TOPICS
TOPICS IN COGNITIVE SCIENCE

This article is part of the topic "Levels of Explanation in Cognitive Science: From Molecules to Cultures," Matteo Colombo and Markus Knauff (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/early view

The Dynamics of Neural Populations Capture the Laws of the Mind

Gregor Schöner

Theory of Cognitive Systems, Institute for Neural Computation, Ruhr-Universität Bochum

Abstract

What would it mean to explain the mind in neural terms? Neural accounts of the mind are often sought in a reductionistic spirit in which neural mechanisms explain cognition. Because an individual's thoughts and behaviors are not reproducible without careful control of task, stimulus, and behavioral history, laws of the mind are the currency of psychology. Reduction may thus have to take the form familiar from physics: deriving macroscopic laws from microscopic laws. I argue that the metaphor of reduction from non-equilibrium physics may be the most appropriate. Macroscopic patterns of neural activity, which cause behavior and thought, are slow dynamical variables that dominate the fast microscopic dynamics of individual neurons and synapses. I outline a theoretical framework in which strongly recurrent neural networks, described by neural dynamics, generate neural representations as attractor states that are embedded in low-dimensional feature spaces. Instabilities of these states are instrumental in decision-making and the generation of sequences of mental states that are the basis for higher cognition. Networks of such neural population dynamics form neural cognitive architectures that capture the laws of the mind.

Keywords: Dynamical systems; Neural dynamics; Neural populations; Embodied cognition; Cognitive architectures; Foundations of cognitive science; Sequence generation

1. Introduction

In cognitive science, we try to understand the mind. Abstract, computational theories of the mind have long been dominant. Neurally grounded accounts must explain how the mind emerges from the brain, is shaped by development, and constrained by evolution. At which level of description these questions can be answered is an empirical issue. Here, I argue that the coupled dynamics of activity in populations of neurons cause behavior and thought. These dynamics reflect the invariance of mental processes against changes in sensory and motor conditions, while also capturing their intermittent coupling to the sensory and motor surfaces. A central issue in this theoretical perspective is how the sequential organization of processing that is characteristic of all thinking and acting may emerge from state- and time-continuous neural dynamics. I will sketch how attractor states of neural dynamics serve as representations of mental states and how sequential processing emerges from their instabilities.

2. Laws, prediction, levels, and reduction

Before diving into this specific proposal, it is important to be clear about what "under-standing," "explaining," and "capturing" might mean here. What are "levels of descrip-tion" and in which sense may mental phenomena be "reduced" to neural phenomena?

In his autobiographical notes, Richard Feynman recounts an episode from his childhood in which he asks his father why the ball on his little cart jumps backwards when he pulls the cart forward (opening chapter of Feynman, 1988). His father told him, he recalls, that the ball was not actually jumping backwards but was just staying in place. Why, his father said, is a big mystery. Nobody knows. It is called "inertia." I have always found this anecdote striking. Feynman's father correctly invokes a physical *law* to account for an observation. But he also points out that naming the law, and thus recognizing it as a general principle, does not explain the observation nor answer "why" questions.

Laws are empirically validated by testing their *predictions*. They have an "if" portion and a "then" portion, the prediction. Laws of motion, for instance, predict the future states of systems given their initial state. The direction of prediction is not unique. Logically, one may "predict" the past from the present, for instance. The predictive power of laws varies considerably. Newton's gravitational law is famous, because its "if" clause is very broad, so that many different systems are covered, including falling apples and planets going around the sun. The narrower the "if" clause, the less powerful the prediction. For instance, a plant may be identified as belonging to a species by paying attention to a small set of features (about 7–9). Predictions about further features can then be made. But those predictions arise merely because someone took this kind of plant apart before and described its features. And those features are reproducible from specimen to specimen. Such narrow forms of prediction amount to *description* (see Chapter 5 of Feynman, 1967, for a discussion that uses geology rather than botany for illustration).

On the other hand, powerful, general laws reveal a lack of constraints, as many different ways to fill the "if" clause are possible. These unconstrained variations are not reproducible and cannot be predicted (Rosen, 1991). Kepler famously searched for a law that would constrain which planetary orbits are possible. He failed and today we say those orbits merely reflect the initial conditions of the solar system. Any other set of orbits could have arisen and different orbits arise in other planetary systems around other suns. So, when certain facets of biological systems are highly reproducible, they are also highly constrained. Given such constraints, extensive description makes sense, as reflected in the tradition of the biological sciences.

Psychology, cognitive science, and neuroscience lie close to a fault line between these two kinds of traditions, the general law oriented tradition of physics, and the description oriented tradition of biology. Human behavior is not predictable in any practical sense, as individual history, the environment, and the moment-to-moment inner state of people's mind matters. But certain behavioral laws can be discovered when conditions are controlled and a task constrains people's freedom. So psychology has had an affinity to the law orientation of physics from its very beginnings in Helmholtz's psychophysics. In contrast, the nervous system in its structure, morphology, its cellular and molecular mechanisms, appears to be more strongly reproducible, so that the descriptive approach prevalent in much of modern neuroscience is useful.

