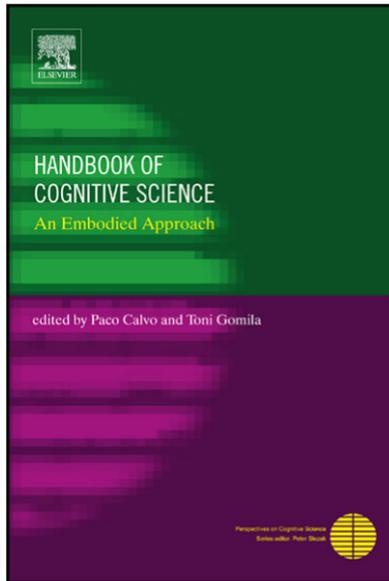


**Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.**

This chapter was originally published in the book *Handbook of Cognitive Science: An Embodied Approach* published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:
<http://www.elsevier.com/locate/permissions>

From: *Handbook of Cognitive Science: An Embodied Approach*, Edited by Paco Calvo and Antoni Gomila
ISBN: 978-0-08-046616-3
© Copyright 2008 Elsevier Inc.
Academic Press.

13

DYNAMIC FIELD THEORY AS A FRAMEWORK FOR UNDERSTANDING EMBODIED COGNITION

SEBASTIAN SCHNEEGANS AND
GREGOR SCHÖNER

Institut für Neuroinformatik, Ruhr-Universität Bochum, Bochum, Germany

Textbooks of cognitive psychology will talk a lot about such things as memory, thinking, deciding, or language, typically with some opening chapters on perception. Movement is often quite secondary in such accounts, and is considered to be a somewhat “low-level” activation of organisms. Yet, all behavior of an organism is ultimately motor behavior. Through motor actions do organisms reveal that they remember something and they have planned something. Visual perception is most commonly supported by motor action that controls where our eyes are pointing or actively supports visual exploration when we take an object into our hands. Conversely, even simple motor acts seem to require the sorts of things that are the stuff of cognition, such as when we must select one of many objects which we want to grasp, or when we must turn our body to bring into our visual array a desired object which we remember is to the right of where we currently look.

Embodied cognition is an approach to cognition that has roots in motor behavior. This approach emphasizes that cognition typically involves acting with a physical body on an environment in which that body is immersed. The approach of embodied cognition postulates that understanding cognitive processes entails understanding their close link to the motor surfaces that may generate action and to the sensory surfaces that provide sensory signals about the

environment. To a certain extent, the embodiment stance implies a mistrust of the abstraction inherent in much information processing thinking, in which the interface between cognitive processes and their sensorimotor support is drawn at a level that is quite removed from both the sensory and the motor systems.

The roots in motor behavior of the embodiment stance manifest themselves also in the emphasis on the real-time autonomy of cognitive processes. These are not typically controlled or triggered by specific inputs to which an “answer” must be generated. Instead, cognition always happens on a background of ongoing behavior. The state of an organism’s nervous systems comes from somewhere and goes somewhere. There is hardly any cognition that does not in some way depend on the recent behavioral and stimulation history as well as the concurrent environmental context. In relation to the environment, this context sensitivity of cognition is sometimes referred to as “situatedness,” a concept we subsume here under embodiment.

Finally, for some (and for us), the embodiment stance also postulates that an understanding of cognition must be based on concepts that are consistent with the fundamental principles of neuronal organization that govern our nervous systems. This means, in particular, that cognition happens in a temporally continuous and asynchronous fashion, without a central controller that clocks computational steps. This also means that a homogeneous language is spoken within the neuronal networks which our nervous system consists of. Neurons interact through their activation levels, be they assessed by firing rates, levels of synchronicity, or intra-cellular potentials. What neurons transmit through their axons and the synapses they form is always the same type of variable. Neurons do not transmit messages beyond these physical signals. The processing of neurons is largely homogeneous across the higher nervous system, and is based essentially on weighted integration. Only through the structure of the neuronal networks, of which neurons are part, may the different functionally relevant states of neurons be brought about. Note, however, that there are no signatures of the temporal discreteness of neuronal spiking events or of the spatial discreteness of individual neurons in cognition or behavior. So the level at which the neuronal substrate provides constraints for an understanding of cognition must be identified rather than fixed a priori. In our review, that level will consist of spatio-temporally continuous neuronal activation patterns. The radical stance within the approach of embodied cognition is that the link to the sensory and motor surfaces, the constraints imposed by the physical body and the structured environment in which it is immersed, the constraints of temporal continuity and autonomy, and the constraints provided by the neuronal substrate are relevant not only for the subset of cognitive processes that control action and perception. Instead, in the radical view, all cognition is hypothesized to be of this kind. Remoteness of cognitive processes from the sensorimotor domain, independence of physical instantiation, forward computation only from given inputs, and abstraction from the neuronal substrate are all illusory. Even the highest form of cognition, thinking, is viewed as a form of motion, characterized by similar constraints as motor behavior, if

not always directly acted out by the motor system (Port & Gelder, 1995). These claims cannot be considered proven at this time but provide a very stimulating research program for a fresh understanding of cognition.

It is clear, that new theoretical tools are needed to address cognition within the embodiment perspective. This chapter reviews one set of theoretical concepts which we believe to be particularly suited to address the constraints of embodiment and situatedness. We refer to this set of concepts as *Dynamical Systems Thinking* or *DST*. The concepts are based on the mathematical theory of dynamical systems, but are not identical with that theory, of course (which is why we resist the term “Dynamical Systems Theory” that is sometimes used to describe this approach). In shortest form, DST is the proposition that the states of the nervous system from which cognition emerges can be described by ensembles of continuous state variables that evolve continuously in time. That evolution is characterized by dynamical laws. Functional states of the neuronal dynamics are attractors, whose stability enables them to persist in the face of perturbations and fluctuating inputs. New solutions and qualitative functional change emerge from instabilities of the neuronal dynamics.

Stability, a core concept of DST, has obvious roots in motor behavior. As every engineer knows, stability is of the essence whenever the control of a physical effector is continuously linked to sensory information as it is during the execution, but also the planning of motor behavior (Goodale et al., 1986). This need to stabilize functional states generalizes, however, to nervous activation other than overt motor behavior, because continuous links to sensory information as well as other, ongoing neural processes is a pervasive feature of neural function. Given the high degree of functional connectivity within the central nervous system, any neuronal subpopulation engaged in a particular functional state receives signals from many other neuronal subsystems that are not contributing to this function. In effect, these signals represent perturbations of the ongoing functional state, against which the state must be stabilized. This is true even for perceptual processes, for which feedforward computation would at first sight seem a reasonable framework. Stability is required, however, to form coherent percepts from the continuous stream of inherently ambiguous sensory signals (Hock et al., 2003).

Once we recognize that functional states of neural systems have stability properties, the question arises how systems may change state to approach the flexibility that characterizes cognition. In the motor domain, such flexibility may appear limited, but cognition and perception are inherently time varying and highly responsive to changing inputs. Flexibility requires that functional states be released from stability. This happens in instabilities (or bifurcations), at which the neuronal dynamics go through qualitative change, leading to new functional states (Schöner & Kelso, 1988; Schöner, 2008).

In the motor domain, the notion of a dynamic state of the neural control systems is easily grounded in biomechanics and physiology. In fact, muscle–tendon systems contribute through their elasticity and viscosity to the stability of effector systems as do peripheral and central reflex loops (Feldman, 1986; Bizzi & Mussa-Ivaldi, 1990;

Hogan, 1990). Applying DST beyond the motor domain requires that those neuronal principles be identified that may endow representations with stability. Stability, as we will illustrate later, requires a metric which distinguishes small from large perturbations and within which resistance to and recovery from perturbations can be defined. Representations can be embedded in metric state spaces through the concept of activation fields that span the continuous, potentially high dimensional spaces of possible percepts, memory states, or action plans (Shepard, 1980). Much of this chapter will review a class of neuronal dynamics, originally inspired by the homogeneous, layered structure of cortical anatomy (Wilson & Cowan, 1973; Amari, 1977), which provides the key to endowing representations with dynamical stability properties as well as the potential for instabilities from which elementary forms of cognition emerge (Spencer & Schöner, 2003; Schöner, 2008). We refer to the conceptual framework that result from combining the concepts of DST with this class of neuronal dynamics of activation fields as *Dynamic Field Theory* or *DFT*.

The major part of this chapter will review DFT, providing first foundations, discussing the units of representation as stable localized patterns of activation, and illustrating some of the instabilities through which different forms of elementary cognition emerge. We will show how this framework connects the graded sensorimotor representations underlying estimation, detection, and motor planning to the seemingly discrete representations underlying categorical behavior. To examine the extent to which DFT is consistent with the embodiment stance, and as a pointer to the achievable complexity of cognitive function, we will review a robotic application of these neuronal ideas to object recognition. Before we start, however, we will ground the ideas in the mathematical theory of dynamical systems through a brief and quite elementary tutorial.

DYNAMICAL SYSTEMS

The theory of dynamical systems has its origins in classical physics, where it was used to understand how physical systems evolve in time. Much of physics involves so-called conservative systems that do not have (asymptotically) stable states, so that perturbations affect the long-term behavior of the systems for ever. A textbook example of a conservative system would be a frictionless pendulum. If hit somewhere along its orbit, its future time course is forever changed. Real pendulums, in contrast, are affected by friction and ultimately come to rest. The resting state is stable, which makes it a point attractor.

