[Chapter 3 of the book: DYNAMIC THINKING — A Primer on Dynamic Field Theory
GREGOR SCHÖNER, JOHN P. SPENCER, AND THE DFT RESEARCH GROUP,
Oxford University Press, New York, 2016]

3

Embedding Dynamic Field Theory in Neurophysiology

SEBASTIAN SCHNEEGANS, JONAS LINS, AND GREGOR SCHÖNER

In the previous chapter, we introduced the dynamic field (DF) as a mathematical concept and as a behavioral model. In particular, we described how peaks of activation constitute attractor states of the dynamical system that serve as units of representation. We then showed how the transitions between different configurations of activation peaks can form the building blocks for generating behavior by implementing different forms of decisions. Moreover, we claimed that the DF is a neural model and that the dynamics of activation peaks can therefore explain biological mechanisms of behavior generation.

At first glance, however, the concept of a continuous activation distribution may not appear very biological. It lacks some of the key components of what is understood to be neural processing in biological systems: There are no actual neurons described in the model, nor axons or synapses, and activity is not expressed through action potentials. Moreover, the form of representation in DFs is conceptually very different from what is typically used in models of neural processing, such as classical neural networks. In neural networks, the representations at each level are typically complex patterns of activation. Learning procedures are often aimed at minimizing the correlation between the activation values of different neurons so as to maximize the amount of information retained in the model representation. The resulting activation patterns are described by high-dimensional vectors and are not easily reducible to a simpler, more comprehensible format. In contrast, in the DF, the neural interaction functions actively create a high correlation of activation values at neighboring positions. And what is represented in a DF can be described—at least at a qualitative level—through a few discrete values that give the positions of the peaks and are easily interpretable in terms of behavioral variables.

This may lead to the impression that the relationship between DFs and biological neural systems takes merely the form of an analogy—that the concept of an activation field is in some way inspired by neural activity, but that it does not actually implement a form of biological neural processing. In this chapter, we will show that this is not the case. First, we will take a closer look at neural representations in biological systems. We argue that the level of population activation is the most appropriate level to elucidate the link between neural processing and behavior. We will show how neural populations represent behavioral variables through the distribution of activation among them and discuss the concept of population coding. We will show some well-studied examples of population representations in sensory and motor areas of the brain and describe empirical results that link experimental manipulations of activation distributions in these areas to behavioral effects.

Next, we will introduce an analysis method of electrophysiological neural data called *distribution of population activation* (DPA). This method takes the firing rates of a group of neurons from a population code representation and transforms them into a continuous distribution of activation over a feature space, using the neurons' measured or estimated tuning curves. We will describe the construction of the DPA in detail for two examples, namely, the activity patterns in cat visual cortex evoked by simple visual stimuli, and preparatory activity for reach movements in the motor and pre-motor cortex of macaque monkeys. The results of DPA analysis show peak-like activation patterns in both the sensory and the motor areas that reflect metric properties of visual stimuli and planned reach movements, respectively.

Moreover, DPA analysis of the population response in visual cortex reveals signatures of

interactions effects. In Chapter 2, we described how such interactions bring about the activation dynamics in DFs that form peaks and create decisions. Here we will show that lateral interactions in DFs are consistent with empirical data and can account for the observed activation patterns in the visual cortex. In this context, we will present an extension of the basic DF model, the two-layer field. The two-layer field reflects more closely the biological connectivity within neural populations and is particularly aimed at capturing the temporal details of population dynamics. With this tool, we can also demonstrate how to fit activation patterns for the preparation of reach movements in the motor cortex with a DF model.

The analysis method of DPA plays a key role in all of this by bringing empirically measured population responses into the same format used in DF models. This makes it possible to directly compare activation patterns in DF models with neural data. In particular, this method allows us to make testable predictions from DF models about activation patterns in biological neural populations. The DPA method thereby provides the neural grounding for the dynamic field theory (DFT), establishing a direct link between the level of neural activity and DF models of behavior and cognition.

LINKING NEURAL ACTIVATION TO PERCEPTION, COGNITION, AND BEHAVIOR

This section concerns the link between neurophysiology and things that actually matter to living, behaving biological agents like you and me. Is this apple green or red? Where do I have to move my hand to grab it? Some aspect of neural activation must reflect the state of affairs on this macroscopic level—the level of perceptual decisions, cognitive states, and overt behavior. As presented in the introduction, we believe that this role is played by patterns of activation in neural populations. To substantiate this claim, we need to take a brief detour to the realm of single neurons, and then work our way up to population-based representations.

To determine the link between the activity of a single neuron and external conditions, neurophysiologists record the spiking of the neuron via a microelectrode placed near (or within) the cell while varying sensory or motor conditions in a systematic fashion. This could mean, for instance, varying the color or position of a visual stimulus or, in the motor case, varying the direction of a limb movement that an animal has to perform. Not all neurons are sensitive to all parameters, so the first step is to determine which parameters cause the neuron to change its activity level. When we find a parameter that reliably affects the spike rate of the recorded neuron, we can proceed to assessing the exact nature of the relationship. In order to do this, the parameter value is varied along the underlying dimension and the spike rate for each sample value is recorded. The results of this procedure can be visualized by plotting spike rate against the parameter dimension. An idealized function may be fitted to the data points, interpolating spike rate between sample values. The resulting curve is called the *tuning curve* of the neuron.

This technique has revealed that, throughout the brain, many neurons share a roughly similar type of mapping between parameter dimension and spike rate, which is characterized by Gaussian-like tuning curves (Figure 3.1). That is, they fire most vigorously for a specific "preferred" parameter value, while spike rate declines with rising distance from that value, reaching the neuron's activity baseline for very distant values.

A classic example for these characteristics can be found in the visual cortex, where many cells respond strongly to bars of light of a particular orientation and reduce their firing as the angle of orientation deviates from that preferred value (Hubel & Wiesel, 1959, 1968). Visual cells show tuning along other feature dimensions as well, such as color (Conway & Tsao, 2009), shape (Pasupathy & Connor, 2001) or the direction of motion (Britten & Newsome, 1998). Neurons in nonvisual areas exhibit similar properties, such as cells in auditory cortex that are tuned to pitch (Bendor & Wang, 2005), or cells in somatosensory cortex that are tuned to the orientation of tactile objects (Fitzgerald, 2006). The most common scheme, however, is tuning to locations in physical space. In sensory areas, most cells are tuned to the

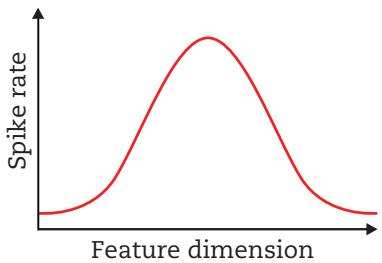

FIGURE 3.1: Schematic illustration of an idealized tuning curve.

position of stimuli on the sensory surfaces, such as the retina or skin. For such spatially tuned sensory neurons, the range where the tuning curve differs from the activity baseline is often referred to as the *receptive field* of the cell, emphasizing that the cell's sensitivity is restricted to a specific region of physical space. In turn, the structure of the tuning curve in that range is called the *receptive field profile* (Jones & Palmer, 1987; Sherrington, 1906). Spatial tuning is found in motor areas as well, where neurons are tuned to locations in motor space, such as hand movement targets (Georgopoulos, Kalaska, Caminiti, & Massey, 1982) or saccade endpoints (Lee, Rohrer, & Sparks, 1988). Generally, neurons tend to be tuned along more than one dimension at the same time (e.g., two dimensions of retinal space and orientation).

Knowing about the typical response schemes of single neurons, only one additional ingredient is missing to make the step to population activation. This ingredient is the scatter of tuning curves across the underlying parameter dimensions. Typically, there are many neurons with disparate preferred values for each of these dimensions, so that the tuning curves collectively cover the entire dimension. Together with the broad extent and large overlap seen in cortical tuning curves, this entails that a single input value to the population, say, a single color in the visual field, activates a large number of neurons. Thus, rather than activating only one neuron, even a single color input evokes a distribution of activation over the population of color-sensitive neurons.