This polarization of a law-oriented versus a description-oriented outlook is a caricature, of course. Neuroscientists are very much interested in general principles and psychologists describe patterns of behavior. The reshaping of the nervous system by experience highlights how the two domains are intertwined. Still, the difference in scientific practice and attitude is palpable. Minimal models that lose strength as detail is added (Batterman, 2002) are sought in law-oriented work, while simulation models that include all known detail are sought in description-oriented work (Keller, 2002). The question of how the mind may be "explained" by neural mechanism uncovers these differences. What would count as such an "explanation" and do such neural explanations reduce the mind to neural mechanisms?

"Explanation" as used in common language does not have an obvious meaning in physics. The logical connections that sometimes link different laws of nature might come closest to such a notion. For instance, Newton's law of gravity implies both Galileo's law of free fall and Kepler's laws of planetary orbits. In common language: "Why do stones fall? Because of gravity." Physicists tend not to emphasize explanation, however. One reason is that the direction of logical derivation is typically not unique. It is not clear that there is a unique set of axioms from which all laws of physics can be derived (Laughlin & Pines, 2000).

This is related to the key problem of *reduction* in which the microscopic explains the macroscopic. The most influential paradigm of reduction comes from the statistical mechanics of equilibrium systems (Landau & Lifshitz, 1979) in which atoms or molecules at the microscopic level explain the physical properties of condensed matter at the macroscopic level. The microstates are characterized by a universal law of motion, described by the Hamiltonian of electrons and nuclei, for instance. Through their initial

conditions, the micro-states vary in combinatorial multitude, described by a statistical distribution. Macrostates are computed as expected values of the distribution in the thermodynamic limit of infinitely large systems. Reduction in this sense does not imply that the macroscopic law and the concepts in terms of which it is framed are eliminated. The viscosity of fluids, for instance, may be derived from the molecular level, but viscosity does not disappear as a concept nor is it a property of the molecules themselves (see, e.g., Chapter 11 of Feynman, 1972).

This program of reduction has been especially successful around the so-called critical points, at which the microscopic details of a system do not matter so that the macroscopic laws are universal (are the same for different materials, Stanley, 1999; Wilson, 1979). Universality leads to an intrinsic limitation of reduction, however (Laughlin & Pines, 2000). Because microscopic details do not matter, the macroscopic law does not actually constrain the microscopic law. Different microscopic laws may lead to the same macroscopic phenomenon.

Less well known is a different metaphor of reduction that comes from the physics of non-equilibrium systems that are characterized as stochastic dynamical systems (Haken, 1983). The laser is a famous example (Haken, 1970). In normal (incoherent) light, the wave packets of individual photons vary randomly in phase. In coherent laser light, these wave packets are synchronized (phase-locked), leading to macroscopic temporal order. The microscopic level consists of atoms that emit photons when they transition from an excited to the ground state. Atoms emit photons spontaneously at random times, leading to random phases between photons. Atoms may be stimulated to emit a photon by an incoming photon. The emitted and the stimulating photon are synchronized. They may stimulate other atoms to emit photons, and so on, leading to an avalanche of synchronized photons, that is, to coherent laser light.

Atoms and photons are "fast," that is, have a short life time. The transition to lasing occurs, when the lifetime of the synchronized photons becomes long. This happens when two mirrors are aligned so that a standing wave of just the right wavelength can build up. Photons that are synchronized with the standing wave have a long lifetime, bouncing back and forth between the two mirrors. The standing wave consists of many such synchronized photons and its amplitude is the slow, macroscopic variable. The slow variable "enslaves" the fast variables: The atoms emit photons as dictated by the standing wave. In fact, the fast variables can be eliminated from the underlying dynamical system, leading to a macroscopic law of motion that depends only on the slow variables. This is a radical reduction of the number of variables (from ca. 10^{22} atoms to 1 laser mode amplitude). In the theory of dynamical systems, this reduction runs under the label "center manifold theorem" (Perko, 2001). In physics, it is known as the "slaving principle" (Haken, 1983), or "adiabatic elimination" (Schöner & Haken, 1987). Clearly, this form of reduction does not eliminate the macroscopic level at all; it actually eliminates the microscopic level!

Some microscopic degrees of freedom are not dictated by the macrostate. For instance, after emitting a photon, the atoms need to be pumped back into the excited state. This happens asynchronously by separate processes and may lead to many different

microscopic states in which different sets of atoms are in the excited state at any moment in time. This variation in the microstate leaves the macrostate invariant, a form of dynamic degeneracy. I will argue that there are analogous forms of degeneracy in neural dynamics.