Dynamical systems that have stable states are called dissipative by physicists and form a special subclass, relevant to understanding neural and behavioral systems. The investigation of the stability properties of such dissipative systems has received considerable attention in mathematics, and we shall introduce some of the most basic terms and interrelationships here. (There are very many textbooks on this field of mathematics. Two examples are Braun (1993) at an elementary level and Perko (1991) for a more advanced level.) To visualize a dissipative dynamical system, imagine a ball rolling in a smoothly sloped landscape

under the influence of gravity. The ball may be coated with something sticky and so it experiences a lot of friction. This will ensure that its movements will be dampened quickly, generating stable states. The state of the system is then completely described by the ball's position in the horizontal plane. The change of this state—the movement of the ball—is determined only by the slope of the landscape locally at the position of the ball. If the slope is zero at the ball's position, the system is at a fixed point, and its state does not change anymore. There are a number of different kinds of fixed points: A stable fixed point or *attractor* is reached if the ball is at the lowest point of a valley. If the ball's position is disturbed within certain bounds by an external force, it will return to this point. Stability is this property of converging to a fixed point from any point in the immediate vicinity of the fixed point. (In mathematical language this is actually called *asymptotic stability* and differs from a weaker condition that mathematicians define as stability. Like most physicists and engineers, we continue to use the term “stability” for the stronger condition of asymptotic stability.)

A stronger disturbance may cause the system to leave the fixed point's *basin of attraction* (a term that can be taken literally in our example), and the ball will come to rest in some other valley, thus putting the system into a new attractor state. Another kind of fixed point, a *repellor*, would exist at the very top of a hill. Exactly on the top the slope is zero, but any disturbance, even a very small one, will cause the system to move away from this fixed point as the ball rolls downhill toward some other valley that might be available. A repellor is not a stable state. There are other ways in which a fixed point could not be stable; for instance by lying exactly at a saddle of the landscape: the system would be attracted only along one route down the saddle but would run away from the fixed point along all other directions. Still another way in which stability could fail would be observed if there was a direction in which the landscape was exactly flat. Along a valley with a perfectly horizontal floor, any point would be a fixed point that would not be perfectly stable because a perturbation would shift the system along the valley. It would still be more stable than a repellor or saddle point, however, because it would stay close to the original position.

Formally, a dynamical system can be described by one or more differential equations of the form $dx/dt = f(x)$, in which the rate of change, dx/dt , depends on the current state, x . For a given initial condition, this equation makes it possible to determine the system's state for all later points in time by integrating the rate of change along time. How attractors and repellors emerge from such equations can be visualized for a one-dimensional state space by plotting the function $f(x)$ as a function of x (Figure 13.1). By definition, fixed points are zero crossings of this function. Fixed point attractors are zero crossings at which the dynamic function, f , has a negative slope: The rate of change is positive for states smaller than the fixed point, leading to increase and thus movement toward the fixed point. The rate of change is negative for states larger than the fixed point, leading to decrease and thus likewise movement toward the fixed point. A zero crossing with a positive slope of the dynamic function is a repellor, at which the analogous logic explains why the system is pushed away from the fixed point. It may be intuitive from these

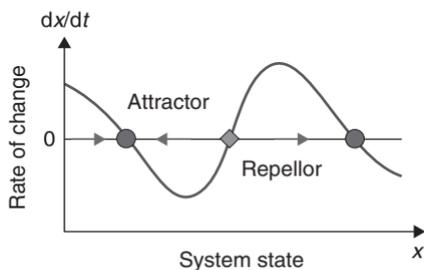


FIGURE 13.1 A one-dimensional dynamical system is described by how the rate of change dx/dt depends on the state, x , of the system. Zero crossings are fixed points that can be either attractors or repellers.

considerations that if f is a continuous function, two attractor states are always separated by a repellor (and vice versa). Thus, to switch from one attractor state to another, some external force must be exerted on the system (represented by a temporary deformation of the dynamic function, f) that is strong enough to move the system to the other side of the repellor. The repellor, therefore, demarks the boundary of the basins of attraction of the two attractors.

A change of the system's state may also occur if some external parameter—one that is not included in the state x —alters the system's dynamics. In the rolling ball metaphor, imagine that the landscape is tilted from the horizontal by a certain angle. This will cause the attractors (minima) and repellers (maxima) to shift (actually, the maxima will move in the opposite direction to the minima). When the incline reaches a critical point, some minima may stop being local minima; typically because they collide with a local maximum (try this out for a one-dimensional landscape!). The ball will track the lowest point of the valley it is in, until that minimum disappears. At this point, the attractor undergoes instability. With just a little more increase in the incline, the ball will move way from the former valley, until it reaches some other valley, which still contains a stable state (this example allows for the unfortunate outcome in which the ball runs off to infinity when no other valleys are left in the downhill direction).

This is a rather abstract view of dynamical systems. How could these terms be used to talk about the evolution in time of patterns of neural activation? What may stable states look like in such neural dynamical systems? How may they arise or disappear through instabilities? We will discuss next how Dynamical Systems Thinking can be combined with neural principles in DFT.

DYNAMIC NEURAL FIELDS AND PEAKS AS UNITS OF REPRESENTATION

The architecture of the *Dynamic Neural Field* or *DNF* is based on the finding that in the central nervous systems of vertebrates metric information is

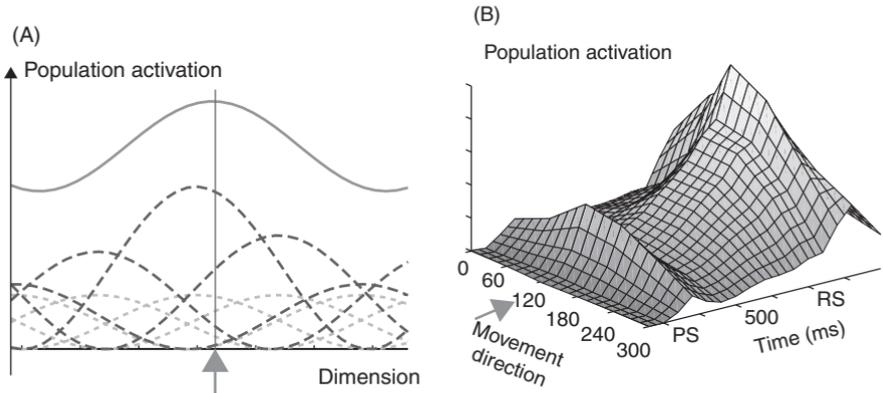


FIGURE 13.2 (A) Neurons tuned to a metric dimension have tuning curves (short dashed lines) with a single hump. These are weighted with the current firing rate of each neuron (long dashed lines) and superposed, generating the distribution of population activation over the metric dimension (solid line). When a specific value of the dimension is specified as in the illustrated case (arrow), the neurons with preferred values close to the specified value contribute more strongly than neurons with preferred values far from that value, because their firing is higher. (B) Time course of a distribution of population activation over the dimension of movement direction constructed in this way from the tuning curves of about 100 neurons in motor cortex (Bastian et al., 2003). The movement direction “120” is first signaled at the time marked as “PS,” followed by the “go” signal at time “RS.” A single peak located at that movement direction emerges.

commonly represented in the form of population codes (Erickson, 1974; Georgopoulos, 1991; deCharms & Zador, 2000). This means that dedicated populations of neurons exist whose activities, taken together, yield a representation of a certain feature. This may be the color of a visual stimulus, the pitch of a sound or a desired hand position in the planning of a motor action. Each neuron within such a population is maximally active when a certain, “preferred” feature value is presented, and its activation decreases as the feature value contained in the stimulus differs increasingly from this preferred value. The response property of a neuron can be visualized by its tuning curve, which plots the neuron’s average activation against the feature dimension. The tuning curves of all neurons in a population cover the represented metric dimension or the relevant part thereof. There is usually a strong overlap between the tuning curves of neurons, so that each stimulus will cause activation in a number of neurons.

For the following theoretical considerations we will assume the neurons are ordered according to their preferred feature value even though this ordering does not necessarily correspond to the spatial layout of the neurons in the nervous system. In this perspective, the information represented by the population can be read out from the spatial distribution of activation (Figure 13.2): A single value along the feature dimension can be represented by a localized peak of activation, that is, by a group of neighboring neurons with high activation levels in an otherwise inactive population. The width and height of a peak may give additional information

about the precision or certainty of this value. Ambiguous information about the feature can be represented by multi-modal distributions of activation, and the absence of information by uniformly low activation levels over the whole population.

These activation patterns are shaped not only by the input that the neurons receive from other structures but are greatly influenced by interactions within the population. A ubiquitous form of connectivity in the central nervous system can be characterized as local excitation and global inhibition in the spatial arrangement of feature sensitive neurons. Neurons that code for similar feature values excite each other, whereas neurons that code for distant feature values inhibit each other (via inhibitory interneurons). This kind of interaction promotes the emergence of localized activation peaks, as will be discussed later.

One key assumption of DNF models is that it is the distributions of activation over neural populations that convey the relevant information and not the behavior of the single neurons. Accordingly, DNFs abstract from the neurons as discrete computational units and model activation over continuous feature dimensions. The evolution of activation patterns is modeled as a continuous process in time, described by a set of differential equations. Special emphasis is put on the internal interactions in the field which are critical for establishing stable states. To model these interactions, an output is calculated over the whole field and fed back into the field as endogenous input. This output can be regarded as a correlate to the mean firing rate of a group of neurons, whereas the activation reflects their mean membrane potential. The field output is usually calculated from the activation via a sigmoid function, which is close to zero for low activation levels, rises around a threshold value and saturates at a constant value for higher levels of activation. The distribution of the endogenous input that originates from one position in the field can be described by an interaction kernel: It consists of an excitatory part, typically modeled as a Gaussian centered at the origin of the output, and an inhibitory part. The inhibitory component may be homogeneous over the whole field, but it may also be a broader Gaussian, resulting in a Mexican hat shape. This type of kernel implements the pattern of local excitation and global inhibition found in neural populations. External input can boost the activation in the field, either locally or globally. Finally, in many models random noise is added to the activation field to account for fluctuations in neural activation that cannot be captured by a deterministic differential equation.

