact that nervous systems consist of elements whose states (e.g., membrane polarization or spiking rate) are in principle independent of each other. Assuming that neurons are the elements of interest,[4] an n-neuron system thus requires a list of n numbers to represent it, which makes it *nominally* n-dimensional (Edelman, 1999, p.97). The instantaneous state of such a system can be treated as a point in a vector space $\mathbb{R}^n$ — a methodological move that makes the formidable formal machinery of geometry applicable to neurobiology (Mumford, 1994).

In practice, however, interactions among the neurons comprising a system constrain its dynamics so as to exclude large portions of the total n-dimensional volume in which this dynamics resides.[5] The *effective* dimensionality of such a system is lower than its nominal dimensionality, often by a very large margin. This is just as well, because the *intrinsic* (outside-world) dimensionality of sensorimotor tasks that neural systems represent and solve is typically low (or else they would be intractable[6]), and it is the intrinsic dimensionality of the task that an animal should care about.

As an example of intrinsic dimensionality, consider the motor task of reaching out and touching a target object that is in front of you. This task is intrinsically three-dimensional insofar as the target location is completely specified by three numbers in the familiar Cartesian coordinate system. The intrinsic dimensionality of a task is, however, always addressed through the lens of representations, whose composition and "grain" may vary. Thus, the reaching task becomes four-dimensional when approached with a two-joint manipulator that has a ball joint at the shoulder (contributing three degrees of freedom) and an angular joint at the elbow (one more degree of freedom). Furthermore, when considered on the level of dynamical control of the joints (rather than their kinematics), the relevant dimensionality is dictated by the number of muscles involved, or, more appropriately, by the number of independently activated muscle fiber bundles, illustrating the notion of representational grain.

3.2 Multimodality

The multimodality aspect of the problem of behavior is due to the qualitative differences that may exist among the dimensions that define sensory, motor, and, a fortiori, sensorimotor tasks. Its best-known guise is the need for "sensory integration" faced by all animals that are equipped with multiple sensory modalities, such as vision and hearing (e.g., Groh and Werner-Reiss, 2002; Fetsch, DeAngelis, and Angelaki, 2013).

[4] As opposed to the level of consideration being parts of neurons (such as ion channels of which each neuron has many), or perhaps cliques of tightly coupled neurons.

[5] For instance, a system of two neurons that inhibit each other cannot stay for long in the corner of its state space corresponding to both neurons being active; cf. (Edelman, 2008a, p.162).

[6] The intractability of learning and control in high-dimensional spaces is known as the curse of dimensionality (Bellman, 1961); see (Edelman and Intrator, 2002; Edelman, 1999) for detailed discussions.

It is crucial to note that the differences among the various dimensions of representation and control of behavior are not in any sense "given," obvious, or even readily apparent to the system in charge. In particular, from the standpoint of a neuron or a circuit that implements integration, there is no a priori difference among the signals that impinge on it: visual and auditory inputs can only be (and feel) different insofar as the statistics of the signals they carry differ (O'Regan, Myin, and Noë, 2004).[7]

The multimodality problem is, therefore, merely a special case of a structural problem that arises *within* the traditional senses as well, and, generally, in any multidimensional system. It is faced not only by a human being, such as Mr. Ford from the above example, but also by the bacteria in his gut, each of which must deal, in parallel, with stimuli from multiple chemical sensors embedded in its membranes, all the while managing, in parallel, multiple internal biochemical processes and controlling, in parallel, multiple cilia that help it move. Moreover, because all animals that sense their environment also act, the integration in question is always sensorimotor rather than merely sensory (for an overview of the brain mechanisms of sensorimotor integration in vertebrates, see Doubell, Skaliora, Baron, and King, 2003).

Whereas cross-modal and sensorimotor integration is widely studied, the full scope of the problem at its core is rarely, if ever, acknowledged. For one thing, as we just noted, the problem arises also within each of the traditionally defined sensory modalities (e.g., when multiple visual cues, such as shape and texture, are "integrated"; Treisman and Gelade, 1980). We thus have no choice but to admit that the problem springs into being, metaphorically speaking, the moment a hitherto single-sensor animal species evolves a second photoreceptor, hair cell, or whatever, to serve alongside the one it had all along (and if a line-up of two or more sensors appears on the scene all at once, the animal is thereby immediately burdened with the problem of multimodality).

This take on the situation may seem extreme but it is the only tenable starting position for reasoning about a developmentally early, or an evolutionarily nascent, stage of sensorimotor integration, when the processes charged with integration are yet to be calibrated (Philipona et al. (2004) pose and address a related computational problem). Moreover, from a meta-theoretical standpoint, it seems reasonable to demand that the details of the key explanatory concepts arising in this situation — dimension and modality — be determined by computational means from the sensorimotor data, instead of through philosophical analysis (or by trying to combine philosophy with neuroscience, as in, e.g., Keeley, 2002). Note that this consideration applies to the predicaments both of the cognitive scientist and, more importantly, of the developing cognitive system that needs to start making sense of the world it finds itself in, and to do so by computationally effective and reliable means (as judged by the usual evolutionary criteria; Dobzhansky, 1973).

3.3 How many dimensions? How many modalities? How many objects?

When pondering the reduction of dimensionality and the integration of modalities, we face three related questions: How many dimensions of interest are there in my world? Which of these are to be grouped together, and apart from others, in distinct modalities? And how many independent sources of information

[7] As O'Regan et al. (2004, p.87) phrase it, "[...] the quality of a sensory modality does not derive from the particular sensory input channel or neural circuitry involved in that modality, but from the laws of sensorimotor contingency that are involved." Philipona, O'Regan, Nadal, and Coenen (2004) discuss computational methods for extracting useful information from such contingencies.

— which is one way of defining what an "object" is — are out there? (We shall see momentarily why this last question belongs here.)

For the dimensionality question, the default starting point for the inquiry is the nominal dimensionality of the sensorium. Thus, a sensory system whose "front end" consists of, say, two photoreceptors and three hair cells spans a five-dimensional data space through which it perceives the world. From there on, the cognitive system must resort to some combination of computational methods for estimating intrinsic (as opposed to nominal) dimensionality (Camastra, 2003; Braun, Buhmann, and Müller, 2008; with regard specifically to time series, as opposed to merely multivariate, data, see Chen and Müller, 2012).

While a discussion of such methods is beyond the scope of the present paper, it is important to note that most of them require that distances among the data points — samples from the sensorimotor representation stream — be known. Distance (or, equivalently, dissimilarity) is, however, in the eye of the beholder: whether or not two stimuli should be considered more similar to each other than to a third one depends on the consequences that treating them as such would have for the agent (Shepard, 1987). In his foundational paper on the psychology of generalization, Shepard showed how distance between two stimuli in a psychological representation space can be estimated from first-principles topological considerations (namely, from the statistical expectation of the degree of overlap between their "consequential regions"). Of the many methods for dimensionality estimation one should, therefore, prefer those that are rooted in topology over those that make prior assumptions about the metrics of the representation space.

Moving on to the second question, that of the number of distinct modalities among the intrinsic dimensions, we note that here too a purely topological method is desirable. One such method, proposed by Clark (1993), is based on the concept of matching, which is, very appropriately, topological rather than metric. Two stimuli "match" if they are physically distinct yet are conflated by the perceptual system; with this definition in mind, "One can get from red to green by a long series of intermediaries, each matching its neighbors; but no such route links red to G-sharp" (Clark, 1993, pp.140-41).[8]

Finally, the question of the number of objects belongs here because the information pertaining to it is confounded with the information regarding the number of modalities. As noted above, modalities can only be distinguished in multidimensional data on the basis of their distinct topological and statistical signatures. The observed distinctions can, however, be due to the presence of several independent sources of variation — objects — in the data. Natural settings typically contain multiple distinct objects, which register simultaneously in visual, auditory, and olfactory modalities. As a rather extreme example, we may think of a rain-forest scene with a multitude of plant and animal life (with the qualification that for many animal species such scenes are mostly "background," against which very few types of distinct objects, such as conspecifics, predators, or prey, need to be discerned). Computationally, this leads to the problem of source separation (e.g., Cardoso, 1998).

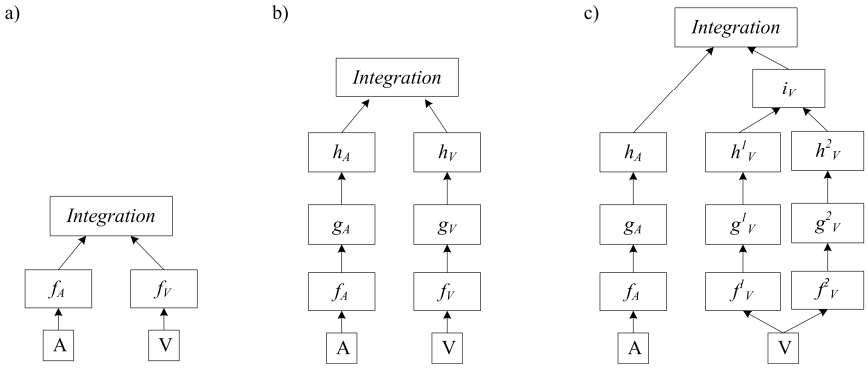

Figure 3: In the literature, sensory or sensorimotor integration is typically depicted as converging to a single arrow/box, implying, perhaps inadvertently, that the resulting space is one-dimensional. In this example, reproduced from (Coen, 2006, fig.6.5), auditory (A) and visual (V) perceptual modalities are processed for a while separately, then integrated, sooner (a) or later (b). In a "hybrid" architecture (c), intermediate within-modality integration stages may be present. Coen (2006) rejects the view that focuses exclusively on the convergence stage, because "perceptual phenomena [...] are complex amalgams of mutually interacting sensory input streams — they are not end-state combinations of unimodal abstractions or features." We concur: the classical view is untenable, if only because neither perception nor action can be boiled down to a single variable (see section 3.5).