3. Toward a neurally grounded theory of cognition

One way to frame the relationship between the mind and neural mechanism is to think of behavior and cognition as a "high" level of description, while neural mechanisms lie at a "low" level of description. Clearly, neurons and their synapses provide a microscopic level of description compared to activation patterns in populations of neurons or whole brain networks, and to the physiological states across the entire organism that accompany behavior. Explaining the mind in neural terms amounts, in this framing, to a reduction in which the laws of behavior and thinking are to be derived from neural laws. That reductionist perspective is largely axiomatic in neuroscience and is debated in psychology and cognitive science. Quite often, I believe, is this reduction thought of in terms of the equilibrium physics metaphor: The macroscopic patterns of neural activity that instantiate thought and behavior arise from the microscopic neural activity through some form of statistical averaging. This is formalized, for instance, when mean field neural models are derived from spiking neural network models (Gerstner, Sprekeler, & Deco, 2012).

This metaphor does not actually make biological sense. It implies a form of redundancy, in which many neurons and synapses can be eliminated by averaging. Neural activity is extraordinarily costly to the organism (Raichle & Gusnard, 2002). Keeping neurons around only to average their activity is wasteful. And although neuronal activity is variable, that variability is functional and can be modulated (Stein, Gossen, & Jones, 2005). Also, this picture is essentially static, while the mind and the brain are essentially dynamic.

The non-equilibrium physics metaphor of reduction may be much more appropriate. In that picture, individual neurons and their synapses are fast dynamical variables evolving on the time scale of milliseconds. The states of neural populations, of brain networks, or of the organism as a whole evolve on slower time scales of hundreds of milliseconds to seconds (estimated, for instance, from reaction times, movement times, the time needed to switch behavior, or the time scale on which correlations in neural populations decay [Murray et al., 2014; Runyan, Piasini, Panzeri, & Harvey, 2017]). The slower macroscopic activation states of populations dictate the faster microstates of individual neurons. Microstates may still vary in ways that are not reflected at the macroscopic level, and are thus not relevant to behavior and cognition (a form of neural degeneracy; Price & Friston, 2002).

In the following, I want to outline a theoretical program that is aimed at understanding the laws of behavior and cognition in ways that are grounded in neurophysiology, but are reductionist only in that second sense.

Hypothesis 1   Neural populations are a privileged level of description to understand cognition and behavior.

There is no a priori reason why the activity of individual spiking neurons and the associated cellular or molecular mechanisms would form the level of description at which cognition and behavior must be understood. Neurons, synapses, and certain of their biophysical properties are necessary for any cognition to arise (Wang, 1999). But so are biomolecules, or even atoms and nuclei. Few of us believe that nuclear forces or even quantum-gravity are relevant levels of description to understand the mind (Penrose's speculations not-withstanding). The most microscopic level is not necessarily the most fundamental. At which level of description the *laws* of behavior and cognition can be understood is an empirical question.

There is an emerging consensus that the activity of small populations of neurons in the brain shows the best correlation with behavior (Cohen & Newsome, 2009; Deadwyler & Hampson, 1995; Georgopoulos, Kettner, & Schwartz, 1988; Rouse & Schieber, 2018). Ultimately, this is due to their patterns of connectivity to the sensory and motor surfaces. When neurons are tuned with respect to dimensions of a perceptual or motor task, their activation levels form *distributions of population activation* (Erlhagen, Bastian, Jancke, Riehle, & Schöner, 1999), from which behavior can be predicted. This is what efforts to "read the mind" in brain–machine interfaces exploit (Schwartz, 2004).

Evidence that population activation causes behavior comes from the observation that perturbations of population activity change behavior (Groh, Born, & Newsome, 1997; Lee, Rohrer, & Sparks, 1988). I argue that the population level is privileged in that it causes behavior and mental states, while the faster, more microscopic level of individual spiking neurons can be eliminated adiabatically.

Hypothesis 2   Neural dynamics at the population level generate stable states.

A strong constraint for neural theories of cognition comes from the fact that most thinking and acting is not directly driven by sensory input. We make three saccades per second, implying that new sensory information comes in every 300 ms. This is not reflected in our visual experience, which is largely invariant under those drastic shifts of input. Across all modalities, sensory input is inherently transient in nature, living on about that same time scale. The perceptual and cognitive phenomena that depend on such input live on a much longer time scale. Voluntary movement and all actions are, of course, generated by neural activity that arises from within the nervous system. Thinking is not typically caused by a sensory stimulus alone, but unfolds as activity is generated and steered within our brain.