In the absence of any input, the activation over the whole field is driven toward a preset resting level, which is usually chosen to be well below the threshold of the output function. This pattern of activation constitutes a first attractor state of the DNF: If the activation is perturbed by noise, it may fluctuate around the resting level, but it does not drift over extended periods of time, and it relaxes toward the resting level when the noise is turned off. If a weak localized input is added to the field, then the activation in the field rises toward a state reflecting the sum of resting level and input (Figure 13.3A). Here, the input strength acts as an external parameter that causes a shift of the attractor states (similar to the tilt of the landscape in the rolling ball example), and the activation distribution follows the attractor. We call this the input-driven state of the DNF.

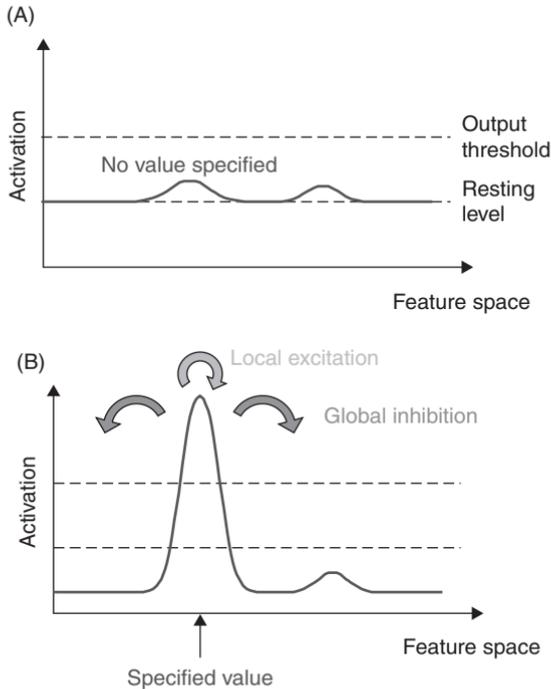


FIGURE 13.3 In DFT, metric information is represented by continuous distributions of activation over metric dimensions that span perceptual or motor feature spaces. (A) Low activation levels across the entire field index the absence of conclusive information for the feature space. (B) A single value along the feature dimension is specified by a peak of activation localized at a particular position in the field, which stands for that feature value. Such activation peaks are the units of representation in DFT and emerge as attractors from the neuronal dynamics of the activation fields.

Consider next the case of a localized input that is strong enough to lift the activation in a small section of the field above the output threshold. In this case, the interactions in the dynamic field must be taken into consideration to determine the stable state of the system. The local excitation will drive activation even higher at the location of the input and global inhibition will depress it elsewhere (Figure 13.3B). If the parameters of the interaction kernel are within a certain range, the result will be a strong localized peak of activation surrounded by a zone of inhibited activation. For this state, the endogenous excitation and inhibition as well as the exogenous input and the forces pushing the system toward resting level reach a balance at every position of the field. This constitutes another attractor state, which we refer to as a self-stabilized state and which is qualitatively different from the input-driven state.

One way to see that this is a qualitatively different state is to decrease the strength of the external input again enough, so that the combined effect of input and resting level are insufficient to reach the output threshold. The field activation in the area of the peak will, however, remain high enough to sustain output

leading to self-excitation by the local excitatory interactions. At such an input level the system is bistable, that is, two attractor states co-exist: The input-driven state that is reached from low levels of activation, does not engage interaction, and mirrors the input signal. The self-excited state is reached from sufficiently high levels of activation, stabilized by interaction, but continues to be influenced by input. Which of these attractor states the system reaches at this input level is determined by the field's activation history. Weak previous levels of activation put the field into the basin of attraction of the input-driven state, strong previous levels of activation put the system into the basin of attraction of the self-excited state. For initial states near the boundary of the basins of attraction, the system may reach either state depending on stochastic perturbations (reflecting, for instance, noisy neural inputs).

The qualitative change of the attractor states when a single localized input increases in strength is illustrated in Figure 13.4. The dynamics generating the attractor solutions sketched in the left column of the figure can be illustrated by plotting the rate of change, $du(x)/dt$ of the activation level at some location, x , within the peak, as function of the activation level, $u(x)$, at that same location. Strictly speaking, this plot is not a mathematically conclusive representation of the dynamics, because the rate of change also depends on activation levels at other field sites. The intuition derived from this plot is corroborated by the correct mathematical analysis, however (Amari, 1977). The rate of change has a negative slope overall, reflecting the fundamental stability of neuronal activation. At large levels of activation, the rate of change is lifted up by the net effect of the excitatory interactions within a peak of activation. The effect of localized input is to shift the rate of change upward across all activation levels. As a result, the single attractor at low levels of activation is joined by a second attractor at high levels of activation, into which the system switches when the attractor at low activation levels becomes unstable for sufficiently strong input. When input levels are then lowered again, the system will remain in this activated state until that state becomes unstable for sufficiently weak inputs. Either switch occurs as an attractor disappears after becoming unstable.

A behaviorally relevant effect of these instabilities is that the bistable regime helps stabilize detection decisions. Consider a simple perceptual detection task and assume that a stimulus is perceived when the relevant neural population creates sufficient activation that exceeds the threshold for output to be generated. In the input-driven regime, the percept would be very unstable for a stimulus that is just strong enough to push the field to the output threshold. Due to sensory noise, the activation would fluctuate around the threshold and the output nonlinearity would produce a signal that alternates on a fast timescale. In the self-stabilized regime, the percept is stabilized once activation reaches the output threshold. The percept persists even when the input strength is reduced (within limits). Empirical support for the stabilization of detection decisions comes from psychophysical experiments demonstrating perceptual hysteresis (Hock et al., 1997).

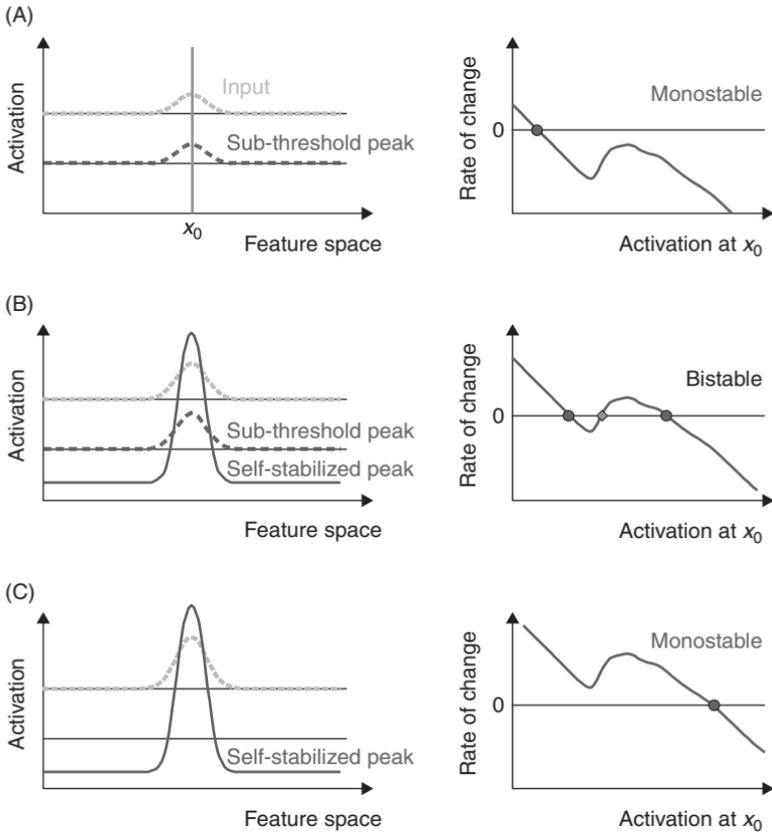


FIGURE 13.4 Stable patterns of activation induced by a single Gaussian input of varying strength are shown (left column) together with the corresponding plots of the rate of change against the current activation level, both taken at the peak position (right column). (A) For weak input (short dashes), the only stable pattern is a matching subthreshold peak (long dashes). The associated dynamics is monostable (dot marks the attractor). (B) At intermediate input strength, the system is bistable. One attractor emerges from the subthreshold peak (long dashes), which is merely shifted toward higher levels of activation (leftmost attractor in the plot on the right). The other attractor is a self-stabilized peak (solid line). It shows up as an additional attractor state of the system (rightmost attractor in the plot on the right), separated from the old attractor by a repeller (diamond). (C) At the highest input levels, the system is monostable again with only the self-stabilized peak surviving. The subthreshold peak has become unstable.

The range of input strengths for which the DNF is in the bistable regime depends on the parameters of the interaction kernel. So far, we have assumed that a DNF would be monostable in the absence of any input so that the activation would always relax to the resting level when the input is removed. In the DNF model, this is not necessarily the case, however. If the excitatory part of the interaction kernel is strong enough and is balanced by sufficient inhibition to stabilize a local peak, then a perfectly stable peak of activation may persist even

without exogenous input. Such self-sustained peaks of activation have been used to model metric working memory (Zipsler, 1991; Durstewitz et al., 2000; Spencer & Schöner, 2003): A peak is created by a single presentation of a localized stimulus. It is then sustained over extended periods of time, representing the former input as a memory item together with its metric value reflected by the location of the sustained peak.

We will make a few additional remarks about the relationship of DNFs to real neural populations and brain structures. Generally, DNFs can be used to describe neural systems at different levels of abstraction. Historically, DNFs were first developed to approximate the cortical neuronal architecture that is characterized by layered sheets of neurons which are relatively homogeneous along the layers with strongly overlapping dendritic trees for nearby neurons (Wilson & Cowan, 1973; Amari, 1977). DNFs can be used to model clearly identified populations of neurons using the concept of a distribution of population activation (Erlhagen et al., 1999), which frees the description of strict anatomical constraints. This has been done, for instance, for the representation of retinal location in primary visual cortex (Jancke et al., 1999), the representation of movement direction in motor cortex (Bastian et al., 1998; Cisek, 2006), and for the representation of saccadic end-points in superior colliculus (SC) (Trappenberg et al., 2001). In these cases, the parameters of DNF models can be tuned to reproduce the experimentally observed patterns of neural activation.