3.4 Respects for similarity

Similarity is useful and perhaps indispensable in guiding the detection of, and reasoning with, objects (Shepard, 1987; Tenenbaum and Griffiths, 2001; Edelman and Shahbazi, 2012), as indicated in particular by the popularity of the "nearest neighbor" methods (Cover and Hart, 1967) in machine learning (e.g., Andoni and Indyk, 2008). It is, however, also deeply problematic, because, being a scalar, similarity can only be arrived at following the most drastic possible reduction of dimensionality — from many to one (cf. Figure 3).

To make this happen, the representational system must fix the contributions (in the simplest case, linear weights; cf. the Ugly Duckling Theorem of Watanabe, 1969, p.376) of various relevant dimensions of the stimuli.[9] These contributions are, however, liable to differ between one case and the next, even when the same objects are involved. Tellingly, when psychologists and philosophers complain about the problematicity of similarity, they often focus precisely on the dangers of pretending that the functional demands with which similarity is saddled can be met by a scalar. Thus, for instance, Eisler (1960, p.77) wrote: "An observer instructed to estimate the similarity of, e.g., two differently colored weights is supposed to ask: in what respect?"

A standard approach to addressing this issue is to specify the *respects* under which the hypothesis of similarity is entertained in each case at hand (Medin, Goldstone, and Gentner, 1993; Edelman and Shahbazi,

[8] Keeley (2002) criticizes this approach for resulting in more than the five traditional sensory modalities (e.g., hue and the direction of visual motion end up being distinct). In our view, this speaks to the inadequacy, not of Clark's method, but of the lay (and philosophical) preconceptions regarding the senses.

[9] This is what Kolodny et al. (2015) did in formulating their measure of similarity between elements comprising a grammar.

2012). For instance, when the agent's task calls for inferring a hypothesized common cause behind a set of measurements (say, a visual object that manifests itself in reflectance, stereo, texture, and motion data), the agent may employ Bayesian inference, or an approximation thereof, to estimate the posterior probability of the object's presence (Kersten, Mamassian, and Yuille, 2004). The same instrumental considerations and the same Bayesian techniques apply when the task is sensorimotor integration and control (Körding and Wolpert, 2006) and, more generally, prediction (Clark, 2013). Integration, however, is not always the right goal to pursue, as we argue next.

3.5 The buck stops here

The funneling of the data into a single variable is justified when the "respect" in question is crystal-clear, as, for instance, when two sets of multimodal cues whose similarity is to be estimated are likely to have been generated by the same kind of object or the same kind of event (i.e., motor program). Even in these cases, however, integration typically involves information loss, e.g., when a decision criterion such as Maximum A Posteriori (MAP) likelihood is applied to the posterior distribution. In general, if carried out in the absence of a well-defined task, dimensionality reduction or integration violates Marr's Principle of Least Commitment, according to which an information processing system should postpone as long as possible undertaking actions that cannot be effectively undone (Marr, 1976, p.485). Moreover, dimensions that do not belong together (as in the case of balance- and movement-related degrees of freedom in walking, mentioned earlier) must never be integrated.

In light of these observations, and seeing that multi-modality/dimensionality is an inherent aspect of the world that animals confront, it seems that the best a cognitive agent can do is employ some means of dimensionality reduction to reduce the complexity of its sensorium, without yet "going all the way." The goal of the reduction should be, generically, not to boil the information down to a single dimension, but rather to discover the typically few intrinsic dimensions of interest in the typically high-dimensional data set (Edelman, 1999; O'Regan et al., 2004; Philipona and O'Regan, 2010).

Because the dimensions of the resulting maximally reduced representation will be mutually irreconcilable,[10] they would have to be used — for instance, mapped onto actions by Bayesian or other inference mechanisms — as they are. This too is just as well: as we just reiterated, the dimensions of motor control in a realistic embodied cognitive system are no more reducible to a single scalar than the dimensions of perception. Synchronically, or instantaneously, therefore, **the problem of behavior is many-to-many-dimensional**.

4 The temporal aspect of the problem of behavior: concurrency

What happens diachronically, when time is allowed to roll? Some types of dynamical systems can be effectively characterized when observed repeatedly (sequentially) over time through the "window" of a single variable, according to a theorem proved by Takens (1981). Even in this case, however, the use of multiple parallel measurements leads to much more effective inference: Deyle and Sugihara (2011), whose

[10]The sense in which some dimensions of a representation space may be irreconcilable or incommensurable is related to the distinction between integral and separable dimensions in psychology (Garner and Felfoldy, 1970).

approach generalizes the Takens theorem, note that it works best when applied "to a wide variety of natural systems having *parallel time series* observations for variables believed to be related to the same dynamic manifold" (our italics). Our argument for the need for many-to-many-dimensional mappings in controlling behavior holds, therefore, even as the representations unfold over time. Moreover, as we shall see this section, the constituent dimensions of such representations — the data in the parallel streams — need not be synchronized, which leads to more complications.

4.1 Physical underpinnings of asynchrony among concurrent processes

In addition to there being multiple processes, both in the world and in the animal's brain, that unfold and must be dealt with in parallel, these processes are typically asynchronous, for deep physical reasons. The ubiquity of asynchrony in systems that require process coordination is due to a combination of two factors: (1) the spatially distributed nature of all physical systems and (2) the finite maximum speed with which information can propagate, the speed of light in vacuum. In this sense, asynchrony is a corollary of special relativity (Edelman and Fekete, 2012, p.83). Enforcing synchrony in an electrical circuit, such as a logical gate array, through the use of a global clock only works because the relevant physical dimensions of a typical circuit are much smaller than the distance traveled by light between clock ticks. In a neural circuit, however, the signal propagation speed relative to the distances involved is too low to be assumed infinite (Izhikevich, 2006), which makes neural computation asynchronous across different signal propagation paths and localities.

Neural processes driven by quasi-global periodic signals (such as respiration driving olfaction) or by a central pattern generator (CPG) are asynchronous too, when scrutinized at a fine enough grain. For instance, while the neural computation in the rabbit olfactory bulb is time-locked to sniffing (Kepecs, Uchida, and Mainen, 2006), each aspiration merely sets off a number of neural processes, which then proceed and wind down at their own pace until the next cycle is initiated. Such subordinate processes may be not only not synchronous, but (deterministically) chaotic (Skarda and Freeman, 1987). Moreover, the inner functioning of CPGs, such as the pyloric and gastric mill circuits in the lobster (Selverston, 2008), is asynchronous, which helps make the timing controllable from the outside, e.g., through dopamine modulation (Harris-Warrick, Coniglio, Levin, Gueron, and Guckenheimer, 1995).

In vertebrates, a similar arrangement is found, on a finer temporal scale, in the distributed neural activity driven by cortical rhythms such as gamma and theta, where functionally significant local phase lead/lag "details" superimpose on global temporal periodicity (Buzsáki and Diba, 2010). Indeed, on one account, the synchronous cortical activity in itself has no functional role in cognitive computation, serving instead as an "infrastructural" mechanism for balancing inhibition and excitation (Merker, 2013a).

4.2 The computational challenge of asynchronous concurrency

Stepping back from specifically neural computation, we note that the problem of asynchronous coordination looms large in any parallel computation framework, such as the one sketched in Figure 4, right, which can be taken to depict the spread of activation in a network of diverging and converging concurrent processes.[11]

[11] The paper from which we adapted this illustration (Briegel, 2012) assumes that activation flows out of a node along one and only one of the several possible connections, but of course, as we pointed out in section 4, this is in general not the case.

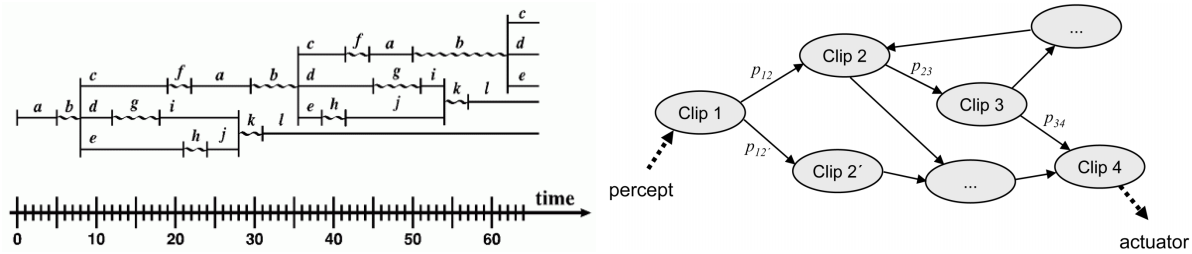

Figure 4: Asynchronous concurrent processes. *Left:* This diagram depicts a system that is in a superposition of *states* (solid horizontal lines), with state *transitions* (vertical lines) driven by *events* (wavy horizontal lines). Adapted from (Sowa, 2000, ch.4, fig.2), where the following definition is offered: "Processes can be described by their starting and stopping points and by the kinds of changes that take place in between. [. . .] In a continuous process, which is the normal kind of physical process, incremental changes take place continuously. In a discrete process, which is typical of computer programs or idealized approximations to physical processes, changes occur in discrete steps called *events*, which are interleaved with periods of inactivity called *states*." *Right:* A more general case, in which there is no central clock. Adapted from (Briegel, 2012, fig.1).