So, most functionally relevant neural activity is generated within the brain's neural networks, not induced by outside stimulation. Behavior and cognition must emerge, therefore, from neural networks that are strongly recurrent, able to create or sustain activity in the absence of external input. This insight is consistent with the observation of quantitative neuroanatomy that much more brain connectivity is part of recurrent loops than part

of a single feed-forward path (Braitenberg & Schüz, 1991). Understanding such networks requires a temporal dimension in which current inputs from recurrent connections derive from earlier outputs to those recurrent connections. Because spiking events in neural populations are not perfectly synchronized, the evolution of activity in strongly recurrent networks is best described in continuous time, asynchronously sampled by discrete spike times. Dynamical systems theory provides the mathematical language for doing so (Grossberg, 1977).

In neural dynamics, the resting state of a recurrent neural network is an asymptotically stable state. Its stability ultimately derives from the dissipative properties of membrane dynamics (see Chapter 1 of Schöner, Spencer, & DFT Research Group, 2016). The stable resting state resists change: It is recovered after small transient inputs.

Given that the activation of populations of neurons evolves on a time scale of 100–350 ms (Murray et al., 2014), the coherence of experience on the time scale of seconds and minutes implies that neural activation patterns other than the resting state must persist over such longer time scales. This requires mechanisms to keep activation from decaying and to stabilize activation against competing inputs. Neurons that form part of the same macrostate must, in effect, be excitatorily coupled to stabilize activation against decay. Neurons that form part of competing activation patterns must, in effect, be inhibitorily coupled to stabilize activation against distractors.

That pattern of coupling is common for neural populations that encode low-dimensional metric spaces. For instance, in much of the visual system, neurons receive input from portions of the retinal surface, as characterized by their receptive fields. Populations span the two-dimensional visual array by their different receptive fields. Neurons whose receptive fields are close to each other in retinal space are found to interact excitatorily while neurons whose receptive fields are further apart interact inhibitorily (Jancke et al., 1999). Such low-dimensional neural representations may take the form of topographical maps in some brain regions, but the anatomical arrangement of the neurons along the cortical or subcortical surfaces does not matter. What matters is the pattern of recurrent connectivity in which neurons with similar tuning are effectively coupled excitatorily while neurons with dissimilar tuning are coupled inhibitorily. This pattern of connectivity within a population ("neural interaction") is the foundation of neural field theory (Amari, 1977; Coombes, beim Graben, Potthast, & Wright, 2014; Wilson & Cowan, 1973).

Hypothesis 3   The neural representations on which cognition is based are low-dimensional.

The notion of low-dimensional feature representations taps into the concept of *invariance*, that is, the idea that behavior and thinking are independent of the precise sensory information that supports perception or the precise motor commands needed to bring about action. There are diverse sources of evidence supporting this hypothesis. That sensory information obtained from the real word is highly constrained by the structure of the environment is the central insight of ecological psychology (Gibson, 1979). Diving gannets, for instance, fold their wings when the visual expansion rate reaches a threshold

(Lee & Reddish, 1981). That visual expansion rate is independent of the detailed visual appearance of the water surface. This is consistent with insights from computer vision that the dimensionality in real images is locally highly reduced due to the continuity of reflecting surfaces in the world (compared to the hypothetical dimensionality in which every pixel would vary independently; see e.g., Roweis & Saul, 2000). The early stages of visual processing in the brain could be viewed as serving the systematic extraction of such low-dimensional feature dimensions (Swindale, 2000). Human observers direct their visual attention to objects based on such low-dimensional feature representations (Treisman, 1998), and working memory is best described in terms of such invariant feature dimensions (Luck & Vogel, 1997).

An analogous argument can be made for movement generation. Motor plans are captured by low-dimensional spatial task parameters (Erlhagen & Schöner, 2002) that are shared with perception (Hommel, Müsseler, Aschersleben, & Prinz, 2001). Mapping these parameters onto motor commands for joints and muscles is a few-to-many transformation that entails the "degree of freedom problem" (Latash, Scholz, & Schöner, 2007). This transformation is not a simple feed-forward projection, but entails recurrent connectivity that sets up compensatory coupling among muscles and joints (Martin, Reimann, & Schöner, 2019; Martin, Scholz, & Schöner, 2009).

I sketched above how low-dimensional neural dynamics may stabilize activation patterns. It is much trickier to stabilize high-dimensional patterns of neural activation. Every neuron in such a pattern needs to "know" which other neurons to excite, which to inhibit. The connectivity required for stabilization is, therefore, specific to each pattern! The Hopfield network (Hopfield, 1999) solves this problem, and it may play a role in the formation of long-term memories (Tsodyks, 2005). Hopfield networks cannot create new instances on the fly nor do they have the capacity for autonomous processing that I will outline below. In vector-symbolic architectures (VSAs) (Levy & Gayler, 2008; Smolensky, 1990), high-dimensional vectors of activation represent instances. VSAs enable key features of higher cognition such as compositionality and productivity, but only if the high-dimensional activation vectors can be sustained independently of ongoing sensory stimulation. Although the neural engineering framework of Eliasmith (Eliasmith, 2005) implements high-dimensional activation vectors in sets of spiking neurons, it does not solve, to my knowledge, the problem of stabilizing these vectors.