The question, then, is how is this distribution "used" by downstream neural areas? Intuitively, the representation of our color could still be a matter of a single neuron, as it seems possible to discern the value from the identity of the most active cell, simply discarding the rest of the distribution as epiphenomenal activation. This winner-take-all scheme faces some problems, however. One is its low robustness against noise. An estimate based on only one or a few neurons would be highly susceptible to the variability of neural firing. Which neuron is most active would change rapidly due to noise, and so would the estimate of the color. The mechanism lacks what seems to be a critical feature of perception, cognition, and behavior—stability. The other major problem is that of ambiguity. With bell-shaped tuning curves, spike rate is ambiguous in that a particular rate may refer to either one of two values (see Figure 3.1). Even worse, most neurons are sensitive to more than one parameter dimension, making their tuning curves multidimensional and their spike rate even more ambiguous. With a two-dimensional Gaussian tuning curve, for example, a particular spike rate may refer to any position on a circle surrounding the cell's preferred value.

So, in sum, single cells do carry some information about the kind of events that interest us, but each neuron provides only a fraction of the full picture. This view receives additional support from explicit measures of the predictive power of single-cell responses for actual psychophysical decisions (Cohen & Newsome, 2009). The single-neuron level is thus not the level we want to consider when trying to find a reliable link between neural activation and the macroscopic neural decisions that bring about concrete, observable behavior.

The alternative is to widen the scope to a multi-neuron or *population* level. This seems a reasonable thing to do, given that both of the above problems stem from basing an estimate on too few neurons. Unsurprisingly, then, the idea has been long-standing that perceptual and behavioral events are captured by patterns of activation within populations of neurons rather than by single neurons. The basic rationale behind *population coding* is that the properties of perceptual, behavioral, and cognitive events are reflected by the distribution of activation over populations of tuned neurons (Erickson, 1974). Figure 3.2 provides a simple outline of this idea.

Figure 3.2a shows the tuning curves of three hypothetical neurons A, B, and C—let's say they are tuned to color. Values 1, 2, and 3 then correspond to different hues that elicit different responses in the three neurons. When hue value 1 is presented, for example, neuron A responds only weakly, but still stronger than the other two neurons. Value 2 is close to neuron A's preferred hue and therefore drives the neuron strongly, while B responds weaker and C is nearly silent. Note that each hue drives multiple neurons. Figure 3.2b illustrates the problem of ambiguity by showing each neuron's spike rate in response to the different hue values. In this example, the response of neuron B is identical for hue value 2 and hue value 3, making it impossible to discern from its activity which of the two colors is present (even in the absence of noise). Figure 3.2c contrasts this by reordering the responses by hue value, that is, by showing the distribution of activation over our toy population for

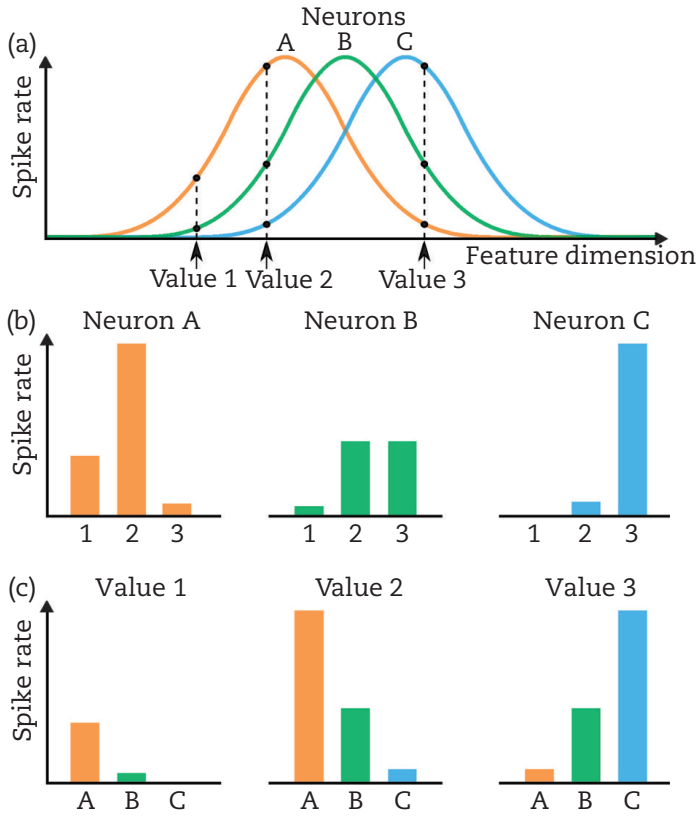

FIGURE 3.2: Neural representations of metric values. (*a*) Tuning curves of three hypothetical neurons A, B, and C. Values 1, 2, and 3 are different values of a sensory or motor parameter that the neurons respond to, according to the tuning curves. (*b*) Responses to the different values, ordered by neurons. On the single-neuron level, different parameter values can evoke identical responses (e.g., value 2 and 3 in neuron B). (*c*) Same schema as in b, but ordered by input values, thus showing activation distributions in the population evoked by each of the three values. The distributions are unique for each value.

each input value. In contrast to the individual neurons' activation, the distribution is unique for each of the three values, thus specifying the respective value unambiguously. So by using the aggregated activation of many neurons it is possible to overcome the problem of ambiguity. In our example, the actual hue can be derived from the activity of multiple differently tuned neurons—just as a target location on a street map can be inferred from its distance to multiple other locations.

Conveniently, the solution to the noise problem comes easily with this scheme, because the random variability of individual neurons tends to be averaged out when activation is integrated across many neurons. Thus, population coding solves both problems at once. However, to see if the principle actually applies in the nervous system, we need to assess whether population activation is really linked to behavior as closely as we claim (where behavior may also indicate the outcome of perceptual decisions or other cognitive processes). The crucial questions are: Does population activation really predict behavior more reliably than single neurons? Do all active neurons impact behavior? A large body of evidence suggests that the answer to both questions is yes (e.g., Cohen & Newsome, 2009; Georgopoulos, Kettner, & Schwartz, 1988; Groh, Born, & Newsome, 1997; Lee et al., 1988; Nichols & Newsome, 2002). We will consider two exemplary experiments.

Lee and colleagues (1988) demonstrated population coding in the superior colliculus, a subcortical structure that plays a decisive role in the preparation and initiation of saccades (rapid gaze shifts that serve to bring a location from the retinal periphery to the fovea). The superior colliculus is organized topographically; that is, visual space is mapped orderly onto its surface. Tuning to the angular direction of saccades varies along its lateral–medial axis, and with respect to saccade amplitude, in an anterior–caudal direction. Unfolding and flattening the superior colliculus thus yields a roughly rectangular map of saccadic motor space, with amplitude on one axis and direction on the other (Figure 3.3).

Following the typical scheme, the tuning of neurons in the superior colliculus is broad, so that a large number of neurons fire for each saccade. Given the topographical layout we can expect that when the metrics of a saccade are specified, the active neurons are clustered together in one spatial region of the superior colliculus. This was exactly what Lee and colleagues found when recording the activity of cells in the superior colliculi of monkeys. Prior to each saccade a circular blob of activation forms in the topographical map. Neurons located in the region of the map that corresponds to the saccade target are most strongly activated, while the level of activation decreases toward the blob's periphery. The red circle in Figure 3.3a outlines the approximate extent of an activation blob that results in the saccade illustrated by vector A (black arrow on the right). B and C mark the centers of activation blobs that result in the saccade vectors labeled accordingly.

It seems intuitively clear that these localized peaks indicate the metrics of saccades, but to test the population coding hypothesis we need to determine whether the actual saccade target really depends on *all* active neurons, including the weakly activated ones at the periphery of the blob. To examine this, Lee and colleagues induced saccades by presenting visual targets to their monkeys while inactivating either peripheral or central portions of the activation blob with a local anesthetic. They then assessed how this deactivation impacted the resulting saccades.

