Eliasmith, C. (in press). Dynamics, control, and cognition. In P. Robbins and M. Aydede (Eds.) Cambridge Handbook of Situated Cognition. Cambridge University Press.

Dynamics, Control, and Cognition

Chris Eliasmith

Once upon real-time

A dynamic object is an object whose properties change over time. A static object is an object whose properties do not change over time. Given such an idealization, the notion of 'static' lies at an extreme end of the spectrum of temporal relations between objects and properties. Indeed, modern physics tells us that no objects are truly static. Nevertheless, many of our physical, computational, and metaphysical theories turn a blind eye to the role of time, often for practical reasons. So, perhaps it is not surprising that in the philosophy of mind – where physical, computational, and metaphysical theories meet – there has been a consistent tendancy to articulate theories that consider function and time independently. As a result, contemporary theories in cognitive science consider time unsystematically (see the next section for specific examples). In this chapter, I suggest that the problem with this 'ad hocery' is that the systems we are trying to characterize are real-time systems, whose real-time performance demands principled explanation (a point on which many of these same contemporary theorists agree). After a discussion of the importance and roots of dynamics in cognitive theorizing, I describe the role of time in each of the three main approaches to cognitive science: symbolicism, connectionism and dynamicism. Subsequently, I outline a recently proposed method, the Neural Engineering Framework (NEF), that, unlike past approaches, permits a principled integration of dynamics into biologically realistic models of high-level cognition. After briefly presenting a model, BioSLIE, that demonstrates this integration using the NEF, I argue that this approach alone is in a position to properly integrate dynamics, biological realism, and high-level cognition.

Historically, many cognitive theories have not been particularly informed by our understanding of biological systems. Arguably, this is because our understanding of the mechanisms driving biological systems was in its infancy until very recently. This suggests that there was little opportunity for theories of mind to gain insight from our understanding of the kinds of systems which putatively have minds. So, there was little inspiration to be drawn from biology regarding mentality. However, recent decades have seen a radical change in this state of affairs. Neuroscience, the subdiscipline of biology which has the most offer theories of mind, only began to systematically explore neural mechanisms quite recently (e.g., after the pioneering experiments of Hodgkin and Huxley (1952) and Hubel and Weisel (Hubel & Wiesel, 1962; Wiesel & Hubel, 1963)).[1] Despite these relatively recent beginnings, the annual conference of the Society for Neuroscience

[1] Of course, much of the groundwork was laid before this. But even as late as 1906, there was still public debate (at the Nobel Prize awards ceremony) regarding the existence of individual nerve cells. As well, intracellular recording techniques were not developed until the 1940s, and basic single cell ion dynamics were not characterized until the 1950s. See Finger (2000) for an extended account of the early history of neurobiology.

features approximately 30,000 attendees, most of whom are directly involved in exploring the mechanisms of the brain. I suspect that all of them, almost without exception, are acutely aware of the dynamics of the mechanisms they are studying.

The importance of the dynamics of neural mechanisms for understanding the brain can be gleaned from the kinds of vocabulary typically employed by neuroscientists. They inevitably speak of “time constants,” “time courses,” “fluctuations,” “firing rates,” “spike timing dependent plasticity,” “theta, gamma, delta, etc. oscillations,” “molecular kinetics,” “membrane dynamics,” “protein dynamics,” “short and long-term plasticity,” “synchrony and temporal correlations,” and so on. In other words, a careful examination of the mechanisms underlying mental phenomena have demanded temporally laden descriptions.

Of course, there is no obvious reason why it is necessary to learn about the brain before gleaning the importance of dynamics. Perhaps the traditional division between cognition and perception/action, reflected neatly in the notion of man as a ‘rational animal,’ suggested to many early cognitive scientists that rationality, a reasonably outside-of-time kind of behavior, is their target of inquiry. Unfortunately, this view relegates many of the dynamical aspects of behavior to the status of an afterthought. No doubt this perspective was bolstered by the development of the von Neumann architecture for computers, which neatly distinguishes input/output functions from central processing, whose temporal properties are determined by a clock that can be sped up or slowed down with little functional consequence. And, it was clear to anyone studying computer systems that the central processor was the most important part of the system. This characterization is efficiently captured in the now famous ‘mind-as-computer’ metaphor that has so dominated the history of cognitive science.

Perhaps one way to move past this metaphor for mind is to learn more about the target of the metaphor. That is, it may be no coincidence that as our understanding of the brain has improved, the ‘standard’ conception of cognition has become more dynamical. In other words, I suspect that the broad ‘dynamical shift’ of cognitive science is widely inspired by the ‘neuro’-izing of the discipline.

While our improved understanding of neural mechanisms has likely cemented the importance of dynamics for understanding cognition, another route to this view can be found in the history of psychology. In particular, the work of psychologist J. J. Gibson and his colleagues at Cornell provides further impetus for taking dynamics seriously (Gibson, 1966; Gibson & Gibson, 1955). In fact, the focus of this research was not on dynamics per se, but on the active role that a perceiver takes in exploiting its own motion to extract relevant information, or underwrite environmental interactions. Gibson described agents as “resonating” with certain information in their environment that is relevant for their potential actions. His well-known notion of an “affordance” captures this theoretical position. It is affordances, after all, which agents are specially tuned to pick up as environmental objects of interest *to them* (e.g., a stump affords living quarters for an insect, but a seat for us).

This emphasis on the environment, and on the relation between agents and environments, has served as a theoretical predecessor to the contemporary concern for the situatedness and embodiedness of agents.[2] Such theories, including Gibson's original characterization, focus attention on movements of an agent within an environment. This necessarily highlights the importance of characterizing both environmental and agent-centered dynamics. For instance, an approach to characterizing visual perception championed by Dana Ballard, called ‘animate vision’, focuses on determining what information agents actively extract from a visual scene through rapid eye movements, rather than taking the traditional Marrian approach of trying to reconstruct the entire visual scene in a 3-D internal representation (Ballard, 1991). It is the dynamics of the agent and the environment that determines what information is available to be acted on.

Despite the fact that traditional computational approaches to understanding cognitive function often label themselves “information processing” approaches, the strongest arguments for the importance of situatedness come from information theoretic considerations. Simply put, there is too much information in an environment for any known sensory organ to extract it all. Sensory organs clearly have limited bandwidth; that is, a limited ability to extract the various information available in natural environments. And while there is evidence that many such systems are near their theoretical limits for extracting information (Rieke *et al.*, 1997), even reaching such limits will not ameliorate the problem of dealing with *all* the information in an environment. Given such ‘hard’ constraints, it is not surprising that biological systems have developed various means of targeting evolutionarily relevant information sources in their environment. It is those sources, after all, that determine if they live or die. As a result of these considerations it becomes clear that *how these systems target information* is as important as *how they pick up that information* once they are oriented towards it. Furthermore, if those methods of targeting are highly sensitive to environmental dynamics, as they clearly seem to be,[3] then it is also essential to understand the dynamics of such an environment. As a result, the dynamics of the agent, the dynamics of the environment, and equally importantly, their interaction, are what need to be understood in order to properly characterize “information processing” in biological systems.

In short, experimental considerations of neural mechanisms and theoretical considerations of agent/environment interactions conspire to suggest that dynamics are an inescapable feature of cognitive systems. This is in contrast to the traditional view that ‘cognitive systems’ are best characterized through a firm theoretical grounding in computational theory. The problem with this traditional picture is that artificial intelligence and computer science researchers are not especially interested in dynamical

[2] Although it should be noted that some symbolicists also seem to have been sensitive to the importance of this interaction: “A proper understanding of the intimate interdependence between an adaptive organism and its environment is essential to a clear view of what a science of an adaptive species can be like.” (Newell and Simon, 1972, p. 870).

[3] This is just the observation that *change* is often an important environmental cue Thus, visual features such as motion are often used to orient an animal towards potentially interesting or dangerous aspects of their environment.

systems. That is, while the design of real-time systems is only one small part of computer science, the only kind of systems ever designed by mother nature are real-time. At the moment, by far the most impressive cognitive systems are natural ones.

Dynamic duels: Dynamics and cognitive architectures

Having briefly argued for the importance of dynamics for understanding cognition, I turn to the issue of how dynamics have been integrated into various theories of cognition. After a brief historical discussion, I describe the strengths and weaknesses of the three main contenders in cognitive science, especially in relation to their incorporation of time into their methods of model construction.