Letting activation propagate unchecked in such an architecture is a recipe for a timing disaster of the same kind that was visited upon Shakespeare's Romeo and Juliet. As the reader will recall, in the play, Friar Laurence set up a seemingly straightforward scheme for orchestrating a happy end to the love story that he was witnessing, a scheme whose success depended critically on the need to coordinate just two concurrent processes.

In the play, a single message miscarried; the relative timing of the two processes went awry; each of the two lovers committed an irreversible act; and the outcome —

> A glooming peace this morning with it brings;
> The sun, for sorrow, will not show his head [. . .]

In the theory of parallel computation, there is a coordination mechanism that could have prevented this tragedy, namely, *guarded command*: an action whose release is predicated on the prior fulfillment of a certain condition — the guard (see (de Bakker and Zucker, 1982) for the formal semantics of such constructs). Specifically, Friar Laurence should have instructed Juliet to go ahead with the sleeping drug scheme only after it becomes certain that Romeo has been informed of it.[12] The key concept in this approach is that of *event*.

[12] In a truly distributed system, where communication between processes, as between Verona and Mantua, is not entirely reliable and involves delays, such coordination is generally intractable (see the discussion and the references in Edelman and Fekete, 2012, section 4.1). Fortunately, the formalism advocated below circumvents this problem.

5 An event-based computational approach to multimodality and concurrency

In this section, we propose that the problem of behavior be considered, both in its structural and temporal aspects, as the coordination of a bundle of concurrent, asynchronous *processes*, interspersed with *events*, as defined and illustrated in Figure 4, left.

5.1 Events as an organizing principle in multimodal, concurrent systems

The concept of event has been discussed extensively in philosophy (Davidson, 1980; see Casati and Varzi, 2014 for a detailed treatment and many references), psychology (Zacks and Tversky, 2001; McAuley, Jones, Holub, Johnston, and Miller, 2006; Zacks, Speer, Swallow, Braver, and Reynolds, 2007; DuBrow and Davachi, 2013; Rubin and Umanath, 2015; Reimer, Radvansky, Lorsbach, and Armendarez, 2015), neuroscience (Damasio, 1989; Eichenbaum, Otto, and Cohen, 1994; Wallenstein, Eichenbaum, and Hasselmo, 1998; Zacks, Braver, Sheridan, Donaldson, Snyder, Ollinger, Buckner, and Raichle, 2001; Fortin, Agster, and Eichenbaum, 2002; Schendan, Searl, Melrose, and Stern, 2003; Paz, Gelbard-Sagiv, Mukamel, Harel, Malach, and Fried, 2010; Allen, Morris, Mattfeld, Stark, and Fortin, 2014), as well as AI and computational linguistics (Park and Aggarwal, 2004; Chambers and Jurafsky, 2008; Elman, 2009; Mehlmann and André, 2012). Here, we propose that events are the key to managing both multimodality and concurrency. A sensorimotor event can be thought of as a "narrowing" or causal nexus at which a bundle of perception and action processes comes together for a while in space and time, before pulling apart again (cf. Spivey, 2006, fig.12-2). Within this conceptual framework, a visual object (say) is a kind of event (the coming into view of a part of the visual world, brought about by a shift of gaze; cf. Dewey, 1896, p.358) and so is, of course, a motor act.[13]

To illustrate the usefulness of the concepts of process and event, consider the task of visually guided grasping in primates. It begins as the gaze disengages from whatever it is that the animal is fixating and lands on the object of interest (we nearly always make a saccade to the place where our hands or feet will be going next). As other body parts start to move in turn, fingers gradually open and the arm rotates, pre-shaping the hand and pre-positioning it for the final approach (it is the sequential instead of concurrent execution of these steps that makes the movement of old-fashioned cinematic robots look so robotic). Finally, the hand makes contact, while the eyes move on and the brain is already a couple of hundred milliseconds into the next stage of motor planning. It is only through the abstract notion of the grasp procedure and the event in which it culminates that the relative timing of the multimodal neural and mechanical processes involved can be understood. It stands to reason, therefore, that events should play a central role in making sense of experience and in the (concurrent) planning and execution of behavior.

Very few of the events that unfold in the system comprising an embodied and environmentally situated brain are global in the sense that they involve (and synchronize) *all* the relevant processes[14] (cf. Figure 4, left). Thus, generally speaking concurrency and multimodality exist and must be dealt with even on the level of events, just as they exist and must be dealt with, recursively, within events. However, a narrower-scope

[13] Interestingly, Hurford (2003) called objects "slow events"; cf. (Edelman, 2008a, p.33).

[14] Death is one example that comes to mind.

situation, such as a task that requires sequential *decisions*,[15] can probably be approximated reasonably well by a single thread of events. In this connection, Edelman (2015b) notes that a temporally extended task that involves repeated "crisp" decisions, each of which requires that any distributed representation or superposition of states be collapsed, is effectively serially local. In other words, the representations behind such a task cannot remain probabilistic and distributed at all times: every now and then, the probability distribution must be collapsed.[16]

Presumably, it is this, relatively high-level and discretized view of behavior that prompted Lashley to formulate his "problem of serial order." So as not to lose sight of the problem of essential concurrency, we stress that Lashley's discussion skirts it entirely — he writes as if a horse's step, to pick just one of his examples, is a unit; a link in a chain, rather than a woven rope that consists of many threads, corresponding to the concurrent activations of many muscles, etc.

Modern formulations of sensorimotor control in behavior likewise tend to skirt the problems of concurrency and asynchronous coordination.[17] For instance, the excellent treatment by Coen (2006) of imitation-based learning of birdsong takes the song's representation to consist, both on the perceptual and on the motor side, of a simple sequence of units ("songemes"), each obtained by categorical clustering in the appropriate multidimensional space. The all-important social aspect of birdsong is, however, often multimodal and asynchronous. One of the many examples here is the interplay between male song (an acoustic signal) and female wing stroke (a visual signal) in cowbirds (Gros-Louis, White, King, and West, 2003).

The importance, in particular, of the relative timing between modalities is illustrated by cases such as that of the female túngara frogs, which are more strongly attracted to a male call if in addition to hearing it they also see the male's vocal sac — but not if the visual cue is shifted in time relative to the natural delay, as would happen in the wild when the male that is being looked at is not the same one that is being heard (Taylor and Ryan, 2013). A game-theoretic analysis carried out by Wilson, Dean, and Higham (2013) suggests that such synergy among multiple modalities promotes the evolution of multimodal communication, which may otherwise be too costly to sustain. Not surprisingly, in primates too the processing of audio-visual signals turns out to be sensitive to the asynchrony between modalities (Perrodin, Kayser, Logothetis, and Petkov, 2015).

Finally, in human cognition, a single-threaded approach to behavior is clearly inadequate for most types of music, nor, as we argued in section 2, does it work for language, where prosody, gaze control, and gesticulation are integral to its realistic use in social situations (cf. Figure 1).

5.2 Computational tools for dealing with concurrency

The best-known or most-intuitive computational tools at the disposal of cognitive scientists and ethologists are not necessarily well-suited for dealing either with multimodality or with concurrency. With regard to the former, we saw earlier that a multidimensional state space indiscriminately lumps together dimensions that

[15] For a construal of decision as a type of event in a dynamical system that describes sequential behavior, see (Rabinovich, Huerta, Varona, and Afraimovich, 2008).

[16] An example of a serially local task offered by Edelman (2015b) is glade skiing, where the skier must every now and then decide whether to pass to the left or to the right of the next tree down the glade.

[17] In a purely sensory setting, the coordination problem has been discussed by Eagleman (2010), in the context of conscious visual awareness.

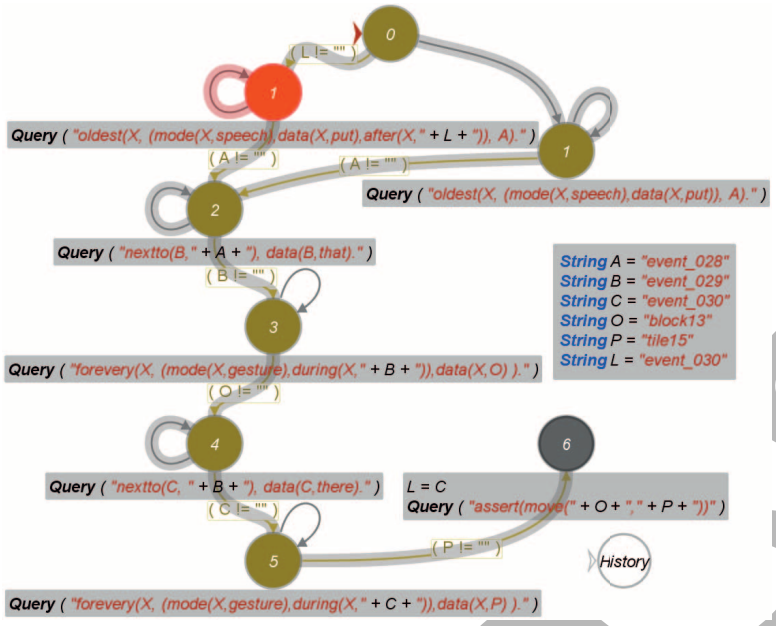

Figure 5: The time course of incremental parsing of Bolt's (1980) "put-that-there" command using the multimodal event logic formalism of (Mehlmann and André, 2012). Note that concurrent processes are in play and that the state transitions are "guarded" by logical preconditions.

may be incommensurable. With regard to the latter, the main challenge is how to coordinate multiple concurrent processes without resorting to a special "central executive" process or at least to a centralized, global clock — options that seem less relevant to biological information processing than layered and subsumption architectures that implement distributed control (Prescott, Redgrave, and Gurney, 1999; Brooks, 1989).