However, DNFs may also be used to explain the results of psychophysical experiments, such as the metrics of performance, reaction times, error rates, frequencies of responses, and other signatures of the underlying processes (Kopeck & Schöner, 1995; Erlhagen & Schöner, 2002; Schutte et al., 2003). In these cases, the feature dimensions over which the fields are defined are usually parameters of the experimental setup. It is often not known exactly where the processes modeled in the DNF take place in the brain, and it may even be doubtful whether any single neural population exists that behaves exactly as the dynamic field does. Instead, the neuronal instantiation of such DNFs could be distributed across multiple areas and populations of neurons. Such functional DNF models may properly capture the net effect of the evolution of stable activation patterns that underlie the observed behavior (Spencer et al., 2007).

INTERACTIONS BETWEEN MULTIPLE ACTIVATION PEAKS

Up to here we looked at the input-driven attractor and at a single, localized peak that is stabilized by interaction and forms a second attractor (see Amari (1977) for a complete mathematical analysis). More complex attractor configurations arise if two or more localized inputs are applied to the field. In such a case, multiple peaks of activation may emerge that influence each other due to the excitatory and inhibitory interactions. In this section we review the different

effects that these interactions generate and show how the resultant solutions can be used to explain experimental results linked to sensorimotor decision-making. The metrics and timing of saccadic eye movements have been extensively investigated and much is known about the underlying neuronal substrate. The neuronal specification of such movements thus provides an excellent model system, which we can use to illustrate key ideas of DFT. We build on detailed DNF models of saccade specification (Kopecz & Schöner, 1995; Trappenberg et al., 2001; Wilimzig et al., 2006).

Saccades are the abrupt eye movements that we use to change the fixation point of our eyes from one location in visual space to another. Saccades are ballistic movements, that is, each saccade's trajectory is determined before the movement starts and normally it remains unaltered during the execution of the saccade. Furthermore, saccadic eye movements are highly stereotyped so that it is sufficient to specify the horizontal and vertical distance of the saccadic target in retinal coordinates, that is, relative to the current fixation point. These two spatial dimensions of the saccadic end-point can thus be considered relevant feature dimensions (for simplicity we will think of only a single dimension in what follows). A peak in an activation field defined over these dimensions thus indicates the metrics of a planned saccade. Such a DNF may be interpreted as a functional description of relevant neural populations, in particular, those in the Superior Colliculus (SC), a mid-brain structure that is involved in saccade planning and initiation. The SC features a topographic map of saccade target positions, in which activation peaks arise before a saccade is initiated. The SC integrates both sensory and cortical inputs and it is assumed that it is in the SC that the final decision about the initiation of a saccade is made. (A more detailed model points to multiple zones within SC and to different layers playing different roles in the specification and initiation of a saccade as well as the opposing function of fixation.)

The presence of a stimulus somewhere in the visual field is modeled by localized input to the corresponding position in the DNF. If the input is strong enough, the system goes through the detection instability and a peak emerges, indicating the metrics of a saccade to the visual target. Under natural conditions, of course, there is never a single unique visual target in the visual array. Instead, typical visual environments provide a rich selection of potential targets of saccadic eye movements, which are most commonly characterized by some high-energy local contrast, edge or corner point. Specifying a saccade under such conditions necessarily involves selection (Ottes et al., 1984). DNFs and their interactions afford such selection.

Consider first a case in which two identical inputs are presented to two field sites that are at a large distance from each other (Figure 13.5A). When these inputs are sufficiently strong, they induce levels of activation in the field that reach the output threshold and thus engage the neuronal interaction in the field dynamics. For perfect symmetry and in the absence of noise, two identical peaks may arise. Mutual inhibition may reduce the total activation in these peaks as compared to a peak induced by a single localized input. This is because

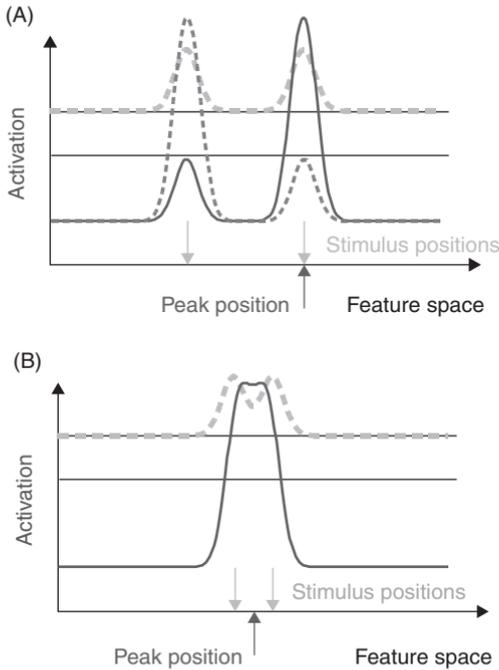


FIGURE 13.5 (A) If two localized inputs (arrows and fat dashed line on top) are applied at distant positions, the emerging peaks compete with each other due to self-excitation and mutual inhibition. This results in the selection of one peak and the suppression of the other. Any of the inputs may be selected depending on their respective strengths and the field's history, allowing two possible stable states for the dynamic field (dashed and solid line below the input). (B) If input positions are close to each other, local excitatory interactions bring about a fusion of the activation peaks, leading to a monostable response at an averaged location.

both peaks contribute to inhibitory interaction that impacts on the entire field, whereas excitatory interaction is local to each peak. The two-peak state is a fixed point of the system but not generally a stable state. If the activation level of one peak is slightly increased by noise or stimulus asymmetry, that peak generates more supra-threshold activation leading to stronger self-excitation and stronger global inhibition. This will diminish the activation in the other peak and, in turn, it reduces the amount of self-excitation within that peak as well as its inhibitory influence on the other peak. An imbalance between the two peaks will arise and grow, which may lead, for sufficiently strong interactions, to the complete suppression of one peak by the other one. Given that all output from the inhibited peak is suppressed, the remaining peak has the same shape and strength as a peak with only one localized input.

Thus, for strong bimodal input, the system is again in a bistable state: If a single peak has been established and activation at the other input location has been suppressed, this pattern is stabilized against noise and also against moderate

increase of the suppressed input. Which one of the two possible peaks is realized depends on the system's activation (and thus stimulation) history as well as on random stochastic fluctuations (e.g., from many uncorrelated neural inputs). If one location receives stronger input than the other then that location has a greater chance at generating a peak, which is suppressed only if a strong, and thus rare, stochastic perturbation favors the other location. The same mechanism of competition takes place for more than two inputs, leading to the selection of one location in many, which is typically the location with strongest input.

So far, we have analyzed the case in which the locations receiving input are distant from each other, so that the associated activation peaks only inhibit rather than excite each other. What about closely spaced inputs? We first remind the reader of experimental observations for metrically close saccadic targets (Ottens et al., 1984). If multiple visual stimuli are presented in proximity to each other, but distant from the current fixation point, the result is typically an averaging saccade made to the center of the group, not to a single item. A smaller saccade to fixate a specific target may follow in a second step.

This averaging behavior can be understood in terms of DNFs as well (Figure 13.5B). Two Gaussian inputs to a dynamic field will overlap if they are close to each other, so that both input sources contribute activation to the area between the two locations. Trivially, this may result in a single localized input to the field, centered already over the averaged input locations. An averaging peak may even emerge, however, when the two inputs do not overlap so strongly that a single-humped input distribution results. Input induced activation at two locations that are close enough to experience mutual excitatory interaction will tend to fuse into a single peak. Supra-threshold activation at either location propagates toward the center. This converges to a merged peak at an averaged position, similar to one that would be created by a single broad input.

For two localized inputs that are applied very close to each other, the merged peak is the only stable activation pattern. If the distance between the inputs is increased continuously, this pattern will remain stable over a certain distance, whereas the same input may create a selection behavior when applied to a previously inactive field. If the distance is increased further, the merged peak attractor is destabilized and the peak quickly shifts to one of the stimulus positions (this is called the fusion/selection instability). Because excitatory interaction can take a direct route between excitatorily coupled neurons, whereas inhibition requires inhibitory interneurons, this account predicts that early saccades tend to fuse inputs, whereas later saccades that occur after more time has been available for the neuronal dynamics to settle, tend to select one target (Wilimzig et al., 2006). This is empirically true.

The time course of sensorimotor decisions has been studied using DFT ideas in a variety of other settings. The timed–movement–initiation–paradigm (Erlhagen & Schöner, 2002) provides access to the preparation of goal-directed hand and arm movements. Infants show reliable patterns of selection when confronted with multiple possible reaching targets in the famous A-not-B task of Jean Piaget

(Thelen et al., 2001). Here the delay across which competition between a motor habit and a cued new movement target occurs can be varied experimentally. More generally, tasks involving metric working memory provide access to the temporal evolution of selection decisions, exposing time-dependent metric biases due to the influence of competing influences. In all of these cases, prior experience within the task plays a critical role in how the selection process unfolds. To understand sensorimotor decision-making, we need to look more carefully, therefore, at how prior experience may have an impact on selection decisions.

PRESHAPE IN DYNAMIC NEURAL FIELDS

We have just seen how small inhomogeneities in a field, a little more input at one location than at another, can have a critical influence on selection decisions. The activation fields underlying the perceptual or motor decisions cannot be generally expected to be perfectly neutral, clean slates. Whenever a particular input arrives that drives the field toward a decision, the activation pattern in the field may be preshaped by other inputs that have been around longer. One source of such preshaping input is the sensed environment, in which there may be rich visual structure including potential movement targets such as graspable objects. Decisions typically take place on such a background of prior activation.

One particular source of such preshaping of activation fields is the recent activation history. Habit formation is perhaps the simplest form of learning in which an organism builds a tendency to repeat behaviors that have been successful before. Habits may be accounted for in DFT by assuming that patterns of activation leave a memory trace, which then in turn contributes to preshaping the field. A simple mathematical formalization is based on an additional layer of activation, in which such a memory trace results from a slow dynamics. This memory layer in turn provides input to the proper activation field. The resultant preshaped activation is generally subthreshold, so that it does not by itself induce decisions. Preshape may, however, exert a great influence on the activation patterns that emerge when stimulus input is added.