Hypothesis 4   Cognition unfolds as sequences of stable neural states are autonomously generated by neural dynamical systems that go through instabilities.

I have proposed that stable patterns of neural activation localized within low-dimensional feature spaces are the neural representations on which cognition and behavior are based. Stability resists change. So how may the sequences of mental states arise without which cognition remains very limited? Change requires release from stability. The hypothesis is that the organized induction of instabilities (or bifurcations, Perko, 2001) is at the basis of cognitive processing.

Deciding which of a set of possible choices to select is an elementary mental act that lies at the basis of much human cognition (Busemeyer & Townsend, 1993; Grossberg & Gutowski, 1987). Models of mathematical psychology conceive of this act as a race between activations representing the different outcomes (Ratcliff & Rouder, 1998), but when embedded in ongoing behavior, the time courses of these activations must be stabilized (Usher & McClelland, 2001), leading to the picture of neural dynamics (Erlhagen & Schöner, 2002; Kopecz & Schöner, 1995). A localized pattern of activation representing one choice is stabilized against competition from the alternate choices by inhibitory interaction within the relevant neural population.

But even responding to a single sensory input is a cognitive act. Detection decisions must be stabilized if they occur in the face of fluctuating sensory inputs. In neural dynamics, the post-decision state is a separate stable state that continues to exist if sensory input is weakened after a detection decision was made (Fig. 1). The detection decision itself arises from an instability, the *detection instability*, in which the resting state of a neural population becomes unstable (Schöner et al., 2016). The underlying mechanism is the threshold property of neural activity that makes neural dynamics nonlinear: Only neurons whose activation is above the firing threshold impact on the neurons to which they project. In a neural population, neural interaction only engages, therefore, once sufficient levels of activation have been reached.

The detection instability amplifies input signals into a macroscopic neural state. The input may be a uniform boost, a "go" signal that does not specify a specific choice. The choice then arises from any inhomogeneity within the population, so that only regions in a population that are slightly closer to threshold than others may become activated. If these are discretely different regions, neural dynamics operate in a categorial regime that can be modeled by zero-dimensional fields or neural dynamic nodes. This may occur, for instance, at the categorization layer of deep feed-forward neural networks.

At the *reverse detection instability*, a self-stabilized blob of activation becomes unstable. This happens at lower levels of input than the detection instability, so that the detection decision is stabilized by hysteresis. Depending on coupling strengths, the reverse detection instability may not be reached even when localized input is removed entirely. The localized pattern of activation is then sustained in the absence of the inducing input, a model of working memory (Simmering, 2016).

Detection and selection uncover the two dimensions of decision-making. Selection is about which of a range of choices is made, that is, about the content of the decision. Detection is about the fact of a decision, the decision to "go." Through the detection instability, a discrete moment in time emerges from the time-continuous neural processes at which this event occurs and macroscopic change is induced. Detection, reverse detection, and selection decisions together organize the autonomous sequential transitions between different activation states from which cognition and behavior emerge (Sandamirskaya & Schöner, 2010). The mechanism builds on the generic excitation–inhibition pairing of neural populations and is illustrated in Fig. 2. Let's say, a meaningful neural activation state has been selected among a range of competing states. This *intentional* state predicts its *condition of satisfaction* (CoS) (a term inspired by Searle's analysis of
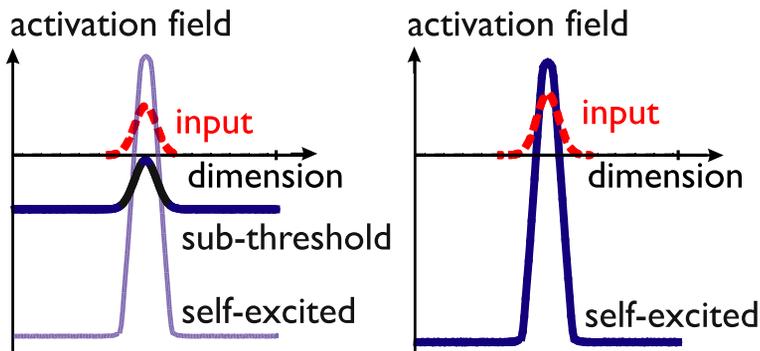
Fig. 1. An activation field defined over a feature dimension is shown below (left panel) and above (right panel) the detection instability. Localized input (dashed line) leaves the system below the detection instability (left panel) when the input-driven sub-threshold activation pattern (fat solid line) remains stable. When localized input increases in strength (right panel), the system goes through the detection instability at which that sub-threshold activation pattern becomes unstable and activation grows into a self-excited activation peak, a macroscopic neural activation state that impacts on downstream processes. This detection decision is self-stabilizing in the sense that the self-excited peak remains stable when input weakens (left panel, thin solid line).

intentionality). It pre-activates an inhibitory population of neurons that represents the CoS. When the CoS system receives input that matches the prediction, an active transient is triggered (Amari, 1977) (somewhat analogously to how spikes are generated). This leads first to the activation of the CoS, then the suppression of the intentional state, and finally, as pre-activation falls away, to suppression of the CoS, clearing the way for any other, competing intentional state to become activated.