In computer science, concurrent programming languages built around the concepts of processes and events have been around for a long time (Diaz, Muñoz-Caro, and Niño, 2012). Some of these support asynchronous parallel processing and offer tools, such the guarded command mentioned in section 4.2, for imposing logical conditions on execution, so as to allow inter-process coordination.[18] Clearly, concurrency can be formalized and operationalized without positing a single, central "master" process that would alone be in charge of integration.

In recent years, the growing demand for multiplayer gaming and social computing has spurred the development of networked and cloud-based architectures and has renewed the interest in parallel asynchronous computation. This, in turn, has resulted in a better appreciation for the full range of challenges arising from what we termed the generalized problem of behavior, including both its multimodality and concurrency aspects. Entire conferences are now devoted to these challenges, such as the 16th ACM International Conference on Multimodal Interaction (ICMI), held in 2014. Our goal here is not to describe the state of the art in theory and practice of parallel asynchronous computation, but rather to facilitate the forging of conceptual

[18]One example of such a language is Concurrent Prolog, used by Edelman (1987) to implement a distributed algorithm for determining topological connectivity in image processing.

links between that work and the cognitive and brain sciences.

For the cognitive scientist setting out to explore the computer science literature on concurrency and multimodality, one useful entry point is the recent paper on *multimodal event logic* (Mehlmann and André, 2012), which aims "to express structural and functional constraints for the unification of partial information distributed over events from multiple devices and modalities" (Figure 5). It is interesting to note that this approach uses a graph-based formalism (cf. Figures 1 and 2) and that it relies heavily on the concept of event.

5.3 Graphical formalisms for language and other types of concurrent processes

Our experience with developing models of language (Solan et al., 2005; Kolodny et al., 2015) suggests that a graph-based approach to representation is the right one here. To fully exploit the potential of this approach, while meeting the constraints that we argued for earlier, the graph grammar must be made to include the multimodal aspects of natural linguistic input, while tolerating — or, even better, putting to good use — the asynchronous concurrent unfolding of the different modalities. Of course, this consideration applies not only to language, but also to the generalized problem of behavior, which is the broader concern of the present paper.

One way of making concurrent use of multiple modalities would be to treat one modality as dominant and the others as subservient to it. In language, seeing how much structure can be learned from bare text alone (Solan et al., 2005; Kolodny et al., 2015), it seems natural to assume that the sequence of phonemes (heard or transcribed and read) constitutes the dominant modality, with prosodic, gestural, and other additional cues helping the learner "annotate" the main sequence by segmenting significant chunks, building up hierarchical representations, and inferring their meaning. The model illustrated in Figure 2, if extended along these lines, would involve learning multimodal collocations, etc. The other extreme would be to keep the modalities separate, conduct within each the kind of pattern extraction just mentioned, and coordinate the resulting representations when possible, perhaps on a need basis. The need could be signaled by a particular type of event (as in, for instance, reaching out and grasping an object that is being spoken of, an act that immediately reduces the uncertainty of reference and of inter-modality coordination). Finally, as a fall-back option suitable for categorization tasks, different modalities can be used to infer an integrated representation (Coen, 2006) or an amodal one (Yildirim and Jacobs, 2015).

Given how general the problem is, which approach would work better depends on the species and the circumstances of the learner. For instance, the behavior of animals whose primary habitat is an open field is likely to be predominantly visually guided, while communication and coordination out of line of sight is expected to make heavier use of the acoustic modality (cf. the "Buena Vista Sensing Club" of MacIver, 2009, sec. 2.2). A "generalist" species such as our own would do well to avoid an exclusive commitment to a single modality, and indeed while some behavioral findings suggest that the auditory modality dominates in human perception of temporal sequencing (Guttman, Gilroy, and Blake, 2005), other evidence points toward a more balanced and perhaps hierarchically staged integration (Keele, Ivry, Mayr, Hazeltine, and Heuer, 2003; van Wassenhove, 2009; Danz, 2011).

As we stressed in section 3.4, any such integration of modalities needs to do the right thing with regard to dealing with multimodal similarity among objects, so as to support informed categorization and gener-

alization. Kolodny et al. (2015, sec. 2.5) define the similarity between two elements of the grammar (that is, vertices in the graph; cf. Figure 2) as a weighted average of three components: proximity between their edge weight vectors, probability of occurring in the same collocation slot, and within-slot interchangeability within a short time window. The problem of choosing the optimal weights — or rather of having to commit to a particular choice of weights, as discussed earlier — is vexing enough; if the elements are made to be multimodal, it would only be exacerbated (notwithstanding the availability of powerful clustering algorithms for multimodal sequential data, such as those of Coen, 2006 or Ghassempour, Girosi, and Maeder, 2014).

In software engineering, there is an existing formalism that seems well-suited for representing multimodal concurrent patterns, without necessarily clumping them together on the basis of some fixed criterion, is Live Sequence Charts (LSC) (Damm and Harel, 2001). The LSC framework includes both formal semantics and software development tools and is based on "multi-modal scenarios, each corresponding to an individual requirement, specifying what can, must, or may not happen following certain sequences of events" (Harel, Marron, and Weiss, 2012). LSC is a dual formalism to that of statecharts (Harel, 1988, 2007), which came up as a candidate formalism in our earlier work (Edelman, 2011; Goldstein et al., 2010; Kolodny et al., 2015; Edelman, 2015b). Whereas statecharts focus on states that a system can occupy (including a superposition or Cartesian product of several states at once), Live Sequence Charts focus on processes and events. As such, LSC seems even better suited to modeling brain and behavior than statecharts.

To construct a biologically relevant model of this type, we need, however, first to form some reasonably specific hypotheses regarding the brain counterparts to the various intuitive and formal concepts that came up so far in this paper. The next section lists some of the initial conditions for this effort.

6 A brain basis for multimodal, concurrent serial behavior

Most of the literature on the brain basis of multimodal information processing deals with multisensory integration — an extremely active research area.[19] Instead of attempting to survey it here, we shall mention some of the methodological trends that strike us as characteristic of the work in this area, while staying in touch with the analysis of the generalized problem of behavior offered earlier.

Perhaps not surprisingly, many of the available papers on sensory integration concentrate each on addressing a specific type of perceptual problem or behavioral task — a setting in which, as we discussed in section 3, combining data from multiple sensors and/or modalities into a single decision variable is justified. Alongside computational modeling (much of it Bayesian), such studies investigate the anatomy and the physiology of convergence of different cues onto multisensory neurons (Angelaki et al., 2009; Fetsch et al., 2013; Chabrol et al., 2015).

Certain modeling efforts concerned with integration do look beyond the level of single neurons to the dynamics of entire circuits. These, however, often make problematic choices both with regard to the interpretation of the behavioral function whose modeling is attempted and with regard to their architecture. To single out just one broader-scope study, consider, for instance, the ambitious and commendably computationally explicit model of temporal integration in working memory, described by Fuster and Bressler (2012). Among its assumptions are the reducibility of behavior to the perception/action cycle and an undue focus

[19] A Google Scholar search for the conjunction of the terms "multisensory integration" and "brain" conducted in May 2015 yielded about $5,300$ publications dated 2011 or later.

on the isocortex at the expense of other brain structures and circuits (fig.5). Moreover, this model uses a uniform all-to-all connectivity in its proposed working memory circuit (fig.4). These design choices, we feel, need to be revisited and scrutinized (see section 1 and Edelman, 2015b).

As a way of setting possible directions for future modeling work, and with the general problem of behavior in mind, we propose to consider the actual circuitry of the (vertebrate) brain not merely as an inspiration but rather as an extensive and complicated body of findings that need to be related to each other and to the relevant functional and computational thinking. Consequently, in the three subsections that follow, we touch upon the possible brain mechanisms that subserve, respectively (i) cue *convergence* and the emergence of spatially anchored objects; (ii) *binding* of objects and their contexts into episodes, or events, and the spatiotemporal *sequencing* of these; and (iii) *coordination and switching* of sequences for behavioral control.[20]

6.1 The superior colliculus: modal/dimensional convergence and spatial anchoring of objects

In the sense that the convergence of certain cues *defines* an object,[21] it is the convergence of pathways carrying information from different modalities that one should be looking for in the brain. This is indeed what modelers do when, like Fuster and Bressler (2012), they study integration in the prefrontal cortex. Most vertebrate species, however, manage very well without isocortex, which suggests that there must be an evolutionarily older structure in the brain that supports object- and action-based integration. Moreover, given that any action is necessarily anchored spatially with reference to the actor (a threat looming from *here*; an escape route leading *there*), we can expect that the structure in question should also directly encode space, in a coordinate system centered on the self. An old, vertebrate-universal structure with the requisite characteristics is found in the midbrain, in the superior colliculus or SC (Meredith and Stein, 1986; Doubell et al., 2003; May, 2006).