The concept of preshape is not meant to model one specific neural mechanism. Instead, preshape is a general functional account for a variety of neural mechanisms that contribute to activation prior to an imperative or specific signal, which triggers an instability leading to a peak being formed and a decision being made. Long-term memory and associations from other cortical areas may represent expectations, predictions, or attention directed at certain parts of the feature space. In other cases, the preshape may be thought of as residual activation from previous behavior. Learning mechanisms may involve changes in synaptic efficacy, either in the afferent or in the lateral connections. In either case, the functional effect is to facilitate the induction of a peak, which we may conceptualize as inhomogeneity or preshape of the field.

How preshape influences selection decisions can be illustrated again in the preparation of saccadic eye movements. In the laboratory, only a limited number of potential target locations are typically used. Participants may acquire prior knowledge about the possible eye movements they will need to perform to acquire these targets. The effects of this knowledge were investigated by Dorris et al. (2007) in rhesus monkeys who were trained to fixate on a light point and then make a saccade to a visual target appearing at a predictable position in the visual array. In electrophysiological measurements in the SC, they found a localized hill of activation in the area representing the target location before the stimulus was actually presented. Furthermore, they investigated the behavioral effect of this preparatory activation by presenting distracters (that differed from the actual targets by their color) at different locations in the monkey's visual field. Distracters presented close to the usual target location tended to attract the eye, leading to erroneous saccades, whereas the metrically distant distracters did not.

These observations can be understood within the DNF model of saccade preparation. In the preshape layer, a constant broad hill of activation is created at the trained target position, moderately increasing the activation level in the associated dynamic field. The distracters may be modeled as transient inputs that are weaker than the target input (because they are not reinforced by the neural systems that performs the target recognition). Such an input will be sufficient to create a peak if applied to a preactivated region of the field, but not in the other regions, explaining the different rates of erroneous saccades. A further effect of the preshape is that peaks are created faster in response to target presentation, predicting shorter reaction times for saccade initiation in situations where the target position is known in advance. This is in accordance with a large range of experimental results (as reviewed in Erlhagen & Schöner, 2002). The same effect of pre-information on neural activation levels and reaction times has been shown for motor cortex when pointing movements were prepared (Bastian et al., 2003).

How preshape may be acquired by a simple learning process is illustrated in Figure 13.6. To this end, the output of the dynamic field is fed into a memory trace layer. Thus, whenever there is supra-threshold activation in the field, a memory trace is laid down at the matching locations. Because the memory trace evolves over a much slower timescale, the pattern within the memory trace layer reflects the statistics of activation in the field, with more activation built in those locations that have repeatedly and consistently been activated.

The memory trace is thus a mechanism through which probability distributions reflecting the activation history can be autonomously acquired and neuronally instantiated, much in the manner of the prior distributions of the Bayesian framework. Metric biases may arise from the preshape pattern induced by such memory traces, which are again consistent with Bayesian estimation. This effect is illustrated in Figure 13.7, where a tiny amount of bias toward one of the pre-shaped locations can be seen (the effect can be larger under appropriate circumstances (see Erlhagen & Schöner, 2002); here we aim to contrast this effect with

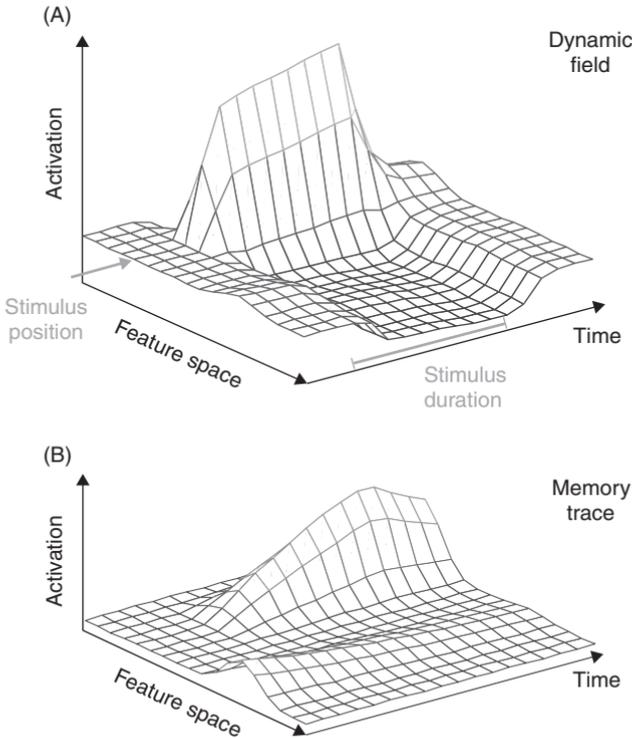


FIGURE 13.6 The evolution in time of the activation pattern in a dynamic field (A) and an associated memory trace field (B) is shown. As long as a peak is present in the activation field (here due to the presence of a stimulus marked by the bar labeled “stimulus duration”), the memory trace slowly builds up at the corresponding field location. Memory traces passively decay in the absence of such activation. In this illustration, a memory trace at a different feature value was assumed to exist initially from earlier peak events. That trace preshapes the activation field at the matching location but then slowly decays because it is not further stimulated by input from the activation field.

a different mode of integration discussed in the next section). One manifestation of the preshaping of the choice seen here is that the time needed to build a peak is shorter when the peak is consistent with the memory trace than when it is not. Note, however, how the DNF goes beyond the fusion of prior and sensory input. The field dynamics suppresses any influence from the other, metrically remote location that has also accumulated preshape. This amounts to something like robust estimation and is one aspect of the stabilization of decisions.

In DFT, memory traces reflect not only the probabilities of different peak events, but also their metrics. Probabilities are essentially encoded by the activation level within a preshaped location, whereas the metrics of prior experience is encoded by the location of the preshaped activation. An experiment and associated simulations highlight, how probability and metrics interact (McDowell et al., 2002). Human participants were asked to make center-out pointing movements to visual targets. In each block of trials, only two movement directions occurred,

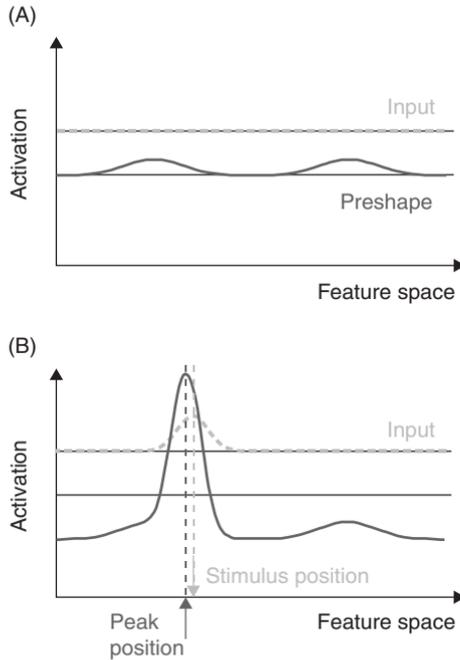


FIGURE 13.7 (A) An activation field (solid line) is preshaped by a memory trace that reflects that two locations have frequently seen activation peaks. This decreases the amount of excitation that is required to generate a peak for the preactivated field locations. (B) When a localized input is applied to the preshaped field, the resulting peak position is slightly biased away from the input specified location toward the closest location specified by the preshape. Metrically distant preactivation does not matter, as it is only within the range of 19 inhibitory interactions.

one of which was elicited frequently and the other rarely. When the two movement directions were metrically far from each other (120°), reaction times to the rare target were longer than reaction times to the frequent target, consistent with the Hyman law (which says that choice reaction time increases with decreasing probability of a choice). When the two targets were metrically close (5°), reaction times for both movement directions were equally fast. The rare movement direction was actually shared across two different blocks. Reaction time to this target was long when it was paired with a metrically far frequent target and short when it was paired with a metrically close frequent target.

Figure 13.8 illustrates the DFT account for this effect. Movement direction is the feature dimension and movement is initiated in the direction encoded by the location in the field at which a self-stabilized peak is generated. Reaction time is predicted by the rate at which activation within the peak rises, shown in the figure through the activation level at the location of the peak. Whenever a peak is created and a response made, the memory trace at the associated location is updated. The frequent movement direction is thus represented by a more strongly preactivated field location than the rare movement direction. This leads to faster buildup of a peak from the

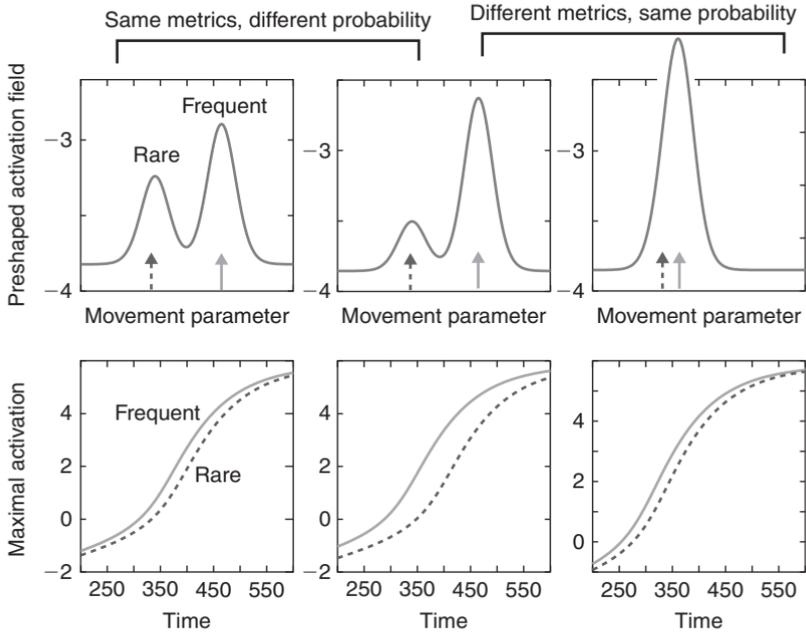


FIGURE 13.8 On top, the preshape of an activation field representing the direction of an upcoming movement is shown. The level of preshape activation at two possible movement directions (arrows) reflects the probability of each movement, higher levels arising for more probable choices (left and middle). When the two movements are metrically close, preshapes overlap, lifting the level of preactivation for the less probable choice. At the bottom, the rise of the maximal level of activation is shown for the case that the frequent (solid) and the rare movement is specified. That rise is faster for the frequent than the rare movement when probabilities and metrics are disparate (middle) but not when probabilities are similar (left) or metrics are close (right).

more strongly preshaped location, explaining the Hyman law (Erlhagen & Schöner, 2002). When the two locations are very close, however, the preshape at the frequent location spills over to the rare location, boosting buildup there and leading to similarly fast buildup time for both choices. The wider implication is that the amount of information (e.g., probability, number of choices, sensory precision) is not the only predictor of choice behavior. The contents of the selection decisions, their metrics, also matter. It is not possible to abstract from the “what,” the specific, embodied and substantive contents of mental representations by focusing only on the “how much,” on the abstract processing of information and its capacity limits.