Behaving in time

In the early part of the last century, the dominant theory in psychology was behaviorism. Famously, behaviorism espoused the view that the only scientifically respectable 'observables' that could underwrite a psychological theory were behavioral events. They argued that only such external events were objectively observable, and thus that only they could be the subject of an objective scientific theory (Watson, 1913). It is somewhat unclear from their collective writings exactly how important dynamics were, or were not, for supporting this understanding of psychological agents.[4] Whatever the case, their behavioral standpoint was unarguably infused with dynamics in the hands of the 'cyberneticists.'

Cybernetics is the study of feedback and control in both artificial in biological systems. It grew from a wartime interest in real-world, goal directed systems – especially enemy-directed systems. Norbert Wiener, who coined the term 'cybernetics,' was a mathematician with interests in communication theory who worked on gun controllers (Wiener, 1948). It has been suggested that Wiener realized that the study of stability and control of the anti-aircraft systems he was working on could be extended to the operator of the system as well (Freudenthal, 1970-1990). As a result, he had the insight that same mathematical tools for understanding goal-directed artificial systems, could be applied to goal-directed natural systems.

The mathematical tools that Wiener used are typically grouped under the heading of 'classical control theory.' Very briefly, classical control theory considers the system under study as implementing a temporal transfer function, which describes how inputs are converted into outputs over time. The point of classical control is to design a control system which can be used to alter the inputs of the system in order to achieve a desired output. In 'open-loop control' the controller simply provides inputs which should, under normal circumstances, achieve the desired outputs. However, since normal circumstances are often difficult to define in advance, and the circumstances themselves are likely to change over time, a more sophisticated form of control called 'closed-loop

[4] Neither Skinner nor Watson, for instance, make special mention of dynamics. However, Hull (1935) in his quest to write Newtonian-like laws for behavior, seems somewhat concerned with the effects of interstimulus delays during learning. However, none of the equations he explicitly writes have a time parameter.

control,' or 'feedback control,' is more commonly employed. In closed-loop control, the inputs provided to the system depend on its current outputs, which are often affected by the current circumstances (e.g., in automobile cruise control, road conditions, hills, etc. greatly affect the effect of various accelerator inputs). The effectiveness of closed loop controllers was demonstrated time and again during WWII, by their inclusion in target trackers, self-guided torpedoes, and various other servomechanisms (Mindell, 1995).

To this day, classical control methods are taught to engineers in order to provide them with strong intuitions about how simple control systems can be analyzed and designed. These methods play this role because they are largely graphical, are easily applicable to simple, single input/single output systems, and introduce a number of useful heuristics for control design. However, when trying to understand a complex control system like the brain, many of these pedagogical strengths become practical weaknesses. For instance, there is no reason to think that a biological system is a single input/single output system. As well, when dealing with complex controllers, graphical methods soon become limiting and clumsy because of their restricted dimensionality. Additional theoretical limitations on classical control include an inability to: quantify optimal control; to characterize adaptive control; and to systematically include considerations of noise.

Despite these limitations, classical control successfully began a practical quantification of real-time systems. As well, the cyberneticist focus on temporal input/output relations (captured by the transfer functions) integrated well with the behaviorist psychology of the day. That is, both classical control theorists and behaviorists did not need to 'look inside' the systems they were interested in understanding. What the control theorists added, of course, was an explicitly dynamical dimension to an otherwise static characterization of cognitive systems.

A cognitive resolution

The famous "cognitive revolution" that took place in the mid-1950s is often hailed as an essential turning point in the history of cognitive science, a turning point without which cognitive science would not have fruitfully developed (Bechtel & Graham, 1999; Thagard, 1996). This may be true in part, but there was also a significant price that was paid for the sweeping adoption of the cognitivist view. This is because the resolution of behaviorist difficulties came in two parts. One was a shift in focus from input/output relations to internal states of cognitive systems. The second was a shift from mathematical models of behavior to computational ones. With this second shift came a general acceptance that the relevant formal theory for characterizing cognitive systems was grounded in abstract entities that have no connection to time: Turing machines. For instance, Newell and Simon (1972) wrote in their historical epilogue that "The formalization of logic showed that symbols can be copied, compared, rearranged, and concatenated with just as much definiteness of process as boards can be sawed, planed, measured, and glued...Symbols became, for the first time, tangible – as tangible as wood or metal. The Turing machine was an all-purpose planar and lathe for symbols" (p. 877-8). A basic assumption of this kind of computational theory is that resources are infinite. So a computable function is one that can be accomplished regardless of temporal, memory, or other constraints. Unfortunately, despite the fact that considering internal states is independent of the formal theory for considering such states, it so happened that by adopting computational theory, time was pushed aside by the cognitive sciences. In other words, *it just so happened* that the formal theory that informed this ‘symbolicist’ characterization of cognition cleaved time from function. An assumption not reflected in natural cognitive systems.

As a result, it is not surprising that in their attack on the temporal deficiencies of these symbolic characterizations of cognitive systems, Port and van Gelder (1995) claim that the symbolicists “*leave time out of the picture*”. But, on the face of it, this is untrue. Consider, for instance, Newell's (1990) paradigmatic symbolicist cognitive model SOAR. In his discussion of this model, and its theoretical underpinnings, Newell includes “operate in real time” as the third of thirteen constraints that shapes the mind (Newell, 1990, p. 19). Thus, it is simply not the case the symbolicists ignore time. However, I believe it clearly is the case that they have great difficulty meeting this essential constraint.

Newell appeals to various neurological data to lend support to his assumption that any particular step (or ‘production’) in a cognitive algorithm operates on the time scale of approximately 10ms (Newell, 1990, p. 127). However, his application of this constraint seems rather contrived. For instance, in one application, SOAR employs a single production to encode whether or not a light is on (Newell, 1990, p. 275). But, in a second application, SOAR uses a single production to encode: “If the problem space is the base-level-space, and the state has a box with nothing on top, and the state has input that has not been examined, then make the comprehend operator acceptable, and note that the input has been examined” (Newell, 1990, p. 167). It seems highly unlikely that both of these productions should ‘fire’ within the same time scale, i.e. approximately 10ms. Time values, the number of productions per step, and the complexity of those productions have clearly been chosen to allow the total ‘reaction time’ of the models to fall within human limits found through psychological experimentation. The claim that SOAR has somehow allowed rough predictions of human reaction time is thus very unconvincing, given this ad hoc methodology. It is rather more likely that the modeller's analysis, experience with psychological results, and chosen time values allowed such predictions (Newell, 1990, pp. 274-282). In sum, there is no mention of how to *systematically* relate productions to neural firings, and worse, the few examples provided are highly unsystematic.

Though it may not be completely futile for the symbolicist to attempt to incorporate realistic time constraints into his or her model, it is undeniably more natural for this constraint to be satisfied by intrinsically dynamic models – that is, models which, in virtue of their underlying formal theory, have time constraints included. In the end, symbolicists have not convincingly described how time in their model of cognitive processes (‘model time’) systematically relates to time in a natural cognizer (‘real time’). This is important since, as Newell himself notes, “minor changes in assumptions move the total time accounting in substantial ways that have strong consequences for which model fits the data” (Newell, 1990, p. 294). This comment reflects two important conclusions of this discussion. First, time is often included in symbolicist models – symbolicists clearly took time very seriously, contrary to some characterizations. And second, there is a massive slippage between cognitive model and real cognitive system for symbolicists. This is the high price symbolicists have paid for considering time *independently* of cognitive function.

Dynamicism: Mind as motion

In the mid 1990s, a movement in cognitive science called 'dynamicism' began to flourish by arguing that these kinds of temporal limitations of symbolicism doomed it to failure (Abraham *et al.*, 1994; Busemeyer & Townsend, 1993; Port & van Gelder, 1995; Robertson *et al.*, 1993; Thelen & Smith, 1994; Tim van Gelder, 1995; T. van Gelder, 1998). The dynamicists espoused what they characterized as a diametrically opposed view, which elevated time to be the single most important constraint on good cognitive models. In doing so, they embraced a different formal theory, 'dynamic systems theory,' which is a branch of mathematics that describes time-varying behavior using sets of differential equations.