Figure 3.3b shows the result of deactivating the center of the blob (blue dot), that is, the most active neurons. The resulting saccade (red arrow) is identical to the one without deactivation. Apparently,

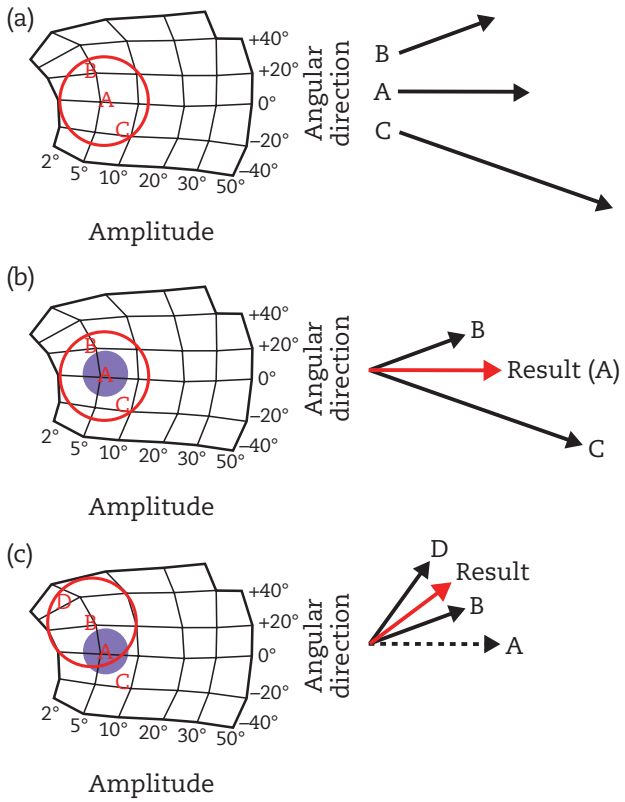

FIGURE 3.3: Results of experiments of Lee et al. (1988). Each subfigure shows a flattened version of the topographical motor map of the left superior colliculus. Red letters mark the centers of activation blobs observed for different saccades, which are depicted by the correspondingly labeled vectors on the right. Red circles mark the approximate extent of activation blobs centered on the middle of the circle. Blue dots mark regions that were deactivated in the experiments. (*a*) Activation centers observed for the saccades on the right, without deactivation. (*b*) A visually evoked saccade to the target described by vector A is not altered by deactivating the blob center. The weighted average of B and C provides a sufficient estimate of A. (*c*) A visually evoked saccade to the target described by vector B is altered when the peripheral blob region that corresponds to A is deactivated. The resulting saccade is now guided by a weighted spatial average of B and D. Adapted by permission from Macmillan Publishers Ltd: *Nature*, Lee, C., Rohrer, W. H., & Sparks, D. L., Population coding of saccadic eye movements by neurons in the superior colliculus, 332(6162), 357–360, copyright 1988.

the average of the remaining ring of activation provides a sufficient, unbiased estimate of the saccade parameters (suggested by the fact that the actual saccade vector is the average of B and C). This is a first hint that weakly activated neurons influence motor outcomes. However, of greater interest is the outcome of deactivating peripheral blob regions, illustrated in Figure 3.3c. Again, the region around A is deactivated, but this time the visual target is at another location, B. Because the neurons at the center of the blob are active as usual, a winner-take-all scheme would predict that the saccade is unaffected and lands at B. Instead, the saccadic endpoint is shifted away from the visual target toward the preferred values of the still active population (red arrow). Thus, the decisive variable seems to be not greatest activation but the overall location of the activation blob, with more active neurons being weighted more strongly when determining it. Taken together, this suggests that a spatial averaging scheme is at work in the superior colliculus, with all active neurons contributing.

Another line of evidence shows that population coding is also employed in areas that are non-topographically organized. Neurons in the arm area of the motor cortex are tuned to a continuous metric dimension, namely, to the direction of arm movement, but their spatial arrangement in the cortex does not follow any obvious spatial scheme. The tuning characteristics were examined by Schwartz, Kettner, and Georgopoulos (1988). They recorded the activity of motor cortical units while monkeys executed an arm movement task. In each trial, the monkey had to move its hand from a central starting button to one of eight target buttons. The target buttons were distributed in three-dimensional space, equidistant from the starting button, sampling the continuum of possible movement directions. Schwartz and colleagues found that each cell responds maximally to a specific preferred direction. As the angle between this preferred direction and the actual movement direction increases, spike rate declines, following a cosine tuning curve. Here, as in the superior colliculus, neurons are tuned very broadly, so that any particular movement direction activates many neurons, including neurons that have preferred directions very different from the current one.

In the next step, Georgopoulos, Kettner, and Schwartz (1988) examined whether movement direction really depends on the entire active population. As the motor cortex is not organized topographically, however, it is not possible to inactivate specific regions of the motor map—anesthesia administered to a patch of cortex would deactivate neurons with very different preferred directions. To overcome this issue, a vector was derived for each neuron from the directional tuning data obtained in the first experiment describing the respective neuron's preferred movement direction. This made it possible to construct a population vector (Box 3.1, Figure 3.4) for each observed movement direction.

The population vector is obtained by summing the preferred direction vectors of all neurons that were active for a movement in the considered direction. Importantly, before summing the vectors, each neuron's preferred direction vector is weighted by the neuron's spike rate. Thus, more active neurons contribute more strongly to the population vector. Finally, the population vector for each movement was compared to the actual arm movement that the monkey performed.

If *all* active neurons are relevant for specification of the movement, then a prediction of that movement should become more accurate the more neurons are included in the population vector. Georgopoulos and colleagues found that this is indeed the case, strongly suggesting that the motor cortex does use population coding.

Although it is not possible to observe a spatially circumscribed blob of activation in the motor cortex, due to its non-topographical layout, a peak can be derived by taking as a basis the dimensions along which the neurons are tuned. Viewed as a distribution over the space of possible movement directions, activation takes the form of a perfectly localized peak that specifies the current value by its position in that space. Thus, although the peak is distributed over physical space in the cortex, it is functionally equivalent to the localized peaks in the superior colliculus.

These examples are prototypical for many areas in the nervous system. The groundbreaking findings have sparked interest in the concept of population representations, and subsequent research has shown that, in addition to increasing robustness and reducing ambiguity, the properties of population representations satisfy basic requirements of perception, behavior, and cognition. For example, neural populations can support multiple activation peaks, indicating several values simultaneously (Harris & Jenkin, 1997; Nichols & Newsome, 2002; Pasupathy & Connor, 2002; Treue, Hol, & Rauber, 2000). This may set the stage for things like

BOX 3.1 COMPUTING THE POPULATION VECTOR

To calculate population vectors for a set of motor cortical neurons, it is first necessary to determine the preferred direction vector of each neuron in the set. Second, one needs to measure the response of each neuron to movement in the direction for which the population vector is to be computed. The population vector can then be obtained by weighting each preferred direction vector with the respective neuron's activity and summing the weighted vectors (Georgopoulos et al., 1986).

More precisely, the weight for the *i*th neuron in the set, $w_i(M)$, is calculated by

$$w_i(M) = d_i(M) - b_i$$

where $d_i(M)$ is the spike rate of the *i*th neuron in response to movement direction *M*, and b_i is the neuron's baseline spike rate (a constant). Thus, only activity above or below the baseline is taken into account.

Next, the vectorial contribution of each neuron, $N_i(M)$, is obtained by multiplying the neuron's preferred direction vector C_i by the corresponding weight:

$$N_i(M) = w_i(M)C_i$$

If a neuron's response to movement direction *M* was above its baseline rate, this vector points in the preferred direction of the neuron, whereas it points in the opposite direction if the response was below baseline. The length of the vector (i.e., how strongly a neuron contributes to the population vector) is scaled depending on the absolute strength of the response.