CATEGORICAL BEHAVIOR FROM CONTINUOUS REPRESENTATIONS

Up to this point we have talked primarily about sensorimotor tasks, in which decisions about continuous feature dimensions needed to be made and the values

of these dimensions estimated. This may entail forms of cognition such as when such estimates need to be stabilized in working memory or when a selection among different possible values is required. Much of cognition, however, may seem to involve primarily categorical behavior and the associated categories. Categorical behavior is required if the environment offers discrete objects such as when we select and reach for one object rather than another, when we name one object rather than another. Words appear categorical in nature and many language tasks seem to require the selection of one of a discrete set of choices, such as when we name an object using one word rather than another.

In the laboratory, categorical behavior is often imposed by asking participants to act on discrete objects (such as pressing one key when stimuli of a certain kind are presented and another when stimuli of another kind are presented) or use discrete responses. Although the sensorimotor tasks reviewed up to now may, in the laboratory, also involve only a small set of discrete possible choices, these are naturally embedded in a continuum (e.g., of possible movement directions). A classification task, in contrast, seems to involve inherently discrete response categories. If we are asked to recognize faces by labeling, we may be compelled to make a discrete selection rather than interpolating between two possibilities (although such interpolation may appear possible at some level of representation, we will come to that). Another way to characterize categorical tasks is to examine the kind of errors that participants can make: are errors graded and metric in nature or are they inherently categorical, for example, "right" or "wrong."

How does DFT deal with inherently categorical behaviors? The key idea is to think of such categories as embedded in underlying continua. In many cases, these may be thought of as arising from the lower level perceptual feature spaces, within which an object may be described. At a neuronal level, cortical feature maps provide a substrate for such an embedding. Population coding has been found in cortical areas as high as IT exactly for the presentations of objects (Young & Yamane, 1992) suggesting that the notion of overlapping neuronal connectivity that gives rise to the notion of peaks along a continuum applies. Psychophysically, most perceptual representations are not strictly categorical, giving access to graded information about the particulars of any given instance of a category (as is true even for the most famous case of categorical perception, the perception of the phonemes of speech, see Massaro, 1987).

Once we recognize that categories may be embedded in this way, the question is how categories may arise from such underlying continua and how categorical behavior may be generated on the basis of continuous DNFs. We have to begin with the latter question to then know what the first question entails. It turns out, there is a simple answer that requires no new mechanisms over those used up to here. Discrete categorical responses may arise from multi-peaked preshape within a continuous activation field. From such graded, subthreshold patterns of pre-activation, self-excited peaks can be generated through the same detection instability discussed earlier. Figure 13.9 illustrates that a simple boost, a homogeneous excitatory input that lifts the activation across the entire field, may push the field

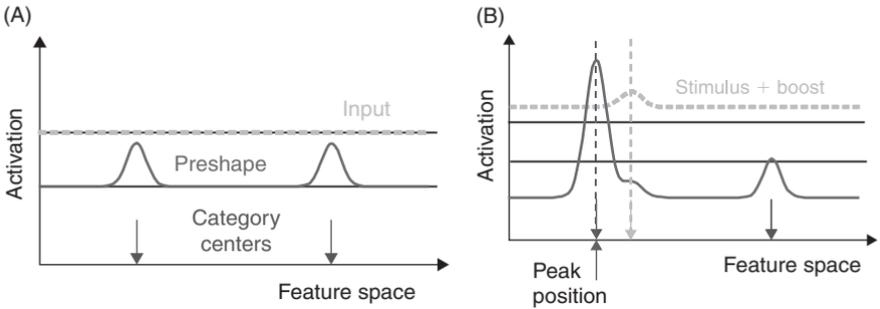


FIGURE 13.9 (A) An activation field (solid line) is preshaped around feature values that correspond to associated categories. In the absence of input (dashed line is at the zero level), the preshaped field remains below the output threshold. (B) Categorical responses are generated by combining a weak localized stimulus input with a homogeneous boost of activation that can be viewed as a “go” signal (dashed line). This lifts all preshape hills above the output threshold and engages both local self-excitation and competition among the potential peaks. The small localized input component biases the competition toward the preshaped category with which it has the greatest overlap. The position of the resulting self-stabilized peak (solid line) is largely determined by the categorical preshape pattern, not by the localized stimulus component.

through the detection instability at one of the preshaped locations. If multiple locations are preshaped, then selection may happen through the same mechanism of lateral inhibition evoked earlier to understand selection in sensorimotor tasks. The input may not be perfectly homogeneous, containing instead some localized structure and thus favoring the selection of the preshaped location that overlaps most with this input. The peak in the field, and thus, the associated behavior, is localized largely over the preshaped locations, however, when this localized component of input is small compared to the amount of preactivation. Comparing the categorical response mode to the response mode used in the sensorimotor scenarios (Figure 13.7), the roles of stimulus input and preshape are reversed. In the latter case, the localized stimulus is dominant, largely determining the peak location and being causal for the initiation of a response (so that there is no need for a separate “go” signal). The preshape makes a minor contribution to the metrics of the representation, biasing the peak toward the preshaped regions. In the categorical response model, in contrast, preshape determines the metrics of the response while the time of response initiation is determined by the homogeneous boost, the “go” signal. The specific, localized stimulus merely biases the competition between the different preshapes, thus selecting the category that will be activated. These two modes are, however, merely limit cases of a continuum, in which the relative strength of preshape and localized stimulus input takes on any intermediate value. The mechanism proposed here to explain categorical responses explains how categorical errors may arise even for unambiguous stimuli. Such errors may arise if the wrong preshape hill, which is not metrically close to the current stimulus, wins the competition. Because the field goes through an instability when the peak is brought up from preshape, it is sensitive to noise and such

an outcome may result due to a fluctuation. This is more likely for smaller differences in input at different possible locations (e.g., because the localized part of the stimulus is weak) or for stronger overlap between preshape patterns (e.g., because the categories are metrically close).

This account also explains why trials in which an error occurs tend to have longer reaction time than trials with a correct response (Luce, 1986). In the dynamic field, the reaction time is determined by the time course of the selection process. If one peak is clearly stronger than all others, it can quickly suppress those others and win the competition. If on the other hand several candidates are almost equally strong, the process of competition starts more slowly, as there is little difference between the forces that act on the single peaks. Even if one of them gains a little advantage, a small amount of noise is sufficient to nullify it. In our account, the strength of different peaks is on average more similar in error trials than in correct ones (as trials with a close competition are more likely to produce errors). Thus, the DNF model will produce longer reaction times for error trials. Furthermore, reaction times of the DNF model will tend to be longer as the number of response categories increases: As more preshape hills compete with each other, the total inhibition gets stronger, slowing down the rise of a single peaks (Erlhagen & Schöner, 2002). Such an influence of the number of categories on the reaction time is experimentally well studied and is captured by Hick's law (Luce, 1986).

The generalization of this result is the Hyman law, of course, according to which reaction time increases with decreasing probability of a choice. We showed earlier how the Hyman law interacts with the metrics of choices when peaks are generated from a localized input representing an imperative signal. In that case, metrically close choices have faster reaction times irrespective of their probability (Erlhagen & Schöner, 2002). This is actually a somewhat counter-intuitive result. More common is the distance effect, in which the decision between two choices takes longer if the choices are more similar and metrically close (Anderson, 1995). As illustrated in Figure 13.10, the distance effect falls out of the DNF account of selection in the categorical mode dominated by preshape (Wilimzig, 2006). Only when the preshape is bimodal are categorical responses possible. Everything else being the same, the stimulus specifying either choice overlaps more with the other choice, so that both choices are activated to a larger extent. Mutual inhibition is more strongly engaged and slows responding down until one of the choices falls below the output level.

The detection instability driven by a homogeneous boost to the field is capable of amplifying small graded inhomogeneities into macroscopic stable states that can begin to impact behavior. This has far reaching implications for what is required to learn categories: Essentially, acquiring a graded preshaping along a feature dimension with local maxima near the centers of categories is sufficient to respond categorically to graded inputs. Categories thus emerge naturally from underlying continuous feature representations through graded, incremental learning rules such as the memory trace mechanism described earlier (Figure 13.6) or a generic Hebbian rule.