In SC, spatial direction — the glue that holds together the features of an object or an action — is represented, in adjacent layers of neural tissue arranged in spatial register, in the visual and auditory modalities, with likewise spatially coded motivational and motor representations residing in nearby brain areas (Merker, 2007, 2013b). This anchoring in space resolves the "binding problem" that plagues corticocentric approaches to multimodal integration, while at the same time explaining the spatial aspect of the first-person experience (Metzinger, 2003; Merker, 2007; see Edelman, 2008a, ch.9 for a synthesis and an extensive discussion).

The special role of SC in bundling together features into objects (and in giving rise to the subject) should not be taken to mean that it is the exclusive location of such information in the brain. The SC is unique in being the last station in the oculomotor pathway where control is exerted before motor commands are issued. And yet, gaze direction is represented, in addition to SC, in the parietal and frontal cortical areas (namely, in intraparietal sulcus and frontal eye fields, respectively). Thus, to come to grips with the generalized problem of behavior, we must try to understand how SC fits into this broader context, both neuroanatomically and

[20] In singling out the three brain areas discussed below, we imply neither that each of these structures is the only one supporting the stated function, nor that its role is exclusively to support the stated function.

[21] That is, in the sense of Hume (1740, IV, 6), who held this view not only with regard to the objects of perception, but also to the perceiver/self, who is "[...] nothing but a bundle or collection of different perceptions."

functionally. Specifically, we need to understand how feature bundles come to persist over time and enter memory.

6.2 The hippocampus: binding and temporal sequencing of events in episodic memory

Compared to rich representations supported by the cortex, the sensorimotor representations of objects, events, and actions in SC are rather minimalist, focusing mostly on spatial direction. While they do serve as an essential basis for behavior in all vertebrates, in mammals and birds these representations are augmented with information mediated, respectively, by the isocortex and by the corresponding parts of the pallium. This information includes hierarchically structured sensorimotor representations, carried by occipito-temporal and parietal pathways, and representations supporting planning and problem solving, likewise hierarchical, in the frontal lobe (see Edelman, 2015b and the many references therein for a review). The allocortical structure where these pathways converge is the hippocampus (Buzsáki, 1996; Merker, 2004).

Whereas in SC the representation encodes space by being arranged topographically with respect to spatial direction, in the hippocampus the spatial information is effectively (if not literally) map-like. The basic functional unit of hippocampal representations is an *episode*: a bundle of sensory and other memories pegged to a specific time and location in the environment (Eichenbaum et al., 1994; Wood, Dudchenko, Robitsek, and Eichenbaum, 2000; Eichenbaum, MacDonald, and Kraus, 2014). These representations are supported by ensembles of "place cells" (better called "episode cells") which respond selectively to various aspects of the animal's experience associated with the given location in space, including just being there. In this sense, episode cells are just the kind of neural basis of object/event/episode representation expected from the discussion in section 5.1.

In the hippocampus, sequences of place-related episodes (events) are encoded (Fortin et al., 2002) and later (in particular during sleep; Wilson and McNaughton, 1994) replayed, as demonstrated in rat foraging and exploration behavior (Davidson, Kloosterman, and Wilson, 2009), birdsong (Dave and Margoliash, 2000), and human event timing (Ekstrom, Copara, Isham, Wang, and Yonelinas, 2011). Interestingly, the replay can be speeded or time-reversed relative to the original sequence (Davidson et al., 2009); it can also precede action rather than recalling it (Muller and Kubie, 1989; Dragoi and Tonegawa, 2013). As expected, the hippocampus activity is coordinated with that of other brain areas, such as visual cortex (Ji and Wilson, 2007) and prefrontal cortex (Peyrache, Khamassi, Benchenane, Wiener, and Battaglia, 2009).

Given the "critical role" of the hippocampus in memory for sequences of events (Fortin et al., 2002), it makes sense that it is critical also for language acquisition and use, as indicated by a growing list of findings (DeLong and Heinz, 1997; Breitenstein, Jansen, Deppe, Foerster, Sommer, Wolbers, and Knecht, 2005; Duff and Brown-Schmidt, 2012; Kurczek, Brown-Schmidt, and Duff, 2013). Because language is a paradigmatic case of the generalized problem of behavior, we take these findings as motivation for taking a closer look at the role of the hippocampus in experiencing, remembering, and planning sequential behaviors. We next consider how the brain might be learning to coordinate sequences of events.

6.3 The basal ganglia: sequence learning, coordination, and switching

Dealing with information streams that comprise an embodied and situated linguistic exchange requires an occasional synchronization of concurrent neural processes. Mechanisms capable of such synchronization

are found in the set of subcortical structures referred to collectively as the basal ganglia (see Atallah, Frank, and O'Reilly, 2004 for a review that stresses the computational functions of basal ganglia and sketches their connectivity with the rest of the brain).

The "backbone" of these mechanisms is a loop that connects the frontal cortex to the striatum, the striatum to the complex consisting of the internal globus pallidus and the substantia nigra pars reticulata (both directly and indirectly, via an "inverting" inhibitory relay in the external globus pallidus), on to the thalamus, and back to the frontal cortex. The multiple parallel pathways comprising this loop are highly specific in their patterns of connections along the way. These connections link them to additional brain areas, both cortical (the entire isocortex is mapped in an orderly fashion onto the striatum) and others — notably, the cerebellum, the superior colliculus, and the hippocampus (a very brief overview of these circuits, along with further references, can be found in Edelman, 2015b).

Of particular interest to us here is evidence for the involvement of the basal ganglia in learning in humans (Seger, 2006). Studies too numerous to be listed here have documented the central role of the basal ganglia circuits in the coding of multimodal sequences, both perceptual and motor, and in switching between motor programs (for a tiny sample of the literature, see Nakahara, Doya, and Hikosaka, 2001; Bullock, 2004; Aldridge and Berridge, 2003; Kotz, Schwartze, and Schmidt-Kassow, 2009; Jin, Tecuapetla, and Costa, 2014). Most interestingly, some of the neurons in the striatum, for instance, seem to respond to events, such as the onset and the termination of a movement phase relative to other movements (e.g., Aldridge and Berridge, 1998; Atallah, McCool, Howe, and Graybiel, 2014). Finally, the circuits connecting the basal ganglia to the cortex have been posited to play a role in language (Ullman, 2001; Lieberman, 2002) and have helped tie theories of birdsong to theories of language (Bolhuis, Okanoya, and Scharff, 2010)).

7 Discussion

As must be the case with any attempt to come to grips with a problem as broad as the generalized problem of behavior, and to bring to bear on it such a wide range of literature, this paper is necessarily both selective and superficial. Our hope is that it can at least serve as a starting point for further investigations. In this closing section, we list some issues to which we believe attention should be directed first and mention one point of contact between the generalized problem of behavior and another topic of foremost concern in psychology: consciousness.

To recapitulate, our main thesis is twofold: (i) certain aspects of the world (which includes the embodied agent situated in it) cannot and should not be integrated and must therefore be represented as distinct; (ii) over time, the resulting representations give rise to concurrent processes, which, moreover, unfold asynchronously rather than in lockstep. This thesis has implications for the project of understanding the brain basis of behavior, and in particular for the choice of a formal computational framework that would enable such understanding, as well as for behavioral neuroscience and experimental psychology.

7.1 Suggestions for empirical studies and modeling efforts

To supplement the questions raised in section 6, we list here some predictions and possible specific targets for empirical inquiry and computational modeling.

Integration and convergence. The limitations of integration-based approaches to multimodality need to be further explored. One notion to examine here is that even on the highest levels of processing in the frontal cortex, distributed representations are the rule, while commitment — a serial bottleneck — is postponed until as late as possible (cf. Rigotti, Barak, Warden, Wang, Daw, Miller, and Fusi, 2013).

Events. On the behavioral level, much evidence (cited elsewhere in this paper) already exists for the key role of what the subject construes as events in orchestrating actions and coordinating their components across perceptual and motor dimensions. On the neurophysiological level, we stress the need to seek neural representations of events. We predict that these would be found in multiple parallel recordings (e.g., Salazar, Dotson, Bressler, and Gray, 2012; Deco, Tononi, Boly, and Kringelbach, 2015) in the form of the functional building blocks familiar to us from theory of parallel computing. Specifically, some processes — that is, signals traveling down a particular pathway — should be waiting on others that unfold in parallel but asynchronously, for purposes of coordination. Furthermore, we predict that neural implementations of guard clauses (another concept from parallel computing, which we mentioned earlier) would be found. These can take the form, for instance, of neurons firing in a sustained fashion until some precondition is fulfilled. A related existing finding here is that of "start/stop activity signaling sequence parsing" by Jin et al. (2014); cf. (Jin and Costa, 2010).

Specific brain structures and issues arising from their study. With regard to the *superior colliculus*, as noted earlier, we need to understand how feature bundles that exist in SC in form of transient multi-laminar activity foci come to persist over time and enter memory. With regard to the *hippocampus*, one needs to take a closer look at its role in experiencing, remembering, and planning sequential behaviors (see (Edelman, 2015a) for a full review of the relevant literature and for some specific proposals inspired by prior work and by the emerging theoretical synthesis). With regard to the *basal ganglia*, whose role in action selection and sequential behavior has been extensively studied, many key questions remain open. Among these are (i) the mechanism whereby BG contribute to implementing the "serial bottleneck" mentioned earlier in the paper; (ii) the functional significance of going once around the cortico-BG loop,[22] and (iii) the representation of events in BG; cf. the reference to (Jin et al., 2014) above.