Finally, to obtain the population vector for a movement direction *M*, *P*(*M*), the vectorial contributions of all neurons are summed:

$$P(M) = \sum_i N_i(M) .$$

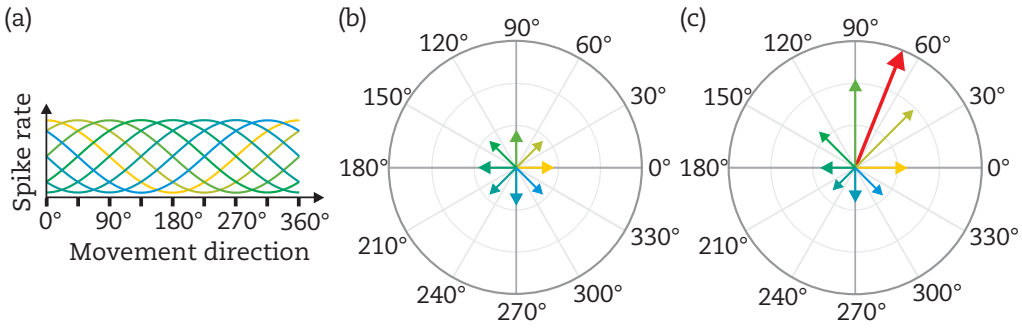

FIGURE 3.4: Schematic illustration of the population vector method. (*a*) Idealized tuning curves over movement direction (reduced to two-dimensional reaching space for simplicity) of eight motor cortex neurons. (*b*) Vector representation of the preferred directions of the eight neurons (arrow color corresponds to curve color in panel a). Note that the preferred direction vectors are normalized to equal length. (*c*) The same vectors, but individually weighted by the respective neuron's spike rate during a reaching movement into the angular direction of about 70°. Each weighted vector represents the respective neuron's contribution to the population vector (large red arrow).

visual stimuli competing for attention, motor acts competing for execution, or multiple items being retained in working memory. Moreover, neural populations are highly sensitive to weak input and respond faster to weak inputs than single neurons (Tchumatchenko, Malyshev, Wolf, & Volgushev, 2011). This again is a property related to noise: The membrane potential of neurons tends to fluctuate randomly, so that a given input might sometimes drive the neuron to threshold (when it happens to be close to threshold), but sometimes might fail to do so (when it happens to be far from threshold). Recalling that in tuned populations a given input potentially impacts many neurons, it is clear that at least some of these neurons will quite probably be in the right state when the input arrives. This is analogous to an array of low-light sensors where each one individually has a low probability of detecting a burglar, whereas with the whole array in place, the burglar will be detected almost certainly.

Taken together, the findings illustrated here argue for the importance of population-based representations in the nervous system. The peak-like structure of the activation distributions hints at parallels with DFs. The next sections elaborate further on this link and complete the grounding of DFT in neurophysiology, by looking more closely at the structure of population activation and how it maps to DFs.

DERIVING CONTINUOUS ACTIVATION DISTRIBUTIONS FROM NEURAL RESPONSES

Motivation for the DPA Approach

The population code representations in the brain form the biological basis for DFs. We contend that this level of analysis—neural population representations—is also the most appropriate level at which to establish formal links between brain and behavior. Dynamic field theory provides a framework that makes this link functional. However, as it stands, there is still a significant gap between biological neural populations and DFs. The formats of representation are fundamentally different. On the one hand, we have a collection of spiking neurons, while on the other hand, there is a distribution of activation, continuous over space and with continuous activation values. This discrepancy makes it difficult to directly compare the DF model with neural data obtained from experiments, or to make any concrete predictions about neural activity patterns from the model.

The first steps to bridge this gap have already been described for the computation of the population vector. The discrete spiking events of biological neurons can be converted into a firing rate to obtain a continuous activation variable. And by interpreting the activity of individual neurons as standing for certain metric feature values, a step is taken toward a representation over feature space. What is still missing here is the transition from a set of discrete values to the continuous activation distributions that form the basis for dynamic field theory. Intuitively, this step from the distributed representations in population codes to actual activation distributions may appear straightforward. However, a mathematically consistent formulation of this transition is not trivial. In the following sections, we will describe a formal method that constructs continuous distributions of population activation (DPA) from experimentally measured neural response properties.

To explore this approach and contrast it with other methods, let us look again at the population vector calculation of Georgopoulos and colleagues (Georgopoulos, Schwartz, & Kettner, 1986), which constitutes one standard approach to analyzing what is encoded in a neural population. In the initial study, the aim of this approach was to estimate the direction of a planned reach movement from the recordings of many motor neurons with different tuning curves which collectively form a population representation of a reach plan. In the population vector calculation, each neuron "stands" for its preferred movement direction. To estimate the movement vector encoded at a certain time by the whole population, these preferred movement directions are weighted with the firing rate of the corresponding neurons, and the average of these weighted direction values is determined. The population vector is a powerful tool for analyzing population activity and has been used successfully under many different experimental conditions to estimate what is encoded by an ensemble of neurons. However, as we shall see, a lot of relevant information is lost when the full distribution of activity over the population is reduced to a single mean value in the computation of the population vector.

The first aspect lost in the reduction to a population vector is the width and shape of the distribution of activation. A movement plan with a particular reach direction, for instance, may be encoded either by a small group of neurons that

have strongly overlapping tuning curves and are all strongly activated or, alternatively, by a larger ensemble of neurons that are only moderately activated and whose tuning curves are distributed over a larger range of movement directions. These different distributions may yield the exact same population vector. However, one of the studies discussed later here (Bastian, Schöner, & Riehle, 2003) found significant correlations between the concentration of activation for a certain movement direction and the time of movement initiation. This strongly indicates that the shape of activation distributions matters for the generation of behavior, and not just where the population vector points. To understand how overt behavior arises from neural processes, we must also capture these details of activation distributions in our models.

The second aspect that is lost when calculating the population vector is multimodal distributions of activity: A neural population can, in general, represent multiple values—such as different movement directions—at the same time. An instance of this has been described by Cisek and Kalaska (2005). Monkeys were presented with two potential reach targets, located in opposite directions from their initial hand positions—for instance, at directions of 90° and 270°. A color signal shown at the end of a delay period indicated which of them would yield a reward when reached toward. During this delay period, a bimodal distribution of activity was found in the investigated neural population in the pre-motor cortex. There was one group of active neurons whose tuning curves overlapped with the 90° direction, so their activity reflected the location of one possible target. A second group of active neurons within the population, with tuning curves covering the reach direction of 270°, reflected the location of the second potential target. When a single population vector is calculated for such a representation, it averages over the prepared movement directions and yields a misleading estimate of the encoded value. If two opposite directions are encoded in the neural population, such as in this example, they may cancel each other out in calculation of the average. The resulting direction of the vector will then be determined by small asymmetries in the activity distribution and be largely random. Alternatively, if two different, non-opposite directions are encoded, the population vector will indicate a direction in the middle between these two, which is not actually supported by the population activity.

In the next sections we present a method for analysis of neural population representations that aims to preserve the full activity distribution. In this approach, a DPA over a feature space is constructed from the tuning curves of neurons. The method can be applied to investigate the shape of unimodal activity distributions and their evolution over time, and likewise deal with multimodal distributions that appear if multiple values are encoded in a population. Beyond its use in analyzing and interpreting neural data, the DPA provides a direct link to DF models. We will describe the derivation of the DPA and its application in the analysis of neural activity patterns for two exemplary cases: the representation of visual stimuli in the primary visual cortex (Jancke et al., 1999) and planning of reach movements with incomplete prior information (Bastian, Riehle, Erlhagen, & Schöner, 1998; Bastian et al., 2003). For both cases, we will show DF models that can reproduce the experimentally observed activation patterns and explain how their shapes come about.

Construction of DPAs from Gaussian Tuning Curves

Jancke and colleagues (1999) recorded activity from neurons in the primary visual cortex of cats and used the DPA method to investigate the effects of neural interactions on early visual representations. To this end, activity distributions in response to single visual stimuli at different retinal locations were compared to the activity evoked by two stimuli presented simultaneously. First we will describe the application of the DPA method for single visual stimuli and then, in a later section, return to this study to discuss further results.