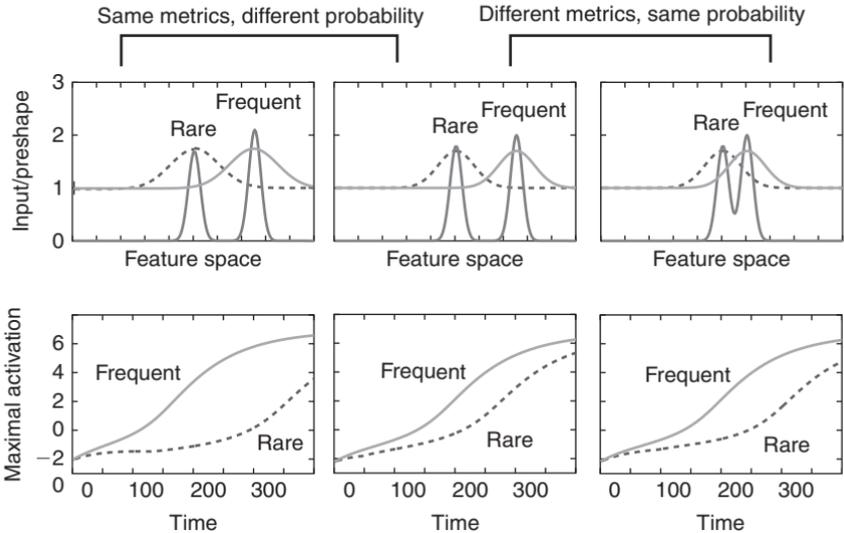


FIGURE 13.10 This figure is analogous to Figure 13.8 but now based on the DNF model in the categorical response mode. Top: Preshape (fat solid line) reflecting a frequent and rare choice along a feature dimension is shown when the difference between the probabilities is small (left), or large (middle and right). The two choices are either metrically far (left and middle) or metrically close. The stimulus specifies either the rare (dashed) or frequent choice (thin solid) and contains a homogeneous boost component. Bottom: The maximal activation at the site specified by the stimulus rises faster for the frequent than for the rare choice. This difference is larger when the probability difference is larger (left compared to middle), but increases again when the choices are metrically close, the opposite effect compared to Figure 13.8 (Figure adapted from Wilimzig (2006)).

EMBODYING DYNAMIC NEURAL FIELDS ON AUTONOMOUS ROBOTS

We have emphasized the concept of stability as a prerequisite for understanding how cognition may emerge in embodied and situated systems that are continuously linked to structured environments through sensory inputs. But how do the concepts fare when a real body is controlled based on real sensory information? One way to evaluate is to implement DNF models on autonomous robots and investigate how DNFs cope with continuously changing and noisy input and how DFT architectures generate consistent and flexible behaviors. If simple, neuronally plausible sensory and motor processes are sufficient to enact a DNF model, then this proves that there are no hidden problems in the interface between the DNFs and the sensory and motor surfaces. This is not trivial. Many a model of cognition makes strong demands on both ends of sensation and motor control. Some connectionist models, for instance, postulate that a specific neuron represents a particular kind of object (e.g., see Munakata, 1998). Recognizing objects on the basis of visual information is, however, a well-known and nontrivial problem. So there is something hidden in the interface here (which may seem

particularly relevant for a model that addresses object permanence such as Munakata, 1998). Another aspect of such implementations is that they probe the real-time autonomy of behavior. Is the robot capable of behaving continuously, going from one state to another, propelled by its inner dynamics and the sensory information it actively acquires from its environment? This entails not only the issue of closed loop control in the real world but also the continuous, asynchronous operation of cognitive processes. This may be contrasted with information processing models, in which behavior is generated only as a response to stimulation, so that time is (implicitly or explicitly) parsed into input–output cycles.

There is more to be gained from robotic implementations beyond such feasibility proof. As a heuristic device, robots may reveal to us all that needs to be specified, and all that can go wrong when a particular behavior is generated. This may motivate new research questions. Examples in kind are calibration and homeostasis, both of which are often left unaddressed in more abstract models of cognitive function. Heuristics also works the other way round: Robot demonstrations of a particular function may be possible without invoking a particular concept. For instance, perseverative reaching can be modeled without using an explicit object representation (Schöner & Dineva, 2006). This does not prove that babies do not have object representations; however, it means that perseverative reaching is not necessarily an index of such representations.

Robotic implementations of DFT may also be pursued simply as a competitive approach to autonomous robotics, evaluated based on the performance of the solutions, on their robustness, ease of design, and so on. It is in this most applied sense that the first robotic demonstrations of DNF models were made (Engels & Schöner, 1995; Schöner et al., 1995; Bicho et al., 2000). Here, we illustrate how DFT can work in a real-world setting using an example close to the issues discussed in the last section, that is, the visual recognition of objects (Faubel & Schöner, in press). This is anchored in a scenario, in which a service robot interacts with a human user within a shared workspace. The robot system learns to recognize a number of objects from a single or a small number of views, associating the object with a label. The ultimate goal is to interact with the user, recognize and name objects, reach for them, manipulate them, and so on. To simplify the task, it is assumed that the number of objects to be memorized is limited, and that the environment is uncluttered and known to the robot.

As first step of the object recognition, a simple segmentation algorithm is applied to the visual input to detect objects on the table surface. Then, several low-level features are extracted for each object: Its size, the aspect ratio as a measure of its shape, and the color, described by a histogram of hue values (Figure 13.11). Each of these features serves as input to a two-dimensional label feature field. In these fields, an association between a feature value and a matching label is realized through hills of preshape: The features are represented along the first dimension of the field, whereas the labels are represented along the second one. For each label, a hill of preshape is created during the teaching procedure around the appropriate feature values. If an object is presented for recognition, its feature

Decision layer

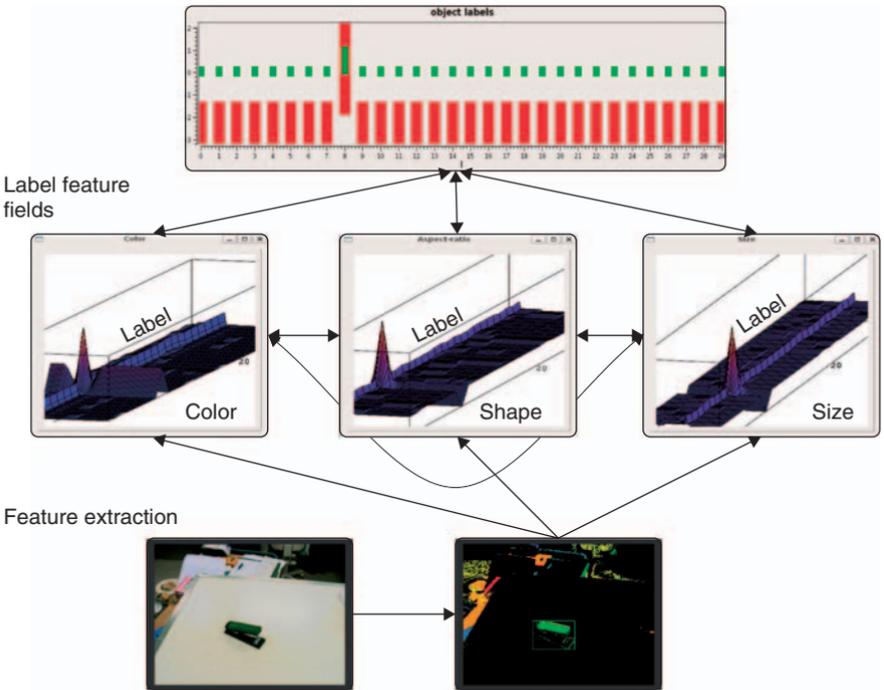


FIGURE 13.11 For the object recognition procedure, three simple object features are extracted from an image and fed into a network of coupled dynamic fields. The interactions within and between these fields result in the selection of one label, which can be read out from the decision layer. (See color plate)

values are fed into the two-dimensional field along the first dimension, creating a ridge of activation that will overlap with some of the preshapes. As for the categorization behavior, the preshape peaks compete with each other through local excitation and global inhibition after the field receives a “go” signal (a homogeneous boost of activation throughout the field), and the preshape that has the greatest overlap with the input is most likely to win the competition. A two-dimensional interaction kernel is used to implement these field interactions. As no metrics is defined for the different labels, the kernel profile is simplified along this dimension such that each label excites only itself and homogeneously inhibits all others.

The output of all label feature fields is fed into a one-dimensional decision layer that has one node for each label. The same simplified interactions are used in this layer to enforce a decision for one label, which is returned as the result of the recognition process once the activation of one node passes a preset threshold. This feedforward processing is augmented by lateral and feedback connections between the fields: The same labels in different label feature fields excite each other, such that the selection of a certain label in one field gives it a competitive

advantage in the other fields. In addition to this, there is inhibitory input from the decision layer: The output of one label in this layer suppresses all other labels in the label feature fields. This ensures that after a decision, all peaks that do not match the selected label are extinguished, which is important for learning.

To teach a new object, its label is associated with one node in the decision field, the corresponding columns in the label feature fields are preactivated, and the object is presented to the robot's camera. Peaks will emerge in the label feature fields for the feature values extracted from the visual input and preshape is laid down at these positions. If similar features are found in later recognition trials, the learned label will be activated through the association mechanism of the two-dimensional fields. It is possible to continue the learning process, that is the buildup of preshape, in later trials, either after a correct recognition or after corrective input from the user. This way, the distribution of the feature values for different views of the same object can be reproduced in the preshapes.

The DFT-based object learning system was tested on 30 objects, presented in several different positions on the table and different orientations during teaching and recognition trials. A mean recognition rate of 88% was achieved after teaching each object in eight different views. In every teaching trial, the object recognition procedure was performed as well, with the correct label being given to the robot if the recognition failed. It is noteworthy that with this setting, on average only 2.8 user interventions (corrections of wrong responses) were necessary per learned object.

One important aspect of this model is the role of the lateral and feedback projections between the fields. Without them, each label feature field would independently select the label that best matches the current input and the final decision would be made between these candidates. With the lateral projections, a label that yields a good match in several features receives extra input, and thus it can win the competition in all fields even if it does not yield the best match in any of them. A second effect aids the selection of the correct label: For those feature dimensions where several labels are closely competing with each other, the selection process is slowed down (as discussed earlier for categorical responses in general). The selection is also slow in those fields where the candidate labels have a broad and flat preshape, which results from a high variance in that feature under different views. Due to these effects, a decision is first made in those fields where the stimulus input is unambiguous, and the other fields are then pushed to select the same feature by inter-field excitation. Once more, this desirable behavior emerges directly from the field interactions, and no superordinate structure is needed to select those features that are most significant in the current situation.