Developing a comprehensive approach to computational modeling of behavior. On a more abstract, computational level, the goal should be to connect the neurobiological speculations offered above to formal computational concepts from section 5. Ideally, this should result in an explicit circuit-and-synapse model that would be, on the one hand, as computationally tractable as the LSC formalism, and, on the other hand, as specific as the wiring diagrams that have been available for decades for the admittedly smaller-scope and structurally much more regular circuits in the CA1-CA3 region of the hippocampus and in the parallel fiber system of the cerebellum. This model may end up using (connection) space to represent time and sequential order, as envisaged by Lashley (1951, p.128), as it is the case in the graph-based approaches to language (Solan et al., 2005; Kolodny et al., 2015). Alternatively, it may turn out that the best way of representing temporal order is to "fold" it into a recurrent network, as per (Levy, 1996; Levy, Hocking, and Wu, 2005).

[22] In particular, it is unclear whether or not a single traversal of the loop corresponds to the execution of a single step in a sequence of steps comprising serial behavior (Edelman, 2015a).

A detailed treatment of those questions is forthcoming (Edelman, 2015a). The next step would be to use the computational model in trying to understand specific multimodal sequential behaviors, notably natural language learning and processing. The goal here should be not merely to replicate psycholinguistic findings, but to do so with a generative model capable of both producing and accepting utterances and of scaling up to realistic corpora and situations (Waterfall et al., 2010). Finally, the resulting model should replicate also those of the many "quirks" of language documented by linguists with regard to whose importance there is a broad consensus in psycholinguistics (e.g., the so-called "island extraction" constraints; Sprouse, Wagers, and Phillips, 2012; Stabler, 2013).

7.2 Some broader implications

It may be interesting to explore the repercussions of the proposed theory of behavior for theories of phenomenal experience, in particular for the so-called unity of consciousness (Bayne and Chalmers, 2003).[23] Philosophical debates surrounding this notion take as a starting point the intuitively obvious and incontrovertible unity of the stream of the first-person phenomenal experience — barring rare pathological conditions, possibly the most fundamental quality of human existence (Metzinger, 2003). At the same time, however, there are numerous considerations that cast doubts on the intuitions regarding cross-modal "integration" (recall section 3). Furthermore, there are the constraints of the distributed, concurrent nature of brain function and of the nonzero physical extent of the brain, which imply that it is physically impossible to instantaneously integrate the information in a brain "state" (as per section 4.1).

We believe that the tension between the phenomenal unity and the difficulties that it runs into must be resolved by rejecting attempts to reconcile the two. Instead, in light of the preceding discussion, we propose that the intuitive conception of the unity of consciousness be relegated to the status of a "folk theory" that need not be given a detailed mechanistic explanation, simply because on the level of process and mechanism it lacks substance.[24]

To facilitate the transition from a unitary conception of phenomenal consciousness to a pluralistic one, which views all of cognition (including consciousness) as a bundle of concurrent processes interspersed with asynchronous events (as per section 5), it may be useful to revisit the familiar simile from William James (1890, p.236):

> When we take a general view of the wonderful stream of our consciousness, what strikes us first is the different pace of its parts. Like a bird's life, it seems to be an alternation of flights and perchings.

This we propose to modify, likening consciousness to the life of a *flock* of birds, which generally moves as one, yet includes both early birds and stragglers, so that the aggregate movement happens without the individual birds necessarily perching or taking off all at the same time.[25]

[23] See (Mudrik, Faivre, and Koch, 2014) for an extensive survey of the concept of, and evidence for, perceptual integration and its relation to consciousness.

[24] We would still have to explain, in terms that do not involve an appeal to "integration," why our experience of the world is normally phenomenally single-threaded, and why special conditions such as synaesthesia (Grossenbacher and Lovelace, 2001) are not more prevalent.

[25] The modified simile brings to mind the Simurgh metaphor for the self (Edelman, 2008a, p.471); cf. (Borges, 1962).

Acknowledgments

Thanks to Arnon Lotem and Barbara Finlay for many discussions, to David Harel for alerting us to the LSC formalism, to Assaf Marron and Björn Merker for detailed comments on an early version of this paper, and to two anonymous reviewers for constructive remarks.

References

Adriaans, P. and M. Vervoort (2002). The EMILE 4.1 grammar induction toolbox. In P. Adriaans, H. Fernau, and M. van Zaanen (Eds.), *Grammatical Inference: Algorithms and Applications: 6th International Colloquium: ICGI 2002*, Volume 2484 of *Lecture Notes in Computer Science*, pp. 293–295. Heidelberg: Springer-Verlag.

Aldridge, J. W. and K. C. Berridge (1998). Coding of serial order by neostriatal neurons: a 'natural action' approach to movement sequence. *Journal of Neuroscience 18*, 2777–2787.

Aldridge, J. W. and K. C. Berridge (2003). Basal ganglia neural coding of natural action sequences. In A. M. Graybiel, M. R. Delong, and S. T. Kitai (Eds.), *The Basal Ganglia VI*, Volume 54 of *Advances in Behavioral Biology*. New York: Springer.

Allen, T. A., A. M. Morris, A. T. Mattfeld, C. E. Stark, and N. J. Fortin (2014). A sequence of events model of episodic memory shows parallels in rats and humans. *Hippocampus 24*, 1178–1188.

Anderson, D. J. and P. Perona (2014). Toward a science of computational ethology. *Neuron 84*, 18–31.

Andoni, A. and P. Indyk (2008). Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of the ACM 51*, 117–122.

Angelaki, D. E., Y. Gu, and G. C. DeAngelis (2009). Multisensory integration: psychophysics, neurophysiology and computation. *Current Opinion in Neurobiology 19*, 452–458.

Atallah, H. E., M. J. Frank, and R. C. O'Reilly (2004). Hippocampus, cortex, and basal ganglia: Insights from computational models of complementary learning systems. *Neurobiology of Learning and Memory 82*, 253–267.

Atallah, H. E., A. D. McCool, M. W. Howe, and A. M. Graybiel (2014). Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. *Neuron 82*, 1145–1156.

Bayne, T. and D. Chalmers (2003). What is the unity of consciousness? In A. Cleeremans (Ed.), *The Unity of Consciousness: Binding, Integration and Dissociation*, pp. 23–58. Oxford: Oxford University Press.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science 33*, 752–793.

Bolhuis, J. J., K. Okanoya, and C. Scharff (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience 11*, 747–759.

Bolt, R. A. (1980). Put-that-there: Voice and gesture at the graphics interface. In *Proc. SIGGRAPH '80*, New York, NY, pp. 262–270. ACM.

Borges, J. L. (1935/1962). The Approach to Al-Mu'tasim. In *Ficciones*. New York: Grove Press. Translated by A. Bonner in collaboration with the author.

Braun, M. L., J. M. Buhmann, and K.-R. Müller (2008). On relevant dimensions in kernel feature spaces. *Journal of Machine Learning Research 9*, 1875–1908.

Breitenstein, C., A. Jansen, M. Deppe, A.-F. Foerster, J. Sommer, T. Wolbers, and S. Knecht (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage 25*, 958–968.

Briegel, H. J. (2012). On creative machines and the physical origins of freedom. *Scientific Reports 2*(522), 1–6.

Brooks, R. A. (1989). A robot that walks: Emergent behaviors from a carefully evolved network. *Neural Computation 1*, 253–262.

Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences 8*, 426–433.

Burgess, N. and G. Hitch (2005). Computational models of working memory: putting long-term memory into context. *Trends in Cognitive Sciences 9*, 535–541.

Buzsáki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral Cortex 6*, 81–92.

Buzsáki, G. and K. Diba (2010). Oscillation-supported information processing and transfer in the hippocampus-entorhinal-neocortical interface. In C. von der Malsburg, W. A. Phillips, and W. Singer (Eds.), *Dynamic Coordination in the Brain: From Neurons to Mind*, Volume 5 of *Strüngmann Forum Report*, Chapter 7, pp. 101–114. Cambridge, MA: MIT Press.

Camastra, F. (2003). Data dimensionality estimation methods: a survey. *Pattern Recognition 36*, 2945–2954.

Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proc. of the IEEE 86*, 2009–2025.

Casati, R. and A. Varzi (2014). Events. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014 ed.).

Chabrol, F. P., A. Arenz, M. T. Wiechert, T. W. Margrie, and D. A. DiGregorio (2015). Synaptic diversity enables temporal coding of coincident multisensory inputs in single neurons. *Nature Neuroscience 18*, 718–731.

Chambers, N. and D. Jurafsky (2008). Unsupervised learning of narrative event chains. In *Proceedings of ACL/HLT 2008*.

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition 113*, 350–364.

Chen, D. and H.-G. Müller (2012). Nonlinear manifold representations for functional data. *The Annals of Statistics 40*, 1–29.

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

Christiansen, M. H. and N. Chater (2008). Language as shaped by the brain. *Behavioral and Brain Sciences 31*, 489–509.

Clark, A. (1993). *Sensory qualities*. Oxford: Clarendon Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences 36*, 181–204.

Coen, M. H. (2006). *Multimodal dynamics: self-supervised learning in perceptual and motor systems*. Ph. D. thesis, Massachusetts Institute of Technology.

Cover, T. and P. Hart (1967). Nearest neighbor pattern classification. *IEEE Trans. on Information Theory IT-13*, 21–27.

Culicover, P. W. and R. Jackendoff (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Dale, R., R. Fusaroli, N. D. Duran, and D. Richardson (2013). The self-organization of human interaction. In B. Ross (Ed.), *Psychology of Learning and Motivation*, Volume 59, pp. 43–95. Elsevier.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation 1*, 123–132.