Often explicitly, the dynamicist movement was a theoretical transition back to the methods and commitments of the cyberneticists. Perhaps reflective of the cyberneticist relation to behaviorism, dynamicists tend to reject both computation and representation (Port & van Gelder, 1995; Thelen & Smith, 1994),[5] despite the fact that cyberneticists had remained largely silent on this point. As well, this concern with representation and computation may have seemed more pressing as a result of the dynamicist discontent with the symbolicist paradigm. In any case, it should be clear that the rejection of computation and representation does not follow from the adoption of dynamic systems theory as a formal means of describing their models. So it may not be surprising that the anti-representationalist stance of dynamicists is generally considered a poorly motivated aspect of dynamicism (Bechtel, 1998; Eliasmith, 2003).[6]

There have been a number of other concerns expressed with the dynamicist approach, including (Eliasmith, 1996, 2000, 2001):

1. The 'lumped' parameters (i.e. parameters that somehow summarize the underlying neural complexity) and variables in the differential equations used by dynamicists are generally not mapped to physical states of the system (except

[5] Van Gelder (1995) is quite explicit in his rejection of representation, noting that "the notion of representation is just the wrong sort of conceptual tool to apply" (p. 353) in describing dynamical systems. Similarly, Van Gelder and Port (1995) state that cognitive systems are not best understood as the result of computing over representations: "a cognitive system is not a discrete sequential manipulation of static representational structures" (p. 3).

[6] In fact, more recent work by van Gelder (1999) and others also begins to back-pedal on the earlier stricture against representation: "Dynamical models usually also incorporate representations, but reconceive them as dynamical entities (e.g., system states, or trajectories shaped by attractor landscapes). Representations tend to be seen as transient, context dependent stabilities in the midst of change, rather than as static, context-free, permanent units" (p. 244). Nevertheless, the original nonrepresentational ideal remains: "Interestingly, some dynamicists claim to have developed wholly representation free models, and they conjecture that representation will turn out to play much less of a role in cognition than has traditionally been supposed" (ibid., p. 244).

inputs and outputs). As a result it is difficult to gain independent empirical support for the models (e.g., there is no role for/relation to neural data).

2. The exemplar dynamical system, the Watt Governor (Tim van Gelder, 1995), is a typical classical control system. This means that the espoused methods are classical input/output analyses, which do not account for considerations of multiple inputs and outputs, noise, multiple loops, optimality, etc.
3. From 2., the concern arises that dynamicism will have all of the same problems that behaviorism has had (e.g., difficulties explaining cognitive behaviors not obviously linked to input states; difficulties explaining recursive processing, etc.). This concern is strengthened by the dynamicist rejection of internal representation.
4. Dynamicists restrict themselves to low-dimensional dynamical systems (in an attempt to distinguish their models from connectionist ones) (van Gelder, 1998).[7] This greatly reduces the flexibility of the system, and opens the possibility that certain natural behaviors will fall outside of the dynamicist approach.

Perhaps the most important of these limitations for this discussion is expressed by 1. Ironically, from 1. it follows that there is no explicit link between dynamicist models and the temporal constraints imposed on real cognitive systems. This is because those temporal constraints are most obvious, and best understood, at the level of single neurons and small networks of neurons. Since there is no mapping between dynamicist model parameters and the physical substrate that these models are trying to explain, it is unclear to what extent the ‘time’ in the models reflects the ‘time’ in the real system. So, while dynamicists have inherently included ‘time’ in their models, it is unclear whether it is the correct, biologically relevant, i.e., ‘real’ time. Since there is no commitment to an explicit mapping between ‘model time’ and ‘real time,’ it is up to each individual modeler to choose some mapping or other that will result in the appropriate outputs. This difficulty, of course, is reminicent of the massive slippage between cognitive model and cognitive system that plagued the symbolicists. So, similarly, this degree of arbitrariness is damaging to the dynamicist claim that they are trying to understand “cognitive phenomena, like so many other kinds of phenomena in the natural world” (van Gelder and Port 1995), given that they have provided no systematic relation between ‘time’ in their explanations and real, natural, cognitive time.

As a result of the variety of difficulties mentioned above, dynamicism, as a cognitive paradigm, has become somewhat marginalized in cognitive science. Neverthless, dynamicism has left a valuable legacy of researchers no longer being able to simply ignore temporal constraints, or assume that those constraints will somehow be taken care of after-the-fact.

[7] For instance, van Gelder (1998) states: “Another noteworthy fact about these models is that the variables they posit are not low-level (e.g., neural firing rates), but rather macroscopic quantities at roughly the level of the cognitive performance itself” (p. 619). Similarly, van Gelder and Port (1995) note that the purpose of a dynamicist model is to "provide a *low-dimensional* model that provides a scientifically tractable description of the same qualitative dynamics as is exhibited by the high-dimensional system (the brain)" (p. 28).

Connectionism in time

The place of time in connectionist modeling is much more complicated than for either symbolicism or dynamicism. This is largely because the label 'connectionism' applies to a wide variety of modeling assumptions. In general, a model is considered to be connectionist if it consists of simple computational units (nodes) connected together in large, usually parallel, networks. The units produce a numerical output based on weighted numerical input from the other nodes to which they are connected. The interpretations of such models have ranged widely, both in terms of what each unit represents, and in terms of the kinds of network topologies that are relevant for understanding the mind. The models range from atemporal localist models (e.g., Thagard, 1992), where each node represents the strength of a concept or sentence, through atemporal distributed models (Elman, 1991), where concepts are represented by the activity of several nodes combined, to (usually distributed) models whose dynamics are of central interest (Lockery *et al.*, 1990). However, it is fair to say that the core of connectionism, represented by the best known connectionist models, is atemporal (Gorman & Sejnowski, 1988; Rumelhart & McClelland, 1986; Sejnowski & Rosenberg, 1986). As a result, the case can be made that the 'spirit' of connectionism is not *essentially* dynamical. This captures at least one concern of the 'dynamicists' regarding the connectionist approach to understanding the mind (Port & van Gelder, 1995).

Nevertheless, the contrary case can be made as well, albeit for a subset of connectionist models: distributed recurrent networks. The timing of such networks is, like any dynamical system, integral to the equations describing the system. Connectionists constructing such models do not need to contrive to include time in a model of cognition, as symbolicists do. Rather, such network models naturally incorporate time constraints. Hence, Churchland and Sejnowski (1992) claim: "A theme that will be sounded and resounded throughout this book concerns time and the necessity for network models to reflect the fundamental and essential temporal nature of actual nervous systems" (Churchland & Sejnowski, 1992, p. 117) – I take this to be a supremely dynamicist sentiment.

As a result, some connectionist models clearly have the potential to be inherently temporal. A connectionist network can be, after all, "a dynamical system, meaning its inputs and internal states are varying with time; it is basically engaged in spatiotemporal vector coding and time-dependent matrix transformations" (Churchland & Sejnowski, 1992, p. 338). The main difficulty for connectionists is not whether or not they can include time, but whether they can do so in a way which can be informative of the systems being studied. To better understand this challenge, consider the vast literature on attractor networks (e.g. see Plaut & McClelland, 1993). Attractor networks are recurrent networks that, as the name suggests, evolve over time in order to exploit the existence of state space attractors (i.e. points or sets of points that are dynamically stable). However, the *particular length of time* it takes a connectionist attractor network to settle is seldom related to the time constraints imposed on real nervous systems. Rather, it is determined by how 'big' the time step that is chosen by the modeler happens to be (where a time step is the length of time it takes to complete one stage in the recurrent processing). As a result, such networks are essentially temporal, but that temporality is not linked to real organisms, i.e. it is not liked to real time. This, of course, is the same problem that, I have already argued, plagues dynamicists and symbolicists: how should ‘model time’ and ‘real time’ be systematically linked? Nevertheless, attractor networks are a useful advance over completely atemporal connectionist networks.