This core mechanism has been used to model a variety of cognitive tasks such as visual exploration (Perone & Spencer, 2013; Schöner & Thelen, 2006), change detection (Johnson, Spencer, Luck, & Schöner, 2009), hypothesis testing in the perceptual grounding of concepts (Richter, Lins, & Schöner, 2017), and mental map formation (Kounatidou, Richter, & Schöner, 2018). The hypothesis is that the sequences of mental states that make up thinking may be neurally realized through this kind of mechanism. Intriguingly, the generation of object-oriented reaching movements can be understood in similar terms, with an active transient at the core (reviewed in Schöner, Tekülve, & Zibner, 2019). In this sense, thinking is similar to moving.

Hypothesis 5   Networks of neural dynamics constitute neural cognitive architectures.

The neural dynamics of low-dimensional fields of population activation create stable states that resist change. Such stability implies robustness: The dynamics of a population do not change qualitatively when small changes in the connectivity are made. As a result, fields can be coupled into larger networks of neural dynamics without a qualitative change of their dynamics. Such networks form neural dynamic architectures that enable increasingly complex forms of cognition. For instance, lower-dimensional fields
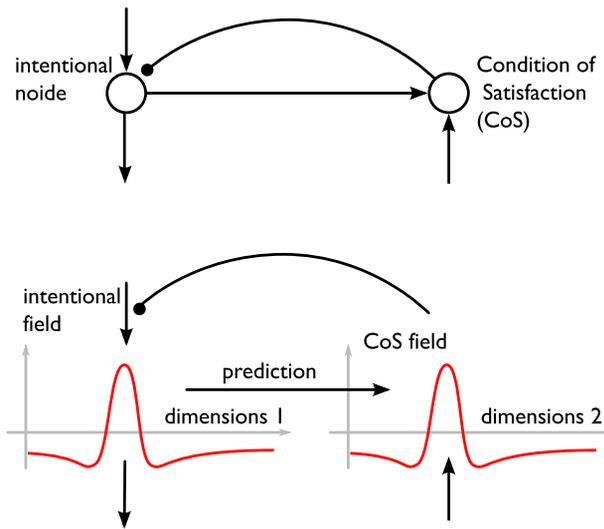
Fig. 2. Autonomous sequential processing entails a pair of neural populations. The excitatory system represents an intentional state, and the inhibitory system represents its condition of satisfaction (CoS). On top, these are illustrated for two neural nodes that represent categorical intentional states. On bottom, the mechanism is illustrated for two neural fields that represent low-dimensional continua of intentional states. Mapping from one set of dimensions to the other reflects predicting the outcome of the intended act. Arrows mark excitatory coupling, and filled circles mark inhibitory coupling. Incoming arrows reflect input from other parts of a neural dynamic network or from sensory surfaces. Outgoing arrows reflect projection onto the motor system or other parts of a neural dynamic network.

projecting into higher-dimensional fields enable visual search or cued recall (Chapter 5 of Schöner et al., 2016), and feature binding through space (Chapter 9 there). Steerable neural maps bring about active reference frame transformations (Chapter 7 there). These are critical to enable neural processes to implement functions and relations (Ballard, Hayhoe, Pook, & Rao, 1997). For instance, a small neural network may be able to detect when a visual target object is "to the right of" a reference object. By transforming the visual array into a reference frame anchored in the reference object, this neural operator can be applied to any reference object (Lipinski, Schneegans, Sandamirskaya, Spencer, & Schöner, 2012; Richter et al., 2017). This is the frontier of the research program I outlined, the goal of which is to show that all of higher cognition can be reached by autonomous neural processing within networks of neural dynamic fields.

4. Conclusion

I have outlined a neurally grounded theory of cognition and behavior at a particular level of description, networks of neural populations. The structure of thinking and behavior provides constraints for this theory, in particular, the need for neural activation to be autonomously generated from within neural networks rather than be induced from the

outside by input. This leads to neural dynamics rather than feed-forward neural networks. It leads to the postulate that neural representations be low-dimensional to enable the regular pattern of neural interaction that support self-generated neural activation. It leads to the identification of dynamic instabilities as the mechanism for decision making and for autonomous sequential transitions between different activation states. And that opens this level of description to processes of higher cognition. Cognition and behavior *emerge* from such networks simply in the sense that the time courses of macroscopic activation patterns that are generated in these networks constitute the time courses of mental states, which are coupled to sensory information and ultimately drive muscles to bring about bodily movement.