The first step in the construction of the DPA is to estimate the tuning curves of the neurons under investigation. Jancke and colleagues only considered the spatial tuning of the neurons, ignoring other visual features like orientation and spatial frequency that are also reflected in the activity of visual cortex neurons. Thus, the tuning curves measured experimentally corresponded to the spatial receptive fields of visually responsive neurons. Neural recordings were performed extracellularly in the foveal part of area 17 of anesthetized cats while visual stimuli were presented on a screen to the contralateral eye. Receptive fields were determined for a total of 178 cells and data were recorded for different stimulus conditions. Note that these 178 cells represent only a small

sample from the complete neural population in that cortical area, but they were sufficient to provide an estimate of the population activity as a whole. The receptive field center of every neuron was first estimated manually by stimulation with a light point and simultaneous observation of the neuron's firing rate. The resulting rough estimate of the neurons' receptive field center was then used as the basis for a more precise assessment, illustrated in Figure 3.5. A 6×6 grid of stimulus positions was placed over the estimated receptive field center, and the neuron's response was recorded while a small disk of light was briefly flashed at each grid location. The response profile obtained in this way was smoothed by a convolution with a Gaussian function, and a more precise estimate of the receptive field center was determined by calculating the center of mass of the smoothed profile.

The tuning curve of each neuron was then approximated by a Gaussian function of fixed width (reflecting the approximate average receptive field width), centered over the cell's receptive field center. A comparable procedure was also used by Cisek and Kalaska (2005), in their work on movement preparation mentioned earlier. We would note, however, that information on the exact shape of the neuron's receptive field is lost in this step. This is a compromise taken due to signal noise and a limited number of measurements on each cell: While it would be desirable to take into account the exact shape of each neuron's receptive field, the approximation by a Gaussian function of uniform shape provides greater robustness of the estimation. A slightly different approach that uses the full measured tuning curves for each neuron is presented in the second exemplary study later in this chapter, and an alternative method for constructing the DPA that avoids this problem is described later in Box 3.4.

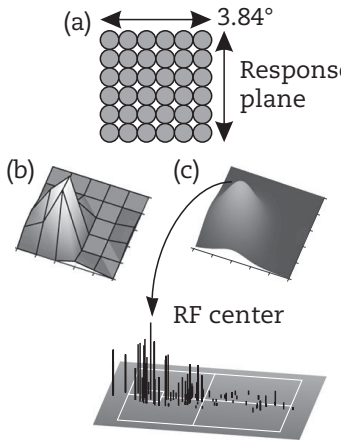

FIGURE 3.5: Determining visual tuning curves of neurons. A grid of 6×6 stimulus locations (*a*) was used to measure the receptive field of each neuron. It was centered on a coarse estimate of the receptive field obtained with the response plane technique. The profile constructed from responses to the grid stimuli (*b*) was smoothed with a Gaussian filter (*c*). The center of mass of this smoothed profile is then used as location of the neuron's tuning curve, modeled by a Gaussian function of fixed width. The firing rate of each neuron (indicated in the figure by bars of different lengths, located at the tuning curve center) is associated with this tuning curve for the construction of the DPA. Adapted from Jancke et al., 1999.

A DPA can now be constructed from the tuning curves for any stimulus condition and any time period of the stimulus presentation for which the neural responses have been recorded. To this end, the average firing rate of each neuron for the selected condition is determined and normalized to a fixed range. The tuning curve for each neuron is then weighted with the neuron's normalized firing rate, and the weighted tuning curves are summed. A schematic illustration of this process for one-dimensional tuning curves is shown in Figure 3.6. The unweighted tuning curves of four neurons are shown in different shades of green. These are then scaled with the neurons' firing rates (indicated by the length of the vertical black bar centered on each curve) to obtain the weighted tuning curves (blue). Finally, all of these weighted curves are summed to obtain the DPA, shown in red. Box 3.2 provides a formal mathematical description of the complete method.

Since each of the estimated tuning curves for the visual cortex neurons is a Gaussian function defined over the two-dimensional visual space, the obtained sum is likewise a distribution over visual space. This distribution yields an activation value for each position, even if no specific neuron has its receptive field center at that position. The activation value reflects how many tuning curves overlap at this point and how strongly the corresponding neurons are activated. It thereby provides a measure of how strongly the population activity supports the notion that a stimulus is present at that location. The sum of the Gaussian curves generally yields a smooth activation distribution in which regions of high activation result from the combined contributions of multiple

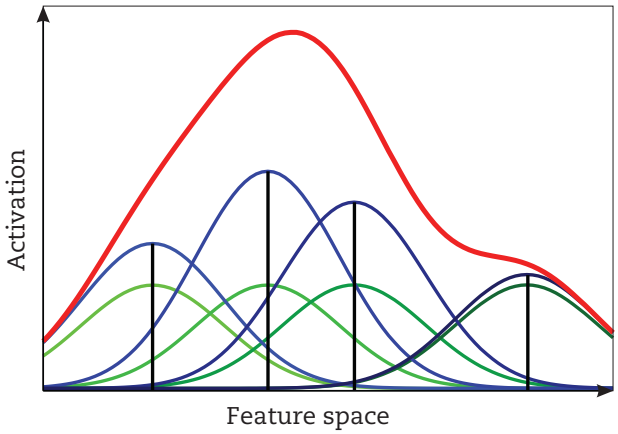

FIGURE 3.6: Schematic illustration of construction of a DPA from neural tuning curves. The normalized tuning curves of individual neurons (different shades of green) are plotted over the feature space under consideration (retinal position or reach direction in the examples treated here). These tuning curves are weighted with the neural firing rate from one experimental condition (black bars centered on the curves, with blue curves showing result of the weighting). The DPA is then computed as the sum of the weighted tuning curves (red). Additional normalization steps are often applied to the DPA to compensate for uneven distribution of tuning curves over the feature space (in this example, tuning curves lying more densely in the center of the depicted space).

BOX 3.2 CONSTRUCTION OF A DPA FROM GAUSSIAN TUNING CURVES

In the work of Jancke et al. (1999), the distribution of population activation (DPA) for visual representations is constructed from idealized Gaussian tuning curves. For each neuron i, the center of its receptive field, $\boldsymbol{m}_i = \left[m_{x,i}, m_{y,i}\right]$, in the two-dimensional visual space is estimated. The neuron's tuning curve f_i over the two-dimensional visual space is approximated by a Gaussian function with fixed width σ:

$$f_i(x,y) = \exp\left(-\frac{\left(x - m_{x,i}\right)^2 + \left(y - m_{y,i}\right)^2}{2\sigma^2}\right)$$

To construct the DPA for a certain stimulus condition a and time interval t, the tuning curve of each neuron is weighted with the neuron's firing rate for that condition and time period. The raw firing rate, $\tilde{r}_i(a,t)$, is first normalized by subtracting the baseline activity b_i and scaling it depending on the maximum firing rate m_i,

$$r_i(a,t) = \frac{\tilde{r}_i(a,t) - b_i}{m_i - b_i}$$

This yields a normalized firing rate, $r_i(a,t)$, that is always in the range [0,1]. A non-normalized activation distribution $\tilde{u}$ is obtained as the sum of the weighted tuning curves:

$$\tilde{u}(x,y) = \sum_i r_i(a,t) f_i(x,y)$$

To obtain the DPA u, the distribution $\tilde{u}$ is again normalized by dividing it by the unweighted sum of all tuning curves (to account for non-uniform sampling of the visual space by the selected neurons):

$$u(x,y) = \tilde{u}(x,y) \Big/ \sum_i f_i(x,y) = \sum_i r_i(a,t) f_i(x,y) \Big/ \sum_i f_i(x,y)$$

activated neurons with overlapping tuning curves, instead of forming only at the receptive field centers of individual neurons. This is shown in Figure 3.7. Figure 3.7b depicts the overlapping receptive field outlines for a small sample of neurons, overlaid over the stimulus display. The resulting smooth DPA (computed from all measured neurons) during the presentation of a single stimulus can be seen in Figures 3.7c and d.