It is important to note that there is another subset of connectionist models that *are* directly constrained by observed temporal properties of organisms. These are the so-called ‘low-level’ connectionist models, where the nodes are mapped one-to-one onto real neurons. However, these kinds of low-level models are considered distinct enough from core connectionism, that there are unique conferences (e.g., CNS, COSYNE, etc.) and journals (e.g., the *Journal of Computational Neuroscience*, *Biological Cybernetics*, etc.) that focus on these far more biologically plausible networks. Most researchers in this domain refer to themselves as ‘computational neuroscientists,’ or ‘theoretical neuroscientists,’ and consider what they do as quite distinct from artificial neural networks or connectionism (although the historical and theoretical relations are clear). It is in these biologically plausible models where real-world dynamics become an inescapable feature of the models. It is here that there is a systematic relation between ‘model time’ and ‘real time.’ In particular, the empirically measurable time constants, voltage and current rate changes, etc., of real neurons are explicitly included in the models. So, as modelers begin to map computational units in their model networks onto computational units in biological systems (i.e., neurons), and as these model units resemble the biological units more and more, dynamics, especially the particular dynamics of natural systems, become crucial for explaining network behavior. This is hardly surprising since these modelers are now directly addressing the same phenomena that gave rise to the dynamics laden vocabulary of neuroscientists. One way of characterizing this important and unique step in understanding cognitive systems is to realize that temporal assumptions regarding the model parameters are *independently* testable assumptions. That is, neuroscientists can go to the system being described and measure those parameters directly. This is not true for firing times of productions, time courses of lumped parameters, or time steps in recurrent networks.

Unfortunately, a new problem arises for these biologically plausible networks. If this biological connectionism, like dynamicism and symbolicism, is to be a paradigm for understanding *cognitive* systems, it is essential to describe how these ‘low-level’ biological models relate to ‘high-level’ cognition. Simply including the dynamics of neurons does not explain how or why those dynamics give rise to complex, higher-level, cognitive dynamics. In general, it is fair to say that the extent to which most such models have included real time is proportional to the extent to which they are noncognitive. What is missing is a systematic method for ‘growing’ extremely complex dynamical models from these well-grounded beginnings.

Dynamic difficulties

Given the preceding discussion, it seems that the history of cognitive science teaches us three main lessons about dynamics. The first, noted most effectively by the dynamicists, is that cognitive systems are organisms embedded in natural environments to which they are dynamically coupled. As a result, it is highly unlikely that addressing the organism's cognitive behaviors independently from temporal constraints on those behaviors will result in explanatorily fruitful theories.

The second lesson is that 'model time' and 'real time' must be systematically related. It is one thing to write down a differential equation over the variable '*t*', but it is another thing to say how that '*t*' relates to the real '*t*,' observed by experimentalists. Because the mapping between 'nodes' for connectionists, or 'parameters' for dynamicsts, and the underlying neural implementation is not systematized by either paradigm, it is a mistake to suppose that time will somehow take care of itself. Despite the switch in formal theories, this problem is closely related to the mistaken assumption of symbolicists that time is somehow independent of function. The difference is that for dynamicists and connectionists the independence is more subtle. While they include 'time' variables in their models, the lack of an explicit relation between model components and the physical system being modeled means that it may well not be the right 'time.'

The third and final lesson is that, even once an explicit mapping has been made between model time and organism time, more work must be done to understand truly cognitive dynamics. This is simply a consequence of the fact that typically cognitive phenomena are the result of complex interactions between millions, if not billions, of neurons. While an explicit, systematic relation between models and physical implementation may exist at the neuron level, to make such models cognitive requires methods for 'growing' this mapping to an appropriate level of complexity.

In the remainder of this chapter, I describe a framework which shows how to resolve these remaining difficulties (see also Eliasmith, 2003).

Dynamics and the Neural Engineering Framework (NEF)

The Neural Engineering Framework (NEF) is a general theory of neurobiological systems proposed in Eliasmith and Anderson (2003). The theory consists of three quantified principles that characterize neural representation, computation, and dynamics. In this discussion, I focus on the third principle. It is stated in Eliasmith and Anderson (2003, p. 15) as:

Neural dynamics are characterized by considering neural representations as control theoretic state variables. Thus, the dynamics of neurobiological systems can be analyzed using control theory.

Though succinct, this principle makes plain how the difficulties faced by symbolicism, connectionism, and dynamicism are addressed. In short, the systematic mapping between 'model time' and 'real time' is accomplished in virtue of the fact that the representations whose dynamics are expressed by control theoretic equations are precisely neural representations. This means that the various time constants of single neurons are mapped onto appropriate time constants in model neurons. In other words, there is a one-to-one mapping between model neurons and modeled neurons, just as for the computational neuroscientific subset of connectionism. However, the NEF goes beyond standard computational neuroscience methods by providing an additional suggestion for how to write modern control theoretic equations *over these neural representations*.

Since control theoretic equations simply *are* sets of differential equations, as in dynamicism, the NEF essentially integrates the biological connectionist view with the dynamicist view of cognitive systems. The benefit is that, unlike dynamicism, the NEF sets up a systematic mapping between 'model time' and 'organism time,' and unlike standard computational neuroscience, the NEF explicitly describes the relation between neuron activity and 'higher-level' variables of the system. So, the NEF simultaneously suggests a method for building towards cognitive dynamics, while remaining responsible to single cell dynamics.

In addition, the kind of control theory adopted by the NEF, modern control theory, suffers none of the limitations of the tools used by the cyberneticists. As suggested by the dynamics principle of the NEF, modern control theory considers the *internal* states of the system (i.e., the state variables) in order to understand the dynamics of the system's output given its input. As well, modern control theory provides for the analysis of multiple input/multiple output systems and multiple-loop systems, as well as incoporating noise, optimality constraints, and adaptive control. In short, modern control theory is an excellent formalism for analyzing and synthesizing real-world physical systems – including the brain.

To better understand how this principle, and modern control theory, is applied in the NEF, let us consider a simple example. One of the most basic, and central properties of recurrent networks is their ability to extend network time constants far beyond the time constants of the individual cells comprising the network (time constants, here, measure how long a signal takes to decay). So, for instance, if we expose a single cell to a brief pulse (e.g., 1ms) of input current, there will be a more slowly decaying current in its cell body (e.g., that lasts, say, 5ms). While this intrinsic current will outlast the length of the actual input, in general it does not last much longer. However, if we take an ensemble of such cells, and connect them appropriately, we can cause a similar injection of current to the population of neurons to be effectively sustained over a very long period of time (e.g., 10s).

This property can be extremely computationally useful. For instance, it can cause a population of neurons to act like a memory, encoding information about an event that occurred in the past. As well, it can be used to accumulate information over time, tracking long-term changes. More generally, such a network acts as one of the basic temporal transfer functions, integration. Integration is so important for understanding dynamical systems, that it is *the* basic transfer function for modern control theory. The ubiquity of recurrent connections in the brain, coupled with the ease of building integrators with recurrent networks, and the importance of integrators for implementing a wide variety of dynamical behaviors suggests that neural integration may be a fundamental neural function. Indeed, the integrator has been used in models of a wide variety of neural systems including working memory (Miller *et al.*, 2003), head direction tracking (Zhang, 1996), eye-position control (Seung, 1996), the vestibular-ocular reflex (Eliasmith *et al.*, 2002), and allocentric position tracking in an environment (Conklin & Eliasmith, 2005b).

Characterizing the precise relation between integration and any one of these specific models would take us too far afield, so let us consider a generic 'neural integrator.' That is, let us assume we wish to build a neural circuit which has the properties described earlier, i.e., a circuit whose network time constant far exceeds the time constants of any of the consitutents. Employing the NEF, we first take the computational units in our model to be single neurons, whose temporal properties are matched to those of the neural system we are studying. This gives rise to a variety of single cell models whose distribution of input response functions[8] reflects the experimentally observed distribution in the relevant part of the brain. These constitute the computational elements of the model, and their dynamics are assumed to be carefully matched to the dynamics of the neural system.

Second, it is generally observed in the brain that many different cells carry information about a given set of internal or external states. As a result, we must determine how the cells in our circuit relate to the states of 'interest' to them. Again, this information can be gathered experimentally. This is a typical step in single cell physiology experiments, when neuroscientists construct what they often term 'tuning curves.' These curves determine which activity states of neurons carry information about which states of the world (e.g., a neuron in the nucleus prepositus hypoglossi is said to carry information about eye position as reflected by its tuning curve, which is a monotonically increasing firing rate as a function of eye position).[9] It is the population-wide neural representation of those states of the world that are considered state variables in our control theoretic description of the behavior of the circuit.