The hypothesis is that the dynamics of neural populations form a privileged level of description, at which the laws and constraints of behavior and cognition manifest themselves. This level of description cannot be eliminated by "reduction" to a more microscopic level. To the contrary, according to this hypothesis, the microscopic level is driven (or "enslaved") by this slower, macroscopic level. In this respect, the argument that a bio-physically detailed level of description is somehow primary misses the point (see Ross, 2014, for a careful discussion).

In addition to implementing the macroscopic neural dynamics, lower levels of brain "hardware" must solve many other problems as well, including homeostasis, plasticity, growth, and the optimization of metabolic cost. Ideally, those low-level constraints have no bearing on the macroscopic level of behavior and cognition. Alas, reality might be more complex. It is possible that the separation of scales is broken by complex nonlinear processes that reach across spatial and temporal scales (Bullmore & Sporns, 2009). Such a scale-free theory would be another way to avoid reductive elimination of the macroscopic scale. A related and important issue that I have not covered here is how the macroscopic neural dynamics are re-shaped by experience through neural mechanisms of learning (McClelland et al., 2010).

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*, 77–87.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–767.

Batterman, R. W. (2002). Asymptotics and the role of minimal models. *British Journal for the Philosophy of Science*, *53*(1), 21–28.

Braitenberg, V., & Schüz, A. (1991). *Anatomy of the cortex*. Berlin: Springer Verlag.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

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

Cohen, M. R., & Newsome, W. T. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *Journal of Neuroscience*, *29*(20), 6635–6648.

Coombes, S., P. beim Graben, R. Potthast, & J. Wright (Eds.) (2014). *Neural fields: Theory and applications*. New York, NY: Springer Verlag.

Deadwyler, S. A., & Hampson, R. E. (1995). Ensemble activity and behavior: What's the code? *Science*, *270*, 1316–1318.

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

Erlhagen, W., Bastian, A., Jancke, D., Riehle, A., & Schöner, G. (1999). The distribution of neuronal population activation (DPA) as a tool to study interaction and integration in cortical representations. *Journal of Neuroscience Methods*, *94*(1), 53–66.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, *109*(3), 545–572.

Feynman, R. P. (1967). *The character of physical law*. Cambridge, MA: MIT Press.

Feynman, R. P. (1972). *Statistical mechanics–A set of lectures*. Reading, MA: Benjamin and Cummings.

Feynman, R. P. (1988). *What do you care what other people think?* New York, NY: WW Norton.

Georgopoulos, A. P., Kettner, R. E., & Schwartz, A. B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space: II. Coding of direction of movement by a neural population. *Journal of Neuroscience*, *8*, 2928–2937.

Gerstner, W., Sprekeler, H., & Deco, G. (2012). Theory and simulation in neuroscience. *Science*, *338*(6103), 60–65.

Gibson, E. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Groh, J. M., Born, R. T., & Newsome, W. T. (1997). How is a sensory map read Out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *Journal of Neuroscience*, *17*(11), 4312–4330.

Grossberg, S. (1977). Pattern formation by the global limits of a nonlinear competitive interaction in n dimensions. *Journal of Biology*, *V4*(3), 237–256.

Grossberg, S., & Gutowski, W. (1987). Neural dynamics of decision making under risk: Affective balance theory. *Psychological Review*, *94*(3), 300–318.

Haken, H. (1970). *Laser theory* (Vol. XXV/2c). Berlin: Springer Verlag.

Haken, H. (1983). *Synergetics–An introduction* (3rd ed.). Berlin: Springer Verlag.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavorial and Brain Sciences*, *24*, 849–878.

Hopfield, J. (1999). Brain, neural networks, and computation. *Reviews of modern physics*, *71*(2), 431–437.

Jancke, D., Erlhagen, W., Dinse, H. R., Akhavan, A. C., Giese, M., Steinhage, A., & Schöner, G. (1999). Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex. *Journal of Neuroscience*, *19*, 9016–9028.

Johnson, J., Spencer, J., Luck, S., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection: Research article. *Psychological Science*, *20*(5).

Keller, E. F. (2002). *Making sense of life—Explaining biological development with models, metaphors, and machines*. Cambridge, MA: Havard University Press.

Kopecz, K., & Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biological Cybernetics*, *73*, 49–60.

Kounatidou, P., Richter, M., & Schöner, G. (2018). A neural dynamic architecture that autonomously builds mental models. In C. Kalish, M. Rau, J. Zhu, & T. T. Rogers (Eds.), *Proceedings of the 40th annual conference of the Cognitive Science Society* (pp. 1–6). Madison, WI: Cognitive Science Society.