To obtain the final DPA, an additional normalization step is necessary. The neural data stem from a random (and quite limited) sample of neurons from a large population, and one cannot generally assume that the tuning curves of these neurons are distributed equally over the visual space. We may, for instance, have one cluster of neurons in the sample with strongly overlapping spatial tunings, such that the corresponding region in visual space is overrepresented. Other regions, by contrast, may be covered only sparsely by recorded neurons. An example of this is also visible in the schematic in Figure 3.6, where the space in the central part of the plot is sampled more densely by neurons' tuning curves. Such uneven sampling can create strong biases in the computed DPA. If we sum the weighted tuning curves of all neurons, those regions in feature space that are covered by a large number of tuning curves will always tend to produce a high activation value, even if the activity of each individual neuron is relatively low. In contrast, more sparsely sampled regions can never reach very high activation values, even if the individual neurons show strong activity, because very few tuning curves contribute to these activation values. If we assume that the population as a whole represents visual space uniformly, we should compensate for such biases. This is achieved in the study of Jancke and colleagues by dividing the weighted sum of tuning curves by the unweighted sum of all tuning curves. This normalizes the DPA by scaling the activation up or down according to the density of the sampling at each point in visual space.

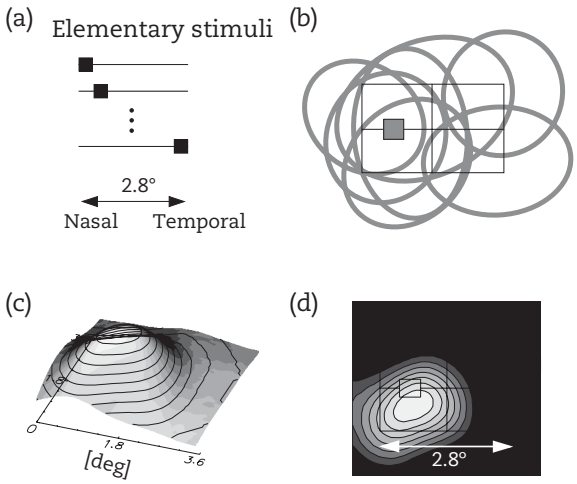

FIGURE 3.7: Stimulus conditions and DPA construction in Jancke et al. (1999). (*a*) Elementary stimuli (0.4° × 0.4° squares of light) were presented at seven horizontally shifted positions in the foveal part of the visual field. (*b*) Receptive field profiles of neurons (gray circles) overlapping and covering the analyzed portion of visual space (black box; gray square illustrates one elementary stimulus). (*c*) DPA constructed as weighted sum of tuning curves. (*d*) DPA derived for one elementary stimulus location overlaid with the stimulus position (small square). Adapted from Jancke et al., 1999.

We would note that even with this normalization, the results will not be meaningful if the number of recorded neurons used in the construction of the DPA is too small. In this case, some regions may not be sampled at all by the neurons' tuning curves. Even though the DPA construction will always yield some activation value for every point in the feature space, these values will not be informative for regions not sufficiently sampled by the recorded neurons. A very small sample size also increases the effects that random noise in the firing rates of individual neurons as well as single neurons with an uncharacteristic response behavior have on the resulting activation distribution. Whether the sample of neurons is sufficient cannot be seen directly from an individual DPA—which will always be a smooth distribution of activation over the feature space—but we may judge it by comparing the DPAs produced for different stimulus conditions.

Let us now look at the results of the DPA construction that Jancke and colleagues obtained for their recordings from cat visual cortex. The elementary stimuli used in the experiment were small squares of light with an edge length of 0.4° of visual angle that were flashed for 25 ms at one of seven horizontally aligned, equidistant locations at intervals of 0.4° (Figure 3.7a). The DPA analysis was applied to the neural response evoked by these stimuli, using the neurons' average firing rates over the whole period that a stimulus was presented at each of the seven locations. For all stimuli, the constructed two-dimensional DPAs over visual space show a single, roughly circular

peak of activation (Figure 3.8). Moreover, the location of the peak center in all cases closely matches the retinal location of the stimulus. This confirms that the neural activity in the cortical area that was recorded from does indeed reflect stimulus location in a population code representation. It also confirms that the DPA method applied on the given sample of neurons is effective in reading out what is being represented by the neural population. Moreover, it supports the assumption in the DF model that properties of sensory stimuli are reflected through activation peaks in neural populations.

In a subsequent analysis, the authors constructed a time series of DPAs for each stimulus presentation. To this end, they determined average neural firing rates for brief time segments and computed a DPA for each of these. The authors found that the peak location remains largely constant as activity rises and falls in response to the flashed stimuli, although representation of stimulus position is less reliable in the latest phases of the response. Interestingly, the width of the activation peaks in the DPA (measured as standard deviation from the center in the normalized distribution) consistently increases over the duration of the neural response. This contradicts earlier findings (e.g., Orban, 1984), which posited that the initial broad representations formed by feed-forward inputs are sharpened over time as a result of recurrent interactions.

These initial results demonstrate the validity of the DPA method and its use in analyzing neural data. However, the main scientific question in this study was whether the activation distributions showed signatures of lateral interactions within neural populations in the visual cortex. The role of lateral interactions in shaping activation patterns is also a central issue in DFs. We will return to this question later in this chapter, where we will present additional empirical results from this study and show how they can be explained in a DF model. Before doing so, however, we will present the DPA construction for a second example from motor and premotor cortex, in order to show how this approach generalizes to cortical populations that do not have a topographical organization on the cortical surface.

Constructing DPAs for Movement Preparation

In the work of Bastian and colleagues (1998, 2003), the DPA method—with slight variations compared to the work of Jancke and colleagues—was used to investigate the formation of movement plans in the motor and premotor cortex of macaque monkeys. This example from a different domain shows the general nature of the DPA approach. For the experiment, monkeys were trained to perform an arm movement from a central location to one of six target locations arranged equidistantly around the center (Figure 3.9). The required reach direction on each trial was indicated by illuminating a red LED at the target location. A preparatory signal, which provided complete or partial information about the upcoming reach direction, was given 1 second before this definite reach cue. It consisted of green LEDs being illuminated at one, two, or three of the potential target locations. These pre-cued locations were always contiguous to each other and included the ultimate reach target. The goal of the experiment was to investigate how the preparatory activity for the reach movement changed with different levels of certainty in the provided preparatory signal.

The feature space over which the DPA was calculated was the direction of the arm movement. The firing rates of neurons in the motor and premotor cortex, described previously to represent movement direction in a population code

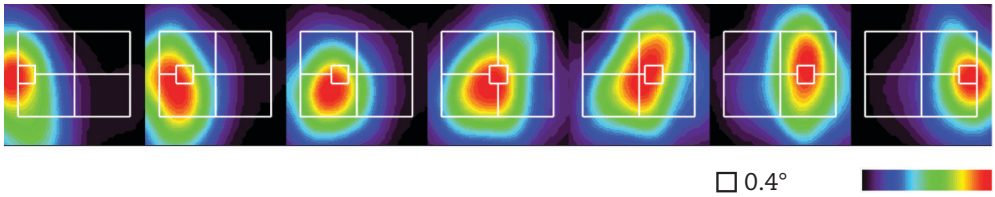

FIGURE 3.8: Two-dimensional DPAs constructed for the neural response to individual visual stimuli, presented at seven horizontally shifted locations. DPAs were computed from neural firing rates averaged over the period from 45 to 60 ms after stimulus onset. The activation level is shown on a color scale normalized to maximal activation separately for each stimulus (calibration bar at bottom right). Adapted from Jancke et al., 1999.