Third, we must express the dynamics of the circuit in control theoretic terms. Simply put, this means writing a set of differential equations that describe the overall circuit dynamics in terms of the state variables. In the case of a single variable neural integrator, we can write the integration as $x(t) = \int u(t)\,dt$, where *x* is the state variable, and *u* is the input to the circuit. As a simple control structure, this can be written as $\dot{x} = \frac{dx}{dt} = Ax(t) + Bu(t)$ where *A*=0 and *B*=1. However, because neurons have intrinsic dynamics dictated by their particular physical characteristics, we must adapt this standard control structure to a neurally relevant one. Fortunately, this can be done in the general case (Eliasmith & Anderson, 2003).

[8] Input response functions are a plot of the input current versus the resultant firing rate. This is like an input/output response function for a cell. More precisely, these curves have a temporal dimension as well, given dynamic single cell effects like adaptation. For simplicity, this will be ignored in the present example.

[9] Again, this is a simplification, since many neurons carry information about internal states, or act largely in a control capacity. This simplification serves a pedagogical purpose and does not speak to a limitation in the generality of the NEF.

Finally, we must use our characterization of single cell representation and circuit dynamics to determine the connection weights between neurons that exploit the single cell properties to realize the defined control structure. The details of the analytical methods to determine the weights are found in Eliasmith and Anderson (2003). It is also demonstrated there that the preceding steps can be carried out in the general case, i.e., for linear or nonlinear control structures, and for scalars, vectors, functions, or any combination of these under noise (see Eliasmith (2005b) for examples of each of these cases). There is no reason to suppose that this degree of generality will, in any way, be limiting to constructing models of cognitive systems.

Even in the simple integrator circuit, we can see how the difficulties faced by past methods are resolved. First, the dynamics of natural systems are mapped directly onto the dynamics of constituents of the circuit. This solves the problem faced by both dynamicists and connectionists regarding adopting natural, realistic dynamic constraints in their models. Second, the description of our model necessarily includes time, as it is written as a set of differential equations. Third, unlike computational neuroscientists, we have an explicit method for relating the activities of individual cells in the circuit to higher-level behaviors of the group of cells (e.g., integration in this case). This simple circuit, of course, does not demonstrate that the method will help build traditionally *cognitive* models. For this reason, in the next section I briefly present an application of the NEF to a more typical cognitive phenomenon.

From neurons to cognition

Fodor and Pylyshyn (1988), and more recently Jackendoff (2002), have suggested that neurally plausible architectures do not naturally support structure-sensitive computation, and that such computation is essential for explaining cognition. Notably, Fodor and Pylyshyn (1988) in particular have further argued to the that extent such architectures could be ‘forced’ into performing this kind of computation, they would turn out to be ‘merely’ implementations of symbolicist cognitive systems. For the purposes of this section, I accept that structure-sensitive processing is fundamental to understanding cognition, but show how neurally plausible architectures can support such processing in a non-symbolicist way. The specific model I present captures the context sensitive linguistic inference exhibited by human subjects in the Wason card task (Wason, 1966). To do so, the model employs biologically realistic neurons to learn the relevant structural transformations appropriate for a given context, and generalizes such transformations to novel contents with the same syntactict structure. Given the salient properties of the model, I refer to it as BioSLIE (BIOlogically-plausible Structure-sensitive Learning Inference Engine).

In the Wason task, subjects are given a conditional rule of the form “if P, then Q”. They are then shown four cards. Each card expresses the satisfaction (or not) of condition P on one side and the satisfaction (or not) of condition Q on the other. The four visible card faces show representations of `P', `Q', `not-P', and `not-Q'. Subjects are instructed to select all cards which must be turned over in order to determine whether the conditional rule is true. A vast majority of subjects (greater than 90%) do not give the logically correct response (i.e., P and not-Q). Instead, the most common answer is to select the P and Q cards, or just the P card (Oaksford & Chater, 1994). However, it became apparent that performance on the task could be greatly facilitated by changing the content of the task to be more realistic or thematic, often by making the rule a permissive one (e.g., "if someone is drinking alcohol then that person is over 21"; Sperber *et al.*, 1995). To distinguish these two version of the task, I refer to them as the 'abstract' and 'permissive' versions of the task respectively. Human performance on the Wason task is an ideal target for providing a neural model of cognition because it is considered a phenomena that can only be explained by invoking structure-sensitive processing. As a result, the task allows BioSLIE to demonstrate its ability to generalize across structures, i.e. to be systematic – an ability that many, including Fodor, Pylyshyn, and Jackendoff, take to be a hallmark of cognitive systems.

The model takes advantage of the NEF, recent advances in structured vector representations, and relevant physiological and anatomical data from frontal cortices. Since the early 1990s, there have been a series of suggestions as to how to incorporate structure-sensitive processing in models employing distributed, vector representations (including Spatter Codes (Kanerva, 1994); Holographic Reduced Representations (HRRs, Plate, 1991); and Tensor Products (Smolensky, 1990)). Few of these approaches have been used to build models of cognitive phenomena (although see Eliasmith & Thagard, 2001). However, none of these methods have been employed in a biologically plausible computational setting. Fortunately, the NEF can be employed to implement the necessary nonlinear vector computations demanded by these solutions.

In particular, BioSLIE employs 100-dimensional HRR vectors to encode linguistic structure. The details of implementing HRRs using the NEF can be found elsewhere (Eliasmith, 2004). In short, we can construct rules, like those needed to understand the Wason task, using vector multiplication and addition in a biologically plausible network. So, for instance, the rule "if *a* then *b*," or Implies(*a*,*b*), can be encoded into a single vector:

$$R = relation \otimes implies + antecedent \otimes a + consequent \otimes b\,,$$

where each variable in this equation is a 100-dimensional vector, and each such vector is represented by neural spiking. It is here, in constructing our repersentation *R* in this manner, that we avoid merely implementing a symbolicist system. This is because this representational format, being a compressed vector representation, does not explicitly include the constitutents of the representation *R* in the representation itself. As a result, the representation is non-compositional, violating a basic constraint Fodor and Pylyshyn (1988) place on symbolicist cognitive systems (see Eliasmith (2005a) for further discussion). Notably, the resulting representation, *R*, can be transformed in various ways to provide information about the contents of that vector representation. In particular, *R* can be transformed to report any of the constituents of the representation, or transformations of those consituents as demanded by a given task. It is precisely such transformations that the system must learn in performing the Wason task. In short, BioSLIE must learn how to transform *R* in different contexts (i.e., the permissive and abstract contexts) to return the appropriate elements of the structure (e.g., *a* and *not b* in the permissive case, and *a* and *b* in the abstract case).

Of course, to use this characterization of structure-sensitive processing in an explanatorily useful model, it is essential to suggest which anatomical structures may be performing the relevant functions. Only then is it possible to bring to bear the additional constraints of (and make predictions relating to) single cell physiology and functional imaging data. Figure 1 shows how BioSLIE is mapped to functional anatomy. Specifically, the network consists of: a) input from ventromedial prefrontal cortex (VMPFC) which provides familiarity, or context, information that is used to select the appropriate transformation (Adolphs *et al.*, 1995); b) left language areas which provide representations of the rule to be examined (Parsons *et al.*, 1999); and c) anterior cingulate cortex (ACC) which gives an error signal consisting of either the correct answer, or an indication that the response was correct or not (Holroyd & Coles, 2002). The neural populations that make up BioSLIE itself model right inferior frontal cortex, where VMPFC and linguistic information is combined to select and apply the appropriate transformation to solve the Wason task (Parsons & Osherson, 2001). It is during the application of the transformation that learning is also presumed to occur in an associative memory. Given this mapping to anatomy, we can appeal to work in frontal cortices that have characterized the kinds of tuning curves pyramidal cells in these areas display.