CONCLUSIONS

At the same time as it illustrates the embodied and situated nature of DFT accounts, the preceding example provided an outline for how the ideas reviewed

in this chapter may scale up from the most elementary forms of cognition toward more complete, “higher” acts of cognition. Other work, not reviewed here, has similarly established the scalability of the concepts by accounting for the emergence of working memory for planned actions (Thelen et al., 2001), spatial working memory (Schutte et al., 2003) and spatial cognition more generally (Simmering et al., 2007), visual working memory (Johnson et al., 2008) and infant habituation (Schöner & Thelen, 2006).

In many of these cases, DNFs have to be combined across different feature dimensions. The dynamic ideas of coupling and stabilization work seamlessly, supporting complex architectures of DNFs. This includes the transformation of sensory into motor representations. Associating different feature dimensions is a natural task of neuronal networks and the dynamics of neural fields accommodate that basic neuronal functionality.

We are expressing some confidence that DFT concepts scale up to forms of cognition more traditionally at the core of concern of cognitive scientists. That said, a subtle, but fundamental issue must be recognized. The complete dynamical system characterizing an organism and its nervous system in a given environment and task context has rich internal structure and includes coupling through the outer world as well. The functionally significant states of such complete systems emerge as attractor solutions from these dynamics under the appropriate circumstances, depending on the behavioral (or activation) history and on an appropriately structured environment (Schöner & Dineva, 2006). These functions are not fixed and they do not “sit somewhere” until activated. They are simply emergent properties of the dynamical system. Individuals may differ in the circumstances that are required to bring about such functions. Individual differences may initially arise from chance events but may then become amplified over time due to the adaptive and learning capacity of dynamical systems. Because dynamical systems can amplify small graded differences into qualitatively different states, this implies a limitation of predictive power. Conversely, the same system may behave differently, exhibiting or not exhibiting a particular function, depending on the task context. Learning does not necessarily install function in a definite and fixed way. Learning may more appropriately be viewed, in Dynamical Systems Thinking, as a process that eases the constraints on the environmental and task conditions under which a function may emerge. Thus, the very nature of Dynamical Systems Thinking makes that the accounts delivered may differ from expectations built on the tradition of information processing or even connectionist thinking. It appears unlikely that there would be something like the ultimate dynamical systems model of the mind, a fixed, if complicated architecture from which behavior can be predicted. A metaphor closer to what Dynamical Systems Thinking may provide is the notion of the brain as a very high dimensional, complex dynamical system, built neuronally, but potentially coupled so closely to the environment through its own effector and sensor systems that these become part of the dynamical system. Understanding such a system amounts to understanding the constraints on its inner structure, flexible

though it may be. What dynamicists try to do is find the tasks and elementary behaviors, in which such structure comes to light and in which projection from the high dimensional state space into much lower dimensional subsystems is possible. In such exemplary situations they identify the principles that describe how such subsystems form, stabilize, and adapt. Those principles provide the basis for extrapolating to the myriad and open-ended ways in which the mind may shape and reshape.

REFERENCES

- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27, 77–87.
- Anderson, J. R. (1995). *Cognitive Psychology and its Implications*. New York: W. H. Freeman and Company.
- Bastian, A., Riehle, A., Erlhagen, W., & Schöner, G. (1998). Prior information preshapes the population representation of movement direction in motor cortex. *Neuroreports*, 9, 315–319.
- Bastian, A., Schöner, G., & Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience*, 18, 2047–2058.
- Bicho, E., Mallet, P., & Schöner, G. (2000). Target representation on an autonomous vehicle with low-level sensors. *The International Journal of Robotics Research*, 19, 424–447.
- Bizzi, E. & Mussa-Ivaldi, F. A. (1990). Muscle properties and the control of arm movement. In D. N. Osherson, S. M. Kosslyn, & J. M. Hollerbach (Eds.), *Visual Cognition and Action* (pp. 213–242). Cambridge, MA: The MIT Press.
- Braun, M. (1993). *Differential Equations and Their Applications*, 4 ed. New York: Springer Verlag.
- Cisek, P. (2006). Integrated neural processes for defining potential actions and deciding between them: A computational model. *Journal of Neuroscience*, 26(38), 9761–9770.
- deCharms, R. C. & Zador, A. (2000). Neural representation and the cortical code. *Annual Reviews of Neuroscience*, 23, 613–647.
- Dorris, M. C., Olivier, E., & Munoz, D. P. (2007). Competitive integration of visual and preparatory signals in the superior colliculus during saccadic programming. *Journal of Neuroscience*, 27(19), 5053–5062.
- Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience Supplement*, 3, 1184–1191.
- Engels, C. & Schöner, G. (1995). Dynamic fields endow behavior-based robots with representations. *Robotics and Autonomous Systems*, 14, 55–77.
- Erickson, R. P. (1974). Parallel “population” neural coding in feature extraction. In F. O. Schmitt & F. G. Worden (Eds.), *The Neurosciences—Third Study Program* (pp. 155–169). Cambridge, MA: MIT Press.
- Erlhagen, W. & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, 109, 545–572.
- 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, 53–66.
- Faubel, C. & Schöner, G. (2008). Learning to recognize objects on the fly: A neurally based dynamic field approach. *Neural Networks*, 21, 562–576.
- Feldman, A. G. (1986). Once more on the equilibrium point hypothesis (λ -model) for motor control. *Journal of Motor Behavior*, 18, 15–54.
- Georgopoulos, A. P. (1991). Higher order motor control. *Annual Reviews of Neuroscience*, 14, 361–377.

- Goodale, M. A., Pélişson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, *320*, 748–750.
- Hock, H. S., Kogan, K., & Espinoza, J. K. (1997). Dynamic, state-dependent thresholds for the perception of single-element apparent motion: Bistability from local cooperativity. *Perception and Psychophysics*, *59*, 1077–1088.
- Hock, H. S., Schöner, G., & Giese, M. A. (2003). The dynamical foundations of motion pattern formation: Stability, selective adaptation, and perceptual continuity. *Perception and Psychophysics*, *65*, 429–457.
- Hogan, N. (1990). Mechanical impedance of single- and multi-articular systems. In J. M. Winters & S. L.-Y. Woo (Eds.), *Multiple Muscle Systems* (pp. 149–164). New York: Springer Verlag.
- Janke, D., Erlhagen, W., Dinse, H. R., Akhavan, A. C., Giese, M., & Steinhage, A. et al. (1999). Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex. *Journal of Neuroscience*, *19*, 9016–9028.
- Johnson, J. S., Spencer, J. P., & Schöner, G. (2008). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology*, *26*, 227–251.
- Kopecz, K. & Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural, dynamic fields. *Biological Cybernetics*, *73*, 49–60.
- Luce, R. D. (1986). *Response Times*. New York: Oxford University Press.
- Massaro, D. W. (1987). *Speech Perception by Eye and Ear: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- McDowell, K., Jeka, J. J., Schöner, G., & Hatfield, B. D. (2002). Behavioural and electro-cortical evidence of an interaction between probability and task metrics in movement preparation. *Experimental Brain Research*, *144*, 303–313.
- Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A pdp model of the ab task. *Developmental Science*, *1*(2), 161–184.
- Ottes, F. P., Gisbergen, J. A. M. van, & Eggermont, J. J. (1984). Metrics of saccade responses to visual double stimuli: Two different modes. *Vision Research*, *24*, 1169–1179.
- Perko, L. (1991). *Differential Equations and Dynamical Systems*. Berlin: Springer Verlag.
- Port, R. & Gelder, R. van (Eds.) (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.
- Schöner, G. (2008). Dynamical systems approaches to cognition. In R. Sun (Ed.), *Cambridge Handbook of Computational Cognitive Modeling*. Cambridge, UK: Cambridge University Press.
- Schöner, G. & Dineva, E. (2006). Dynamic instabilities as mechanisms for emergence. *Developmental Science*, *10*, 69–74.
- Schöner, G. & Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science*, *239*, 1513–1520.
- Schöner, G. & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*(2), 273–299.
- Schöner, G., Dose, M., & Engels, C. (1995). Dynamics of behavior: Theory and applications for autonomous robot architectures. *Robotics and Autonomous Systems*, *16*, 213–245.
- Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development*, *74*, 1393–1417.
- Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*(4468), 390–398.
- Simmering, V. R., Schutte, A. R., & Spencer, J. P. (2007). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research*, (doi:10.1016/j.brainres.2007.06.081).
- Spencer, J. P. & Schöner, G. (2003). Bridging the representational gap in the dynamical systems approach to development. *Developmental Science*, *6*, 392–412.
- Spencer, J. P., Simmering, V. R., Schutte, A. R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory

- of spatial cognition. In J. Plumert & J. P. Spencer (Eds.), *The Emerging Spatial Mind*. New York, NY: Oxford University Press.
- Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Brain and Behavioral Sciences*, *24*, 1–33.
- Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, *13*(2), 256–271.
- Wilimzig, C. (2006). *Dynamische Feldtheorie der kognitiven Informationsverarbeitung*. Unpublished doctoral dissertation, Ruhr-Universität Bochum.
- Wilimzig, C., Schneider, S., & Schöner, G. (2006). The time course of saccadic decision making: Dynamic field theory. *Neural Networks*, *19*, 1059–1074.
- Wilson, H. R. & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*, 55–80.
- Young, M. P. & Yamane, S. (1992). Sparse population coding of faces in the Inferotemporal cortex. *Science*, *256*, 1327–1331.
- Zipser, D. (1991). Recurrent network model of the neural mechanism of short-term active memory. *Neural Computation*, *3*, 179–193.