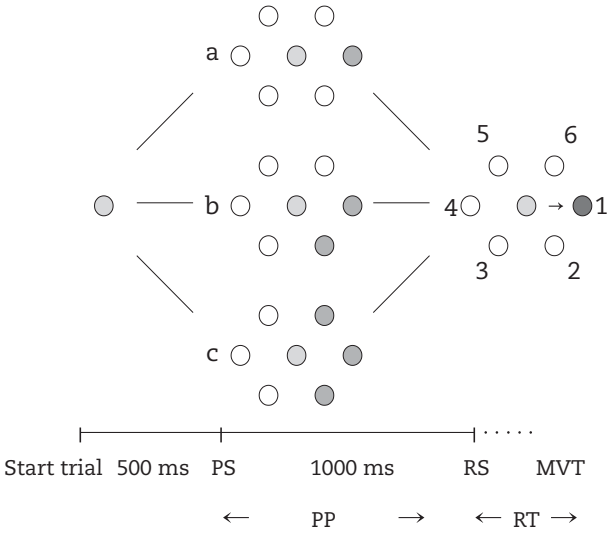

FIGURE 3.9: Reach task with pre-cues. Macaque monkeys were trained to make reach movements from a central manual fixation position to six possible target positions arranged on a circle around the fixation point. After the monkey held its hand on the central point (light gray circle, left), pre-cues were presented at one (*a*), two (*b*), or three (*c*) contiguous target locations (darker gray circles, middle). After an additional delay, a definite reach cue was shown at one of the pre-cued locations and the monkey had to execute a reach movement (dark gray circle, right). PS: preparatory signal, RS: response signal, MVT: movement onset, PP: preparatory period, RT: reaction time. From Bastian, Schöner, & Riehle, 2003.

(Georgopoulos, 1995), were measured by extracellular recording. Unlike in the first example, the tuning curves of the neurons were not estimated in a separate procedure but were instead determined directly from the neural responses in the main task. To this end, the reaction time period of the experiment was used as a reference condition. This was done on the basis of the assumption that during this time period—after the definite movement goal has been presented, until the start of the actual arm movement—an arm movement to the uniquely indicated target is prepared. By averaging over all trials (with different preparatory signals) and over the whole duration of this reaction time period, a single average firing rate is obtained for each of the six target directions.

These firing rates are assembled into a tuning curve (see Box 3.3 for a formal description). Each of the six reach directions in the experiment serves as a node or sampling point for the tuning curve over the space of movement directions, and the measured average firing rate for that direction yields the tuning curve value. These raw tuning curves are then normalized so that they range from 0 to 1. Note that in using this approach, the tuning curves do not all have a uniform shape, and individual properties like the width of a neuron's tuning are preserved.

Using these tuning curves, we can now again construct the DPA for any time period and any condition of the experiment from the measured firing rates of the neurons. The tuning curve of each neuron is weighted with the neuron's firing rate in the condition under consideration, and all tuning curves are summed. Again, another normalization step is necessary to account for the non-uniform sampling of the feature space by the neurons' tuning curves. If there is a higher density of tuning curves for one reach direction than for others, this will introduce a bias in the resulting DPA, even if each tuning curve itself is normalized. In a situation where each contributing tuning curve is assigned the same weight, the activation would still be higher for the more densely sampled region. Bastian and colleagues employed a subtractive normalization (or baselining) in which they determined a DPA for a baseline condition (before the presentation of any stimuli) and subtracted it from the DPAs in all other conditions.

Examples of DPAs constructed in this way are shown in Figure 3.10a–c. Note that these DPAs appear less smooth than those constructed from idealized Gaussian tuning curves for visual cortex neurons (e.g., Figure 3.8). This is because

BOX 3.3 CONSTRUCTION OF A DPA FROM FIRING RATES IN REFERENCE CONDITIONS

In the work of Bastian et al. (2003), the tuning curves over the space of reach directions are obtained directly from the neural firing rates in the reference conditions (reaction time phase of each trial). For each neuron i, the raw tuning curve $\tilde{f}_i$ is defined at the six possible reach directions $x_k, k \in \{1,\ldots,6\}$ as

$$\tilde{f}_i(x_k) = \left\langle r_i\left(x_k, t_{rtp}\right)\right\rangle$$

Here, $r_i\left(x_k, t_{rtp}\right)$ is the mean firing rate of neuron i during the reaction time period in a single trial with reach direction x_k, and $\langle . \rangle$ denotes the average over all trials. The tuning curves f_i for the construction of the distribution of population activation (DPA) are derived from these raw tuning curves by normalization to the interval [0,1].

The non-normalized DPA $\tilde{u}$ is then determined for any condition a and time interval t as weighted sum of the tuning curves:

$$\tilde{u}(x) = \sum_i r_i(a,t) f_i(x)$$

Here, $r_i(a,t)$ is the mean firing rate of neuron i for the given condition and time interval, averaged over trials. Note that the activation distribution is only defined at the original reach directions x_k used in the reference conditions; for other points along the space of possible reach directions an estimate can only be obtained by interpolation.

As a form of normalization (or, more precisely, baselining), another DPA is subtracted from this distribution, one that is computed from the neural firing rates in the same condition during a 200 ms time window t_{pre} before any stimuli are presented:

$$u(x) = \sum_i r_i(a,t) f_i(x) - \sum_i r_i\left(a, t_{pre}\right) f_i(x)$$

the neural tuning curves used here only specify the firing rates for six movement directions, corresponding to the six reference conditions in the experiment. No interpolation or function fitting was employed to estimate firing rates for intermediate movement directions. The resulting DPA then yields activation values only for these six directions, rather than providing a continuous distribution over the space of movement directions. In order to increase the spatial resolution of the DPA, we would have to increase the number of reference conditions. Adding more neurons, by contrast, would produce a more reliable estimate of the actual activity distribution in the whole population but would have no effect on the spatial resolution of the DPA.

While a DPA constructed directly from measured neural firing rates appears less smooth than one that is based on idealized Gaussian tuning curves, it can nonetheless provide a representation of the neural population activity. As in the previous example, it can form flat distributions in the absence of strong activity, localized peaks, or multimodal distributions. It is in fact a more accurate representation of population activity, since the individual shape of each neuron's tuning curve is preserved in the computation of the DPA.

It is informative to first look at the DPAs for the reference conditions themselves, that is, the reaction time periods for reaches to the six target locations. If the neural population sampled from does indeed provide a population code representation of movement direction, then the resulting DPAs should reflect the actual reach direction in those conditions. This was indeed the case: The constructed DPAs showed a single peak at or close to the reach direction for all target locations (averaged

over trials of all conditions). The same was true for the early and late preparatory period in the condition with a definite preparatory signal (only a single potential target illuminated). This indicates that the same neurons are also involved in the earlier planning stage of the movement and consistently reflect the planned reach direction throughout the trial.

Following this confirmation of the analysis method, Bastian and colleagues (2003) used DPAs to describe the differences in activation patterns under the different trial conditions. The evolution of the activation distribution for different pre-cue conditions is shown in Figure 3.10a–c. When two (Figure 3.10b) or three locations (Figure 3.10c) were indicated in the preparatory signal as potential reach targets, the peak in the DPA was located approximately at the center of these locations during the preparatory period. It then shifted toward the actual reach direction once the definite target cue was given. Furthermore, the width of the activation peak in the DPA during the preparatory period increased with the number of pre-cued locations: It was narrowest in the condition with complete target information (Figure 3.10a), wider in the condition with two potential targets (Figure 3.10b), and widest for three pre-cued target locations (Figure 3.10c). This indicates that the activity pattern in this neural population does not simply encode a single direction value. Instead, the full activity distribution contains information about additional aspects of the movement plan,