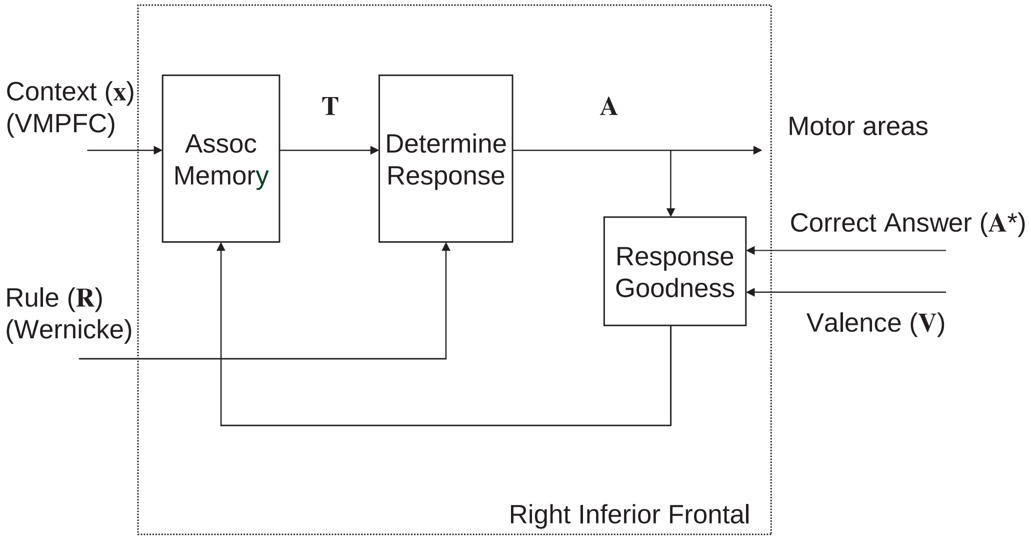

Figure 1: Functional decomposition and anatomical mapping of the model. The letters in bold indicate the vector signals in the model associated with the area.

To perform the needed HRR vector operations, learning, and so on, BioSLIE further decomposes this high-level functional mapping into neural subsystems responsible for these tasks. The resulting set of subnetworks is shown in figure 2, which is a model that consists of ten interconnected neural populations, for a total of approximately seventeen thousand neurons.

When run, the model is able to reproduce the typical results from the Wason task under both the abstract and permissive contexts (not shown). Simply put, this means that the model is taught, and successfully reproduces the transformation ‘if a then b → {b, a}’ in the abstract context, and the transformation ‘if a then b → {~b, a}’ in the permissive context. So, when the context signal is switched, the model applies a different transformation, as expected. The point of mentioning these results is simply to emphasize that this is done using biologically plausible neurons in a complex neural network, not by having a computer peform these logical transformations directly. And, while simple, this model does show rudimentary structural transformations. This, however, is not enough to support the claim that the model is structure sensitive. The obvious concern is that the model is simply 'memorizing' a mapping it has seen (i.e. it is constructing a look-up table). If this were true, the model would not truly be generalizing over the appropriate syntactic structures, as demanded by systematicity.

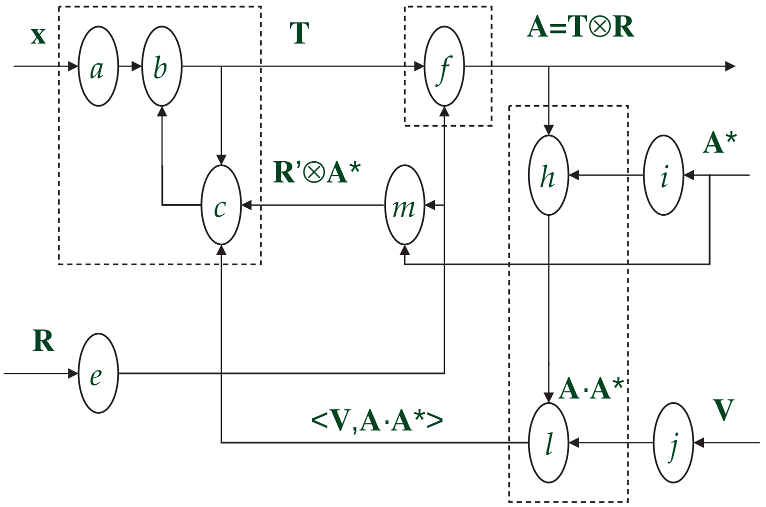

Figure 2: The complete network at the population level. The lower case letters indicate populations of approximately 2000 neurons each. Upper case letters indicate the signals being sent along the relevant projections. The dotted boxes indicate how this diagram relates to the functional decomposition of figure 1, and hence the anatomical mapping discussed earlier.

To demonstrate that the network is truly learning a language-like transformation in a context, figure 3 shows that it does in fact generalize learned, structure-sensitive transformations to unfamiliar contents (i.e., "if someone votes then that person is over 18") in a familiar context (i.e., the permissive context). This demonstrates that the system has learned a systematic syntactic regularity. That is, it can transform novel structured representations based solely on the syntax of the representation.

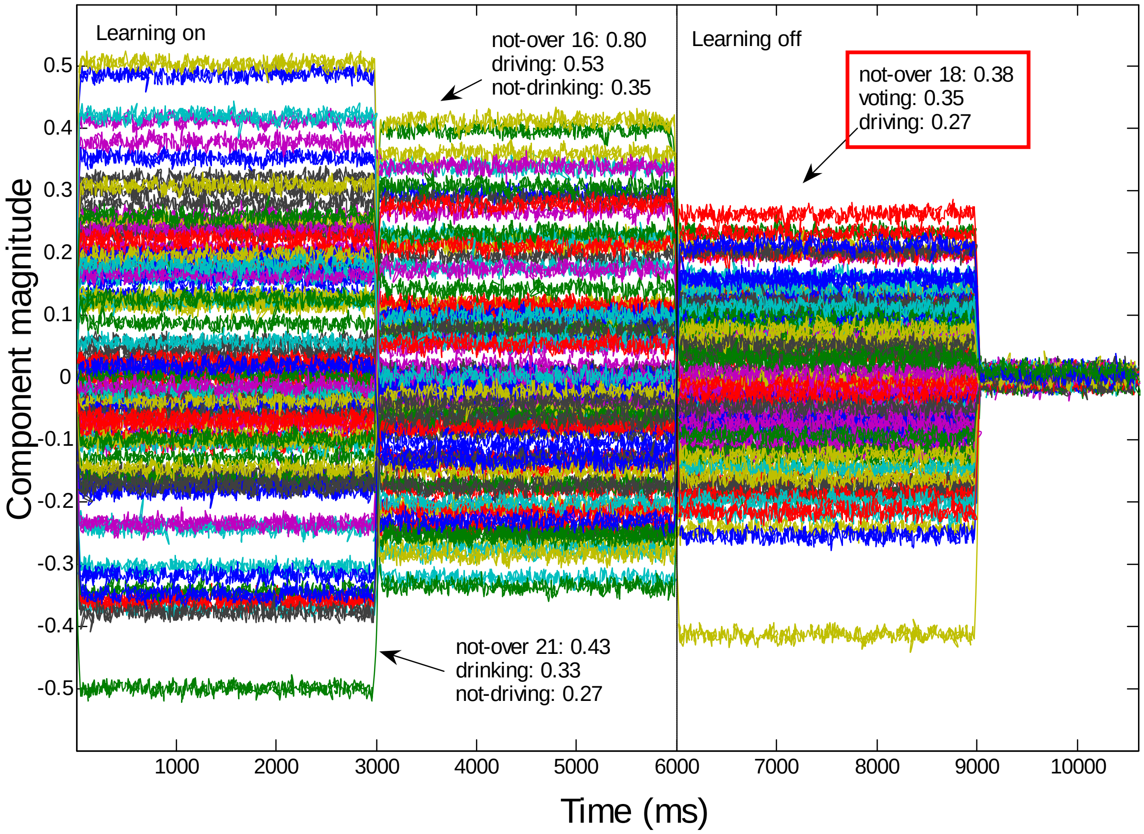

Figure 3: Generalization across different rules in the same context. Each line indicates the value of one dimension of the 100-dimensional vector encoded in neural spiking in the *f* population from figure 2. The top three similarity results of each transformation are shown, to demonstrate that simple thresholding results in the correct answer. See text for further discussion.