Landau, L. D., & Lifshitz, E. M. (1979). *Statistical physics. Vol. V of the Course of theoretical physics* (5th ed.). Oxford, UK: Pergamon Press.

Latash, M. L., Scholz, J. P., & Schöner, G. (2007). Toward a New Theory of Motor Synergies. *Motor Control*, *11*, 276–308.

Laughlin, R. B., & Pines, D. (2000). The theory of everything. *Proceedings of the National Academy of Sciences*, *97*(1), 28–31.

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, *332*, 357–360.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. *Nature*, *293*, 291–294.

Levy, S. D., & Gayler, R. (2008). Vector symbolic architectures: A new building material for artificial general intelligence. In P. Wang, B. Goertzel, & S. Franklin (Eds.), *Proceedings of the 2008 conference on artificial general intelligence 2008: Proceedings of the first AGI conference* (pp. 414–418). Amsterdam, The Netherlands: IOS Press.

Lipinski, J., Schneegans, S., Sandamirskaya, Y., Spencer, J. P., & Schöner, G. (2012). A neuro-behavioral model of flexible spatial language behaviors. *Journal of Experimental Psycholog: Learning, Memory and Cognition*, *38*(6), 1490–1511.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Martin, V., Reimann, H., & Schöner, G. (2019). A process account of the uncontrolled manifold structure of joint space variance in pointing movements. *Biological Cybernetics*, *113*(3), 293–307.

Martin, V., Scholz, J. P., & Schöner, G. (2009). Redundancy, self-motion and motor control. *Neural Computation*, *21*(5), 1371–1414.

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

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., & Wang, X. J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, *17*(12), 1661–1663.

Perko, L. (2001). *Differential equations and dynamical systems* (3rd ed.). Berlin: Springer Verlag.

Perone, S., & Spencer, J. P. (2013). Autonomy in action: Linking the act of looking to memory formation in infancy via dynamic neural fields. *Cognitive Science*, *37*(1), 1–60.

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*(10), 416–421.

Raichle, M. E., & Gusnard, D. A. (2002). Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(16), 10237–10239.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.

Richter, M., Lins, J., & Schöner, G. (2017). A neural dynamic model generates descriptions of object-oriented actions. *Topics in Cognitive Science*, *9*, 35–47.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia University Press.

Ross, L. N. (2014). Dynamical models and explanation in neuroscience. *Philosophy of Science*, *82*(1), 32–54.

Rouse, A. G., & Schieber, M. H. (2018). Condition-dependent neural dimensions progressively shift during reach to grasp. *Cell Reports*, *25*(11), 3158–3168.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *289*, 2323–2326.

Runyan, C. A., Piasini, E., Panzeri, S., & Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature*, *548*(7665), 92–96.

Sandamirskaya, Y., & Schöner, G. (2010). An embodied account of serial order: How instabilities drive sequence generation. *Neural Networks*, *23*(10), 1164–1179.

Schöner, G., & Haken, H. (1987). A systematic elimination procedure for Ito stochastic differential equations and the adiabatic approximation. *Zeitschrift für Physik B Condensed Matter*, *68*(1), 89–103.

Schöner, G., Spencer, J. P., &DFT Research Group. (2016). *Dynamic thinking: A primer on Dynamic Field Theory*. New York, NY: Oxford University Press.

Schöner, G., Tekülve, J., & Zibner, S. (2019). Reaching for objects: A neural process account in a developmental perspective. In D. Corbetta & M. Santello (Eds.), *Reach-to-grasp behavior: Brain, behavior and modelling across the life span* (pp. 281–318). New York, NY: Taylor & Francis.

Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*(2), 273–299.

Schwartz, A. B. (2004). Cortical neural prosthetics. *Annual Reviews in Neuroscience*, *27*, 487–507.

Simmering, V. (2016). Working memory capacity in context: Modeling dynamic processes of behavior, memory and development. *Monographs of the Society for Research in Child Development*, *81*(3), 1–158.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*(1–2), 159–216.

Stanley, H. E. (1999). Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Reviews of Modern Physics*, *71*(2), S358–S366.

Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: Noise or part of the signal. *Nature Neuroscience*, *6*, 389–397.

Swindale, N. V. (2000). How many maps are there in visual cortex? *Cerebral Cortex*, *10*(7), 633–43.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society (London) B Biological Sciences*, *353*, 1295–1306.

Tsodyks, M. (2005). Attractor neural networks and spatial maps in hippocampus attractor. *Neuron*, *168*(2), 168–169.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550–592.

Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *Journal of Neuroscience*, *19*(21), 9587–9603.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*, 55–80.

Wilson, K. G. (1979). Problems in physics with many scales of length. *Scientific American*, *241*(2), 158–179.