Damm, W. and D. Harel (2001). LSCs: Breathing life into message sequence charts. *Formal Methods in System Design 19*, 45–80.

Danz, A. D. (2011). Multimodal temporal processing between separate and combined modalities. In B. Kokinov, A. Karmiloff-Smith, and N. J. Nersessian (Eds.), *European Perspectives on Cognitive Science*.

Dave, A. S. and D. Margoliash (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science 290*, 812–816.

Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.

Davidson, T. J., F. Kloosterman, and M. A. Wilson (2009). Hippocampal replay of extended experience. *Neuron 63*, 497–507.

de Bakker, J. W. and J. I. Zucker (1982). Processes and the denotational semantics of concurrency. *Information and Control 54*, 70–120.

Deco, G., G. Tononi, M. Boly, and M. L. Kringelbach (2015). Rethinking segregation and integration: contributions of whole-brain modelling. *Nature Reviews Neuroscience 16*, 430–439.

DeLong, G. R. and E. R. Heinz (1997). The clinical syndrome of early-life bilateral hippocampal sclerosis. *Annals of Neurology 42*, 11–17.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review 3*, 357–370.

Deyle, E. R. and G. Sugihara (2011). Generalized theorems for nonlinear state space reconstruction. *PLoS ONE 6*, e18295.

Diaz, J., C. Muñoz-Caro, and A. Niño (2012). A survey of parallel programming models and tools in the multi and many-core era. *IEEE Transactions on Parallel and Distributed Systems 23*, 1369–1386.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher 35*, 125–129.

Doubell, T. P., T. Skaliora, J. Baron, and A. J. King (2003). Functional connectivity between the superficial and deeper layers of the superior colliculus: an anatomical substrate for sensorimotor integration. *Journal of Neuroscience 23*, 6596–6607.

Dragoi, G. and S. Tonegawa (2013). Distinct preplay of multiple novel spatial experiences in the rat. *Proceedings of the National Academy of Science 110*, 9100–9105.

DuBrow, S. and L. Davachi (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General 142*, 1277–1286.

Duff, M. C. and S. Brown-Schmidt (2012). The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience 6*, 69.

Eagleman, D. M. (2010). How does the timing of neural signals map onto the timing of perception? In R. Nijhawan (Ed.), *Problems of space and time in perception and action*. Cambridge: Cambridge University Press.

Edelman, S. (1987). Line connectivity algorithms for an asynchronous pyramid computer. *Computer Vision, Graphics, and Image Processing 40*, 169–187.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.

Edelman, S. (2008a). *Computing the mind: how the mind really works*. New York, NY: Oxford University Press.

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

Edelman, S. (2011). On look-ahead in language: navigating a multitude of familiar paths. In M. Bar (Ed.), *Prediction in the Brain*, Chapter 14, pp. 170–189. New York: Oxford University Press.

Edelman, S. (2015a). Brain grammar: computational nature and possible brain mechanisms of sequential behavior. In preparation.

Edelman, S. (2015b). The minority report: some common assumptions to reconsider in the modeling of the brain and behavior. *Journal of Experimental and Theoretical Artificial Intelligence 27*, –.

Edelman, S. and T. Fekete (2012). Being in time. In S. Edelman, T. Fekete, and N. Zach (Eds.), *Being in Time: Dynamical Models of Phenomenal Experience*, pp. 81–94. John Benjamins.

Edelman, S. and N. Intrator (2002). Models of perceptual learning. In M. Fahle and T. Poggio (Eds.), *Perceptual learning*, pp. 337–353. MIT Press.

Edelman, S. and R. Shahbazi (2012). Renewing the respect for similarity. *Frontiers in Computational Neuroscience 6*, 45.

Eichenbaum, H., C. J. MacDonald, and B. J. Kraus (2014). Time and the hippocampus. In D. Derdikman and J. J. Knierim (Eds.), *Space, Time and Memory in the Hippocampal Formation*, pp. 273–301. Vienna: Springer.

Eichenbaum, H., T. Otto, and N. J. Cohen (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences 17*, 449–517.

Eisler, H. (1960). Similarity in the continuum of heaviness with some methodological and theoretical considerations. *Scand. J. Psychol. 1*, 69–81.

Ekstrom, A. D., M. S. Copara, E. A. Isham, W.-C. Wang, and A. P. Yonelinas (2011). Dissociable networks involved in spatial and temporal order source retrieval. *NeuroImage 56*, 1803–1813.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science 33*, 547–582.

Fetsch, C. R., G. C. DeAngelis, and D. E. Angelaki (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience 14*, 429–442.

Fodor, J. A., T. G. Bever, and M. F. Garrett (1974). *The psychology of language*. New York: McGraw Hill.

Fortin, N. J., K. L. Agster, and H. B. Eichenbaum (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience 5*, 458–462.

Frank, M. C., J. B. Tenenbaum, and A. Fernald (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language, Learning, and Development 9*, 1–24.

Fuster, J. M. and S. L. Bressler (2012). Cognit activation: a mechanism enabling temporal integration in working memory. *Trends in Cognitive Sciences 16*, 207–218.

Garner, W. R. and G. L. Felfoldy (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology 1*, 225–241.

Ghassempour, S., F. Girosi, and A. Maeder (2014). Clustering multivariate time series using Hidden Markov Models. *Int. J. Environ. Res. Public Health 11*, 2741–2763.

Goldstein, M. H., H. R. Waterfall, A. Lotem, J. Halpern, J. Schwade, L. Onnis, and S. Edelman (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences 14*, 249–258.

Groh, J. M. and U. Werner-Reiss (2002). Visual and auditory integration. In V. S. Ramachandran (Ed.), *Encyclopedia of the Human Brain*, pp. 739–752. San Diego, CA: Academic Press.

Gros-Louis, J., D. J. White, A. P. King, and M. J. West (2003). Female brown-headed cowbirds' (*Molothrus ater*) social assortment changes in response to male song: a potential source of public information. *Behav. Ecol. Sociobiol. 53*, 163–173.

Grossenbacher, P. G. and C. T. Lovelace (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends in Cognitive Sciences 5*, 36–41.

Guttman, S. E., L. A. Gilroy, and R. Blake (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological Science 16*, 228–235.

Harel, D. (1988). On visual formalisms. *Commun. ACM 31*, 514–530.

Harel, D. (2007). Statecharts in the making: a personal account. In *HOPL III: Proceedings of the third ACM SIGPLAN conference on History of programming languages*, New York, NY, pp. 5–1–5–43. ACM.

Harel, D., A. Marron, and G. Weiss (2012). Behavioral programming. *Communications of the ACM 55*, 90–100.

Harris-Warrick, R. M., L. M. Coniglio, R. M. Levin, S. Gueron, and J. Guckenheimer (1995). Dopamine modulation of two subthreshold currents produces phase shifts in activity of an identified motoneuron. *J. of Neurophysiology 74*, 1404–1420.

Henson, R. N. A. and N. Burgess (1997). Representations of serial order. In J. A. Bullinaria, D. W. Glasspool, and G. Houghton (Eds.), *4th Neural Computation and Psychology Workshop*, pp. 283–300. London: Springer.

Hilliard, C., E. O'Neal, J. Plumert, and S. Wagner (2015). Mothers modulate their gesture independently of their speech. *Cognition 140*, 89–94.

Hockett, C. F. (1960). The origin of speech. *Scientific American 203*, 88–96.

Hume, D. (1740). *A Treatise of Human Nature*. Available online at http://www.gutenberg.org/etext/4705.

Hurford, J. R. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences 26*, 261–316. Available online at http://www.isrl.uiuc.edu/~amag/langev/paper/hurford01theNeural.html.

Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Computation 18*, 245–282.

James, W. (1890). *The Principles of Psychology*. New York: Holt. Available online at http://psychclassics.yorku.ca/James/Principles/.

Ji, D. and M. A. Wilson (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience 10*, 100–107.

Jin, X. and R. M. Costa (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature 466*, 457–462.

Jin, X., F. Tecuapetla, and R. M. Costa (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nature Neuroscience 17*, 423–433.

Keele, S. W., R. Ivry, U. Mayr, E. Hazeltine, and H. Heuer (2003). The cognitive and neural architecture of sequence representation. *Psychological Review 110*, 316–339.

Keeley, B. (2002). Making sense of the senses: individuating modalities in humans and other animals. *The Journal of Philosophy XCIX*, 5–28.

Kepecs, A., N. Uchida, and Z. F. Mainen (2006). The sniff as a unit of olfactory processing. *Chemical Senses 31*, 167–179.

Kersten, D., P. Mamassian, and A. Yuille (2004). Object perception as Bayesian inference. *Annual Review of Psychology 55*, 271–304.

Kersten, D. and A. Yuille (2003). Bayesian models of object perception. *Current Opinion in Neurobiology 13*, 1–9.

Kolodny, O., A. Lotem, and S. Edelman (2015). Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cognitive Science 39*, 227–267.

Körding, K. P. and D. M. Wolpert (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences 10*, 319–326.

Kotz, S. A., M. Schwartze, and M. Schmidt-Kassow (2009). Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex 45*, 982–990.

Kurczek, J., S. Brown-Schmidt, and M. Duff (2013). Hippocampal contributions to language: Evidence of referential processing deficits in amnesia. *Journal of Experimental Psychology: General 142*, 1346–1354.