Let us consider this figure in more detail. In the simulation the ‘permissive’ context signal is kept constant and there are three separate rules that are presented to BioSLIE. While learning is on, the rules Implies(*drinking-alcohol*, *over-21*) and Implies(*driving*, *over-16*) along with their expected answers are presented to the network. The learning is then turned off, and it is presented with the novel rule Implies(*voting*, *over-18*). Notably, since the context is the same in the novel case as for the previous examples, the same transformation should be applied. Indeed, BioSLIE infers that *voting* and *not_over-18* are the expected answers (i.e., the cards that need to be checked to ensure the rule is not violated). In the last quarter of the simulation, no rule is presented and thus no answer is produced (i.e., all similarity measures are very low).

This graph thus demonstrates that BioSLIE is systematically processing language-like structures with biologically realistic computational components. As a result, not only does it provide an explicit counterexample to Fodor, Pylyshyn, and Jackendoff’s claims, it also demonstrates how the NEF can relate single neuron dynamics to the dynamics of cognitive behavior. Admittedly, BioSLIE most directly addresses the issue of how the appropriate representations and transformations for accomplishing cognitive tasks can be understood in a neurally plausible way. It does not directly map on to the observed dynamics of human performance on the Wason task (the model is much faster, although it is appropriately constrained by single neuron dynamics). This, no doubt, is because far more than the few brain areas modelled by BioSLIE are employed by human subjects to perform the task. Nevertheless, timing constraints on certain aspects of the task can be inferred from BioSLIE's performance (e.g., minimum transformation times). And, more importantly for this discussion, the methods provided are general enough to address a wide variety of cognitive tasks in a way that directly incorporates underlying neuro-dynamical constraints.

Embeddedness and the NEF

To this point I have discussed how the NEF relates low-level neural dynamics, with higher-level circuit dynamics, and demonstrated that it is possible to build rudimentary cognitive systems using the NEF. Earlier, I briefly touched on the shared inspiration for taking dynamics seriously and for being concerned with the embeddedness, or situatedness, of cognitive agents. Here I want to discuss what, if any, consequences the NEF has for our understanding of cognitive embeddedness.

Note that for some dynamicists, taking dynamics seriously means holding a fairly strong embedded view: "In this vision, the cognitive system is not just the encapsulated brain; rather, since the nervous system, body, and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three" (Tim van Gelder, 1995, p. 373). For dynamicists, then, a distinction between the system and the system's environment becomes very difficult – system boundaries become obscure. Dynamicists often claim that this result is a unique strength of the dynamicist approach, and an accurate reflection of the true state of cognitive systems (van Gelder and Port 1995). Similarly, those focused on the situatedness of cognitive systems have argued that the traditional boundaries between an agent and its environment, provided by the skin, are unreasonably hegemonic and that, instead, "the mind extends into the world" (Clark & Chalmers, 2002, p. 647).

I suspect that such conclusions are misguided, and we can turn to the NEF to see why. As discussed, the NEF adopts modern control theory as a means of specifying dynamics. Control theory, as opposed to dynamic systems theory, has a number of benefits for describing cognitive systems. First, control theory explicitly acknowledges system boundaries, in virtue of identifying state variables with subsystems of the overall system of interest. Second, control theory explicitly introduces the central notion of 'control' and related notions such as 'controlability'. These notions help underwrite distinctions between systems whose dynamics are fixed or otherwise independent of one another. And finally, control theory has its roots in engineering, a discipline concerned with implementational aspects of physical systems, including noise and other component limitations. These concerns contrast with dynamic systems theory whose roots are in mathematics. This is not to say that either control theory or dynamic systems theory is somehow more mathematically powerful, but rather it is to point out that the methods have different emphases, one of which is more appropriate for understanding physically realized, natural, cognitive systems.

Let us consider each of the first two benefits in more detail. The importance of acknowledging system boundaries cannot be overstated when pursuing system analysis. Decomposition of complex systems is essential for our understanding of such systems, whether they be biological, ecological, economic, meteorological, or what have you. As Bechtel and Richardson (1993) have argued at length, "a mechanistic explanation identifies these [system] parts and their organization, showing how the behavior of the machine is a consequence of the parts and organization… A major part of developing a mechanistic explanation is simply to determine what the components of a system are and what they do" (p. 17-18). Blurring, shifting, or removing system boundaries, as dynamicist and embedded agent theorists often advocate, is seriously detrimental to making progress in our explanations of such systems. This is especially true if there are no theoretical principles for determining which shifting or removing of boundaries is justified, and which is not. As a result, considering cognitive systems (constituted by brains, body, and world) as a "single unified system," is both impractical and uninformative from a scientific point of view – it in no way helps determine what the components are. Notice that advocating the identity of system components does not imply that such decompositions should not be 'reassembled' for explaining certain properties. Rather, it is the observation that to explain a large, complex system requires identifying and explaining both its subsystems and their interactions. And, to do that, those subsystems must *themselves* be identified and well-understood.

This leads naturally to the second point, that the introduction of the notion of 'control' helps to categorize different kinds of subsystems. A typical dynamical system in control theory consists of a plant and a controller. The plant is a physical system whose inputs we would like to change in order to result in particular outputs from that system. The controller plays the role of producing the necessary inputs to result in those particular outputs. This basic distinction is one which helps us understand the different roles brain, body, and world play in an overall explanation of a behaving agent in an environment. With this distinction, we can see what is special about the brain. We have fairly good physical theories that can be used to explain the kinetics and dynamics of bodies and of the world. However, we have little idea how to understand the more complex dynamics found in the brain. As a result, it is natural to consider the brain as the controller of the body as a plant, together acting as controller for the environment as a plant (see figure 4).

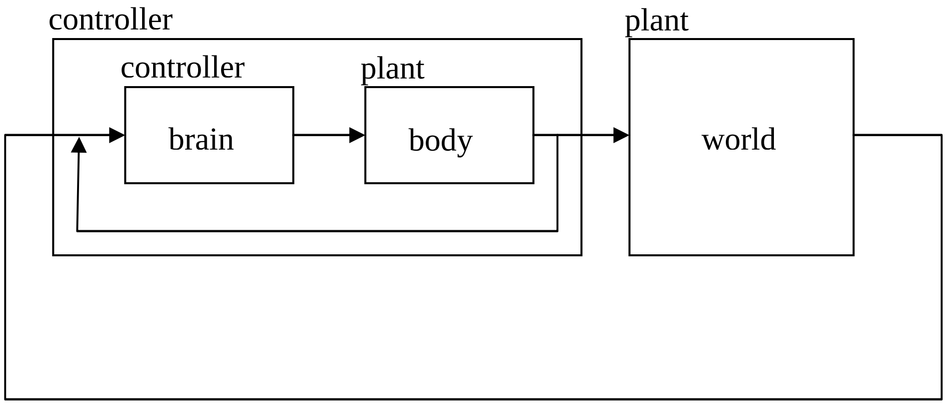

Figure 4: Brain, body and world as controllers and plants. Drawing such system boundaries, and making plant/controller distinctions makes clear the differences between subsystems and their interactions.

Our goal in understanding a cognitive system is to elucidate the qualitatively different dynamics internal to the brain. The most obvious differences are the speed of information flow (i.e., bandwidth), and the degree and kind of coupling. Because bodies have mass, they tend to slow down the transfer of information to the world from the brain (i.e., they effectively act as a low-pass filter). However, no such impediment to information flow exists between brain areas. This results in a huge difference between the kinds of coupling that can be supported between brain subsystems and between the brain and the external environment. In short, interactions with the environment are slower than intra-brain interactions. I find it rather ironic, or perhaps surprising, that researchers who embrace the importance of dynamics for understanding cognitive function, and who argue that differences in dynamics are cognitive differences (when confronting symbolicists; van Gelder (1998, p. 622)), then suppose that differences in dynamics between brain-brain and brain-world interactions can be overlooked when arguing for embeddedness (Clark & Chalmers, 2002, p. 648; van Gelder, 1995, p. 373). I think it is much better to consistently claim that differences in dynamics *often* result in distinct properties and behaviors. If we adopt that view, it becomes clear that the suggestion that "nothing [other than the presence of skin] seems different" between brain-brain and brain-world interactions (Clark & Chalmers, 2002, p. 644), is plainly false.

I should note that I do not want to suggest that determining the appropriate system boundaries will be an easy task (nor that it stops at the skin). Indeed, it is unclear whether or not we will be able to identify general, consistent principles for identifying system boundaries. Nevertheless, it is essential to realize that this is a task worth pursuing, and that simply blurring systems over boundaries, or suggesting that such boundaries do not really exist is bad for both practical (i.e., trying to do science) and theoretical (i.e., appropriate conceptual application) reasons.