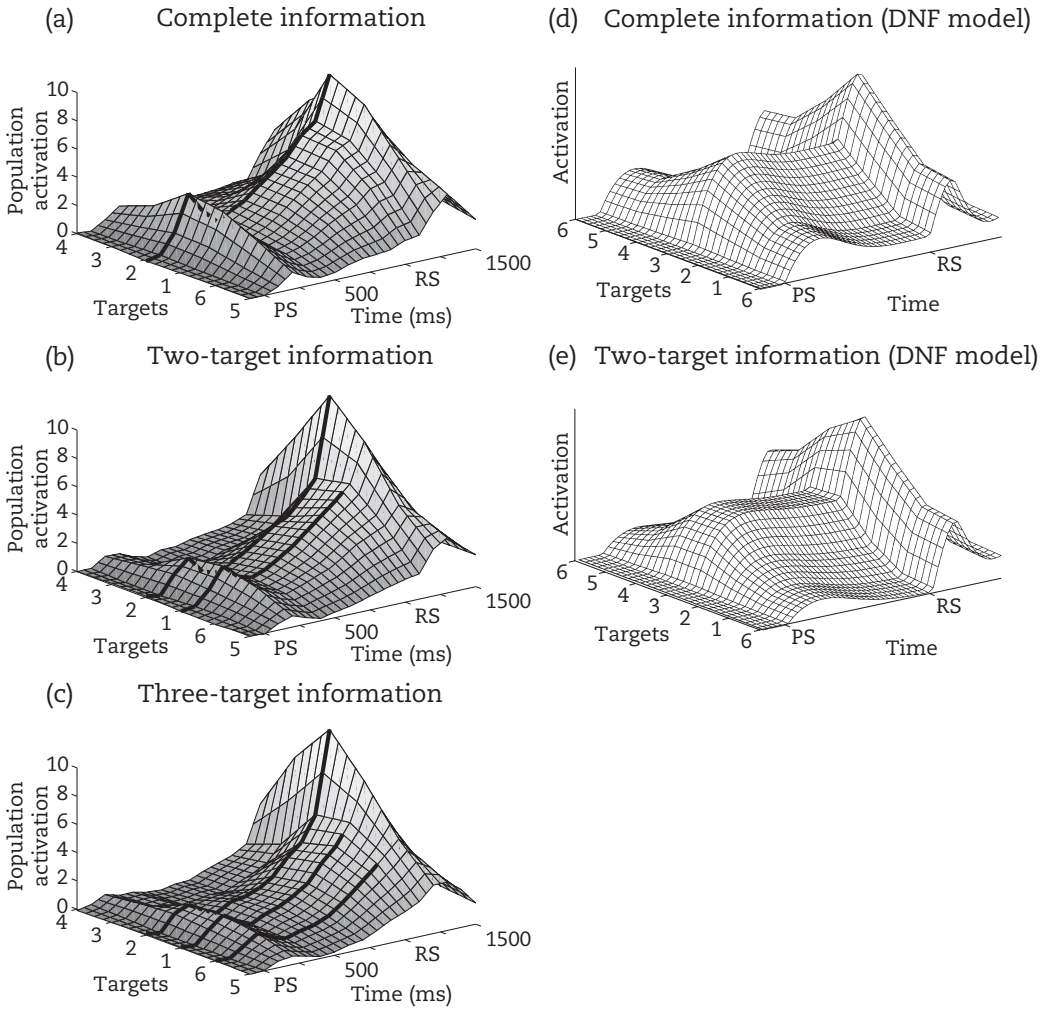

FIGURE 3.10: Temporal evolution of neural activity in monkey motor cortex during movement preparation analyzed with a DPA, and DF model fit. The plots on the left show the preparatory activation distribution over time for a reach movement given one (*a*), two (*b*), or three (*c*) pre-cued locations. On the right, the activation patterns in the excitatory layer of a two-layer DF model performing the same task are shown for one (*d*) and two (*e*) pre-cued locations. PS: preparatory signal onset, RS: response signal onset. Adapted from Bastian, Riehle, Erlhagen, and Schöner (1998), and Bastian, Schöner, and Riehle (2003).

such as the certainty of an upcoming movement in a specific direction.

The DPA analysis can in particular be used to describe the detailed time course of the evolution of activity patterns, which can also be seen in Figure 3.10. To this end, the duration of the trial is divided into short time segments. The neurons' firing rates are then determined for each segment individually, and a series of DPAs is constructed from these values. For the reaching experiment, this analysis showed an initial rise of activity in response to the preparatory signal, reaching a first maximum briefly after the cue onset. Activity then transiently decreased, but subsequently grew again over the course of the preparatory period and more quickly following the presentation of the definite reach cue. The concentration of the activity likewise increased after the reach cue, and both reached a maximum approximately 100 ms before movement initiation.

To assess the functional significance of the DPA time courses, Bastian and colleagues tested whether there was a correlation with reaction times. Trials were separated into two groups—reaction times higher than the median value and lower than the median value—and separate DPAs were constructed for the two groups. The total activity was found to be larger and rise earlier in the fast trials than in the slow trials. In addition, the concentration of activity was higher in the fast trials, especially toward the end of the preparatory period. These results establish a direct link between the shape of the DPAs and a behavioral variable—in this case, reaction times.

These results highlight, once again, that the *distribution of activation* is important, not just a mean value as used in the population vector approach. The shape of activation distribution for the preparatory activity in the motor cortex reflects the certainty of a movement plan and is functionally relevant for movement initiation. Moreover, this example shows that the DPA method can reliably create meaningful activation distributions even if the neurons recorded do not form a topographical map. In the visual cortex example discussed previously, the physical arrangement of the neurons in the cortex preserves the neighborhood relations of their spatial receptive fields, such that simply plotting their activity over the cortical surface often yields activation patterns that are comparable to the DPA results (Markounikau, Igel, Grinvald, & Jancke, 2010). In motor cortex, however, there is no such topographical map. Since the DPA method describes activation over the space of a perceptual or behavioral variable (i.e., movement direction), the results are independent of the physical arrangement of neurons on the cortical surface.

The DPA examples from visual and motor cortex demonstrate the utility of this approach for understanding neural population representations. In the next section, we ask how the DPA approach relates to DFT. In particular, we describe how DF models can be used to simulate results from the DPA approach in detail and how this sheds light on the neural dynamics that underlie activity of neural populations.

DYNAMICS OF ACTIVATION DISTRIBUTIONS IN NEURAL POPULATIONS AND DYNAMIC FIELDS

Signatures of Lateral Interactions in Primary Visual Cortex

The DPA study of movement preparation showed how different stimulus patterns shape the activation distribution in a neural population. A single pre-cue induces a relatively sharp activation peak, while multiple adjacent cues create a broader distribution of activation over the space of possible movement directions. But the stimuli alone cannot fully explain the activation time courses found during movement preparation. While there was an initial activation maximum during the presentation of pre-cues, activation did not fall back to its resting state after the visual cues were turned off. Instead, the general pattern of activation over the feature space was retained, and activation rose again over the period of movement preparation.

These observations indicate the presence of interactions within the neural populations. These interactions create, retain, and modulate activation patterns beyond what is directly induced by external stimulation. These interactions were discussed in the previous chapter as the source of cognitive processes in DF models. Interactions can produce detection decisions, selection decisions, and working memory, and thereby move the DF models beyond passive representations of the input. In this section, we discuss these neural interaction effects in the context of the DPA examples introduced earlier in the chapter.

We begin with the study of Jancke and colleagues (1999), which was designed to find empirical evidence of such interactions in neural populations of

cat primary visual cortex. The effects of interactions in this sensory cortical area can be expected to be merely modulatory in nature (since these areas are not assumed to be directly involved in selection decisions or working memory). Nonetheless, clear signatures of the types of interactions employed in DF models have been identified.

To identify interaction effects, Jancke and colleagues compared the responses to elementary stimuli to the activation patterns evoked by composite stimuli (Figure 3.11a). The elementary stimuli—which formed the reference conditions for the comparison—are the small squares of light described previously (Figure 3.7a). For the composite stimuli—the test conditions—two of these squares were presented simultaneously. One stimulus was always presented at the most nasally located position, while the other occupied one of the six remaining locations, yielding six different distances between the two stimuli. DPAs were constructed as