Lamb, S. M. (2004). *Language and reality*. London: Continuum.

Langacker, R. W. (1987). *Foundations of cognitive grammar*, Volume I: theoretical prerequisites. Stanford, CA: Stanford University Press.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior*, pp. 112–146. New York: Wiley.

Levy, W. B. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus 6*, 579–590.

Levy, W. B., A. B. Hocking, and X. Wu (2005). Interpreting hippocampal function as recoding and forecasting. *Neural Networks 18*, 1242–1264.

Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology 45*, 36–62.

MacIver, M. A. (2009). Neuroethology: From morphological computation to planning. In P. Robbins and M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition*, pp. 480–504. New York, NY: Cambridge University Press.

Marr, D. (1976). Early processing of visual information. *Phil. Trans. R. Soc. Lond. B 275*, 483–524.

May, P. J. (2006). The mammalian superior colliculus: laminar structure and connections. *Prog. Brain Res. 151*, 321–378.

McAuley, J. D., M. R. Jones, S. Holub, H. M. Johnston, and N. S. Miller (2006). The time of our lives: life span development of timing and event tracking. *Journal of Experimental Psychology: General 135*, 348–367.

Medin, D. L., R. L. Goldstone, and D. Gentner (1993). Respects for similarity. *Psychological Review 100*, 254–278.

Mehlmann, G. and E. André (2012). Modeling multimodal integration with event logic charts. In *Proc. ICMI'12*, Santa Monica, CA.

Meredith, M. A. and B. E. Stein (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology 56*, 640–662.

Merker, B. (2004). Cortex, countercurrent context, and dimensional integration of lifetime memory. *Cortex 40*, 559–576.

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behavioral and Brain Sciences 30*, 63–81.

Merker, B. (2013a). Cortical gamma oscillations: the functional key is activation, not cognition. *Neuroscience and Biobehavioral Reviews 37*, 401–417.

Merker, B. (2013b). The efference cascade, consciousness, and its self: naturalizing the first-person pivot of action control. *Frontiers in Psychology 4*(501), 1–20.

Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the Institute of Radio Engineers 49*, 8–30.

Mudrik, L., N. Faivre, and C. Koch (2014). Information integration without awareness. *Trends in Cognitive Sciences 18*, 488–496.

Muller, R. U. and J. L. Kubie (1989). The firing of hippocampal place cells predicts the future position of freely moving rats. *Journal of Neuroscience 9*, 4101–4110.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In C. Koch and J. L. Davis (Eds.), *Large-scale neuronal theories of the brain*, Chapter 7, pp. 125–152. Cambridge, MA: MIT Press.

Nakahara, H., K. Doya, and O. Hikosaka (2001). Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences — a computational approach. *Journal of Cognitive Neuroscience 13*, 626–647.

O'Regan, J. K., E. Myin, and A. Noë (2004). Towards an analytic phenomenology: the concepts of "bodiliness" and "grabbiness". In A. Carsetti (Ed.), *Seeing, Thinking and Knowing*, Volume 38 of *Theory and Decision Library A*, pp. 103–114. Springer.

Park, S. and J. K. Aggarwal (2004). A hierarchical Bayesian network for event recognition of human actions and interactions. *Multimedia Systems 10*, 164–179.

Paz, R., H. Gelbard-Sagiv, R. Mukamel, M. Harel, R. Malach, and I. Fried (2010). A neural substrate in the human hippocampus for linking successive events. *Proceedings of the National Academy of Science 107*, 6046–6051.

Perrodin, C., C. Kayser, N. K. Logothetis, and C. I. Petkov (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proceedings of the National Academy of Science 112*, 273–278.

Peyrache, A., M. Khamassi, K. Benchenane, S. I. Wiener, and F. P. Battaglia (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience 12*, 919–929.

Philipona, D. and J. O'Regan (2010). The sensorimotor approach in CoSy: The example of dimensionality reduction. In H. I. Christensen, G. J. M. Kruijff, and J. L. Wyatt (Eds.), *Cognitive Systems*, pp. 95–130.

Philipona, D., J. K. O'Regan, J.-P. Nadal, and O. J.-M. Coenen (2004). Perception of the structure of the physical world using unknown sensors and effectors. *Advances in Neural Information Processing Systems 15*, 945–952.

Phillips, C. (2003). Syntax. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*, Volume 4, pp. 319–329. London: Macmillan.

Prescott, T. J., P. Redgrave, and K. N. Gurney (1999). Layered control architectures in robots and vertebrates. *Adaptive Behavior 7*, 99–127.

Rabinovich, M. I., R. Huerta, P. Varona, and V. S. Afraimovich (2008). Transient cognitive dynamics, metastability, and decision making. *PLoS Comput. Biol. 4(5)*, e1000072.

Reimer, J. F., G. A. Radvansky, T. C. Lorsbach, and J. J. Armendarez (2015). Event structure and cognitive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. In press.

Rigotti, M., O. Barak, M. R. Warden, X.-J. Wang, N. D. Daw, E. K. Miller, and S. Fusi (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature 497*, 585–590.

Rosenbaum, D. A., R. G. Cohen, S. A. Jax, D. J. Weiss, and R. van der Wel (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science 26*, 525–554.

Rubin, D. C. and S. Umanath (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review 122*, 1–23.

Salazar, R. F., N. M. Dotson, S. L. Bressler, and C. M. Gray (2012). Content-specific fronto-parietal synchronization during visual working memory. *Science 338*, 1097–1100.

Schendan, H. E., M. M. Searl, R. J. Melrose, and C. E. Stern (2003). An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron 37*, 1013–1025.

Seger, C. A. (2006). The basal ganglia in human learning. *The Neuroscientist 12*, 285–290.

Selverston, A. (2008). Stomatogastric ganglion. *Scholarpedia 3*(4), 1661.

Senghas, A., S. Kita, and A. Özyürek (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science 305*, 1779–1782.

Shatz, D. (2014). 'Yeah, Yeah': Eulogy for Sidney Morgenbesser, Philosopher With a Yiddish Accent. *Tablet*. Available at http://tabletmag.com/jewish-arts-and-culture/books/177249/sidney-morgenbesser.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science 237*, 1317–1323.

Skarda, C. and W. J. Freeman (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences 10*, 161–195.

Solan, Z., D. Horn, E. Ruppin, and S. Edelman (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Science 102*, 11629–11634.

Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing Co.

Spivey, M. J. (2006). *The continuity of mind*. New York: Oxford University Press.

Sprouse, J., M. Wagers, and C. Phillips (2012). Working-memory capacity and island effects: A reminder of the issues and the facts. *Language 88*, 401–407.

Stabler, E. P. (2013). Two models of Minimalist, incremental syntactic analysis. *Topics in Cognitive Science 5*, 611–633.

Sutton, R. S. and A. G. Barto (1998). *Reinforcement Learning*. Cambridge, MA: MIT Press.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand and L.-S. Young (Eds.), *Dynamical systems and turbulence*, Lecture Notes in Mathematics, pp. 366–381. Berlin: Springer.

Taylor, R. C. and M. J. Ryan (2013). Interactions of multisensory components perceptually rescue Túngara frog mating signals. *Science 341*, 273–274.

Tenenbaum, J. B. and T. L. Griffiths (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences 24*, 629–641.

Treisman, A. and G. Gelade (1980). A feature integration theory of attention. *Cognitive Psychology 12*, 97–136.

Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews Neuroscience 2*, 717–727.

van Wassenhove, V. (2009). Minding time in an amodal representational space. *Phil. Trans. R. Soc. B 364*, 1815–1830.

van Zaanen, M. (2000). ABL: Alignment-Based Learning. In *Proceedings of the 18th International Conference on Computational Linguistics*, pp. 961–967. Available online at http://citeseer.nj.nec.com/article/vanzaanen00abl.html.

Vigliocco, G., P. Perniss, and D. Vinson (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Phil. Trans. R. Soc. B 369*, 20130292.

Wallenstein, G. V., H. Eichenbaum, and M. E. Hasselmo (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences 21*, 317–323.

Watanabe, S. (1969). *Knowing and Guessing: A Quantitative Study of Inference and Information*. New York: Wiley.

Waterfall, H. R., B. Sandbank, L. Onnis, and S. Edelman (2010). An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language 37*(Special issue 03), 671–703.

Wilson, A. J., M. Dean, and J. P. Higham (2013). A game theoretic approach to multimodal communication. *Behavioral Ecology and Sociobiology 67*, 1399–1415.

Wilson, M. A. and B. L. McNaughton (1994). Reactivation of hippocampal ensemble memories during sleep. *Science 265*, 676–679.

Wood, E., P. A. Dudchenko, R. J. Robitsek, and H. Eichenbaum (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron 27*, 623–633.

Yildirim, I. and R. A. Jacobs (2015). Learning multisensory representations for auditory-visual transfer of sequence category knowledge: a probabilistic language of thought approach. *Psychonomic Bulletin & Review 22*, 673–686.

Zacks, J. M., T. S. Braver, M. A. Sheridan, D. I. Donaldson, A. Z. Snyder, J. M. Ollinger, R. L. Buckner, and M. E. Raichle (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience 4*, 651–655.

Zacks, J. M., N. K. Speer, K. M. Swallow, T. S. Braver, and J. R. Reynolds (2007). Event perception: a mind-brain perspective. *Psychological Bulletin 133*, 273–293.

Zacks, J. M. and B. Tversky (2001). Event structure in perception and conception. *Psychological Bulletin 127*, 3–21.