Dynamics + control = cognition

It is important to take the critical considerations of this paper in their appropriate context. While I have expressed serious concerns with both a dynamicism and embedded approaches to understanding cognitive systems, it should be clear that the positive view I have espoused is highly sensitive to the concerns which gave rise to these positions. The NEF undeniably draws inspiration from dynamicism, as it includes at its core an acknowledgment of the importance of time for understanding natural cognitive systems. While the NEF rejects the noncomputationalism and antirepresentationalism of dynamicism, it does so in a way that is consistent with dynamicist arguments against the symbolicist treatment of time.

As well, the fundamental insights of those interested in the embeddedness of cognitive systems is not lost in the NEF. Characterizing the brain as a control system means understanding the dynamics of its inputs and its coupling to the environment. However, I have suggested that this can be done in such a way that traditional distinctions between brain, body, and world are preserved. In other words, consideration of ecological (i.e. 'real') operating environments is imperative for trying to comprehensively understand a dynamical system interacting with that environment. This is true regardless of how that system might be broken into subsystems. In fact, there are good reasons, even dynamical reasons, for performing a decomposition consistent with traditional boundaries. It is evidently a mistake, then, to rule out decomposition merely because of dynamic coupling. Unfortunately, this seems to have been the tendency of those espousing the embodied, embedded, and extended views of cognition.

In sum, the intent of the NEF is to provide a suggestion as to how we might take seriously many of the important insights generated from cognitive science: insights from symbolicists, dynamicists, and connectionists. I have argued that it embraces realistic neural dynamics, can help us understand high-level cognition, and is consistent with traditional boundaries between brain, body, and world. I suspect it is far from a complete theory, but perhaps it is a useful start.

References

Abraham, Abraham, & Shaw. (1994). *Dynamical systems for psychology*.

Adolphs, R., Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1995). Neuropsychological approaches to reasoning and decision-making. In A. R. Damasio, H. Damasio & Y. Christen (Eds.), *Neurobiology of decision-making*. New York: Springer Verlag.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence, 48*, 57-86.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist challenge in cognitive science. *Cognitive Science, 22*, 295-318.

Bechtel, W., & Graham, G. (Eds.). (1999). *A companion to cognitive science*. London: Blackwell.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432-459.

Churchland, P. S., & Sejnowski, T. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Clark, A., & Chalmers, D. (2002). The extended mind. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings*: Oxford University Press.

Conklin, J., & Eliasmith, C. (2005). An attractor network model of path integration in the rat. *Journal of Computational Neuroscience, 18*, 183-203.

Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. *Philosophical Psychology, 9*(4), 441-463.

Eliasmith, C. (2000). Is the brain analog or digital? The solution and its consequences for cognitive science. *Cognitive Science Quarterly, 1*(2), 147-170.

Eliasmith, C. (2001). Attractive and in-discrete: A critique of two putative virtues of the dynamicist theory of mind. *Minds and Machines, 11*, 417-426.

Eliasmith, C. (2003). Moving beyond metaphors: Understanding the mind for what it is. *Journal of Philosophy, 100*(10), 493-520.

Eliasmith, C. (2004). Learning context sensitive logical inference in a neurobiological simulation. In S. Levy & R. Gayler (Eds.), *AAAI fall symposium: Compositional connectionism in cognitive science* (pp. 17-20): AAAI Press.

Eliasmith, C. (2005a). Cognition with neurons: A large-scale, biologically realistic model of the Wason task. In G. Bara, L. Barsalou, and M. Bucciarelli (Eds)., *Proceedings of the 27 th Annual Meeting of the Cognitive Science Society*. Stresa , Italy.

Eliasmith, C. (2005b). A unified approach to building and controlling spiking attractor networks. *Neural Computation, 17*(6), 1276-1314.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Eliasmith, C., M. B. Westover, & Anderson, C. H. (2002). A general framework for neurobiological modeling: An application to the vestibular system. *Neurocomputing, 46*, 1071-1076.

Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science, 25*(2), 245-286.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. In D. Touretzky (Ed.), *Connectionist approaches to language learning* (pp. 91-122). Dordrecht: Kluwer.

Finger, S. (2000). *The minds behind the brain: A history of the pioneers and their discoveries*: Oxford University Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

Freudenthal, H. (1970-1990). Norbert weiner. In C. C. Gillespie (Ed.), *Dictionary of scientific biography*. New York: Scribners.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62*, 324-341.

Gorman, R. P., & Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks, 1*, 75-89.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology, 117*, 500-544.

Holroyd, C., & Coles, M. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-rleated negativity. *Psychological Review, 109*, 679-709.

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London), 160*, 106-154.

Hull, C. (1935). The conflicting psychologies of learning − a way out. *Psychological Review, 42*, 491-516.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*: Oxford University Press.

Kanerva, P. (1994). The spatter code for encoding concepts at many levels. In M. Marinaro & P. G. Morasso (Eds.), *Proceedings of the international conference on artificial neural networks* (Vol. 1, pp. 226-229). Sorrento, Italy: Springer-Verlag.

Lockery, S., Fang, Y., & Sejnowksi, T. (1990). A dynamical neural network model of sensorimotor transformation in the leech. *Neural Computation, 2,* 274-282.

Miller, P., Brody, C. D., Romo, R., & Wang, X. J. (2003). A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cerebral Cortex, 13*, 1208-1218.

Mindell, D. (1995). Engineers, psychologists, and administrators: Wartime control systems research, 1941-1945. *IEEE Control Systems Magazine, 15*(4), 91-99.

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

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*(4), 608-631.

Parsons, L., & Osherson, D. (2001). New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cerebral Cortex, 11*, 954-965.

Parsons, L., Osherson, D., & Martinez, M. (1999). *Distinct neural mechanisms for propositional logic and probabilistic reasoning.* Paper presented at the Proceedings of the Psychonomic Society Meeting.

Plate, A. (1991). *Holographic reduced representations: Convolution algebra for compositional distributed representations.* Paper presented at the Proceedings of the 12th International Joint Conference on Artificial Intelligence.

Plaut, D. C., & McClelland, J. L. (1993). *Generalization with componential attractors: Word and nonword reading in an attractor network.* Paper presented at the Proceedings of the 15th Annual Conference of the Cognitive Science Society, University of Colorado.

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

Rieke, F., Warland, D., de Ruyter van Steveninick, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Robertson, S. S., Cohen, A. H., & Mayer-Kress, G. (1993). Behavioural chaos: Beyond the metaphor. In L. B. Smith & E. Thelen (Eds.), *A dynamic systems approach to development: Applications* (pp. 120-150). Cambridge: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of english verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 216-271). Cambridge MA: MIT Press.

Sejnowski, T. J., & Rosenberg, C. R. (1986). Nettalk: A parallel network that learns to read aloud. *Cognitive Science Quarterly, 14*, 179-211.

Seung. (1996, November 1996). *How the brain keeps the eyes still.* Paper presented at the National Academy of Science USA, Neurobiology.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence, 46*, 159-217.

Sperber, D., Cara, E., & Girotto, R. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31-95.

Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.

Thagard, P. (1996). *Mind: Introduction to cognitive science*. Cambridge, MA: MIT Press.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action* (Vol. 2). Cambridge: MIT Press.

van Gelder, T. (1993). What might cognition be if not computation? *Cognitive Sciences Indiana University Research Report 75.*

van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy, XCI*(7), 345-381.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences, 21*(5), 615-665.

van Gelder, T. J. (1999) Dynamic approaches to cognition. In R. Wilson & F. Keil ed., *The MIT Encyclopedia of Cognitive Sciences*. Cambridge MA: MIT Press, 244-6.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology*. Harmondsworth: Penguin.

Watson, J. (1913). Psychology as the behaviorist views it. *Psychological Review, 20,* 158-177.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. New York: John Wiley & Sons, Inc.

Wiesel, T. N., & Hubel, D. H. (1963). Effects of visual deprivation on morphology and physiology of cells in the cat's lateral geniculate body. *Journal of Neurophysiology, 26*, 978-993.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience, 16,* 2112-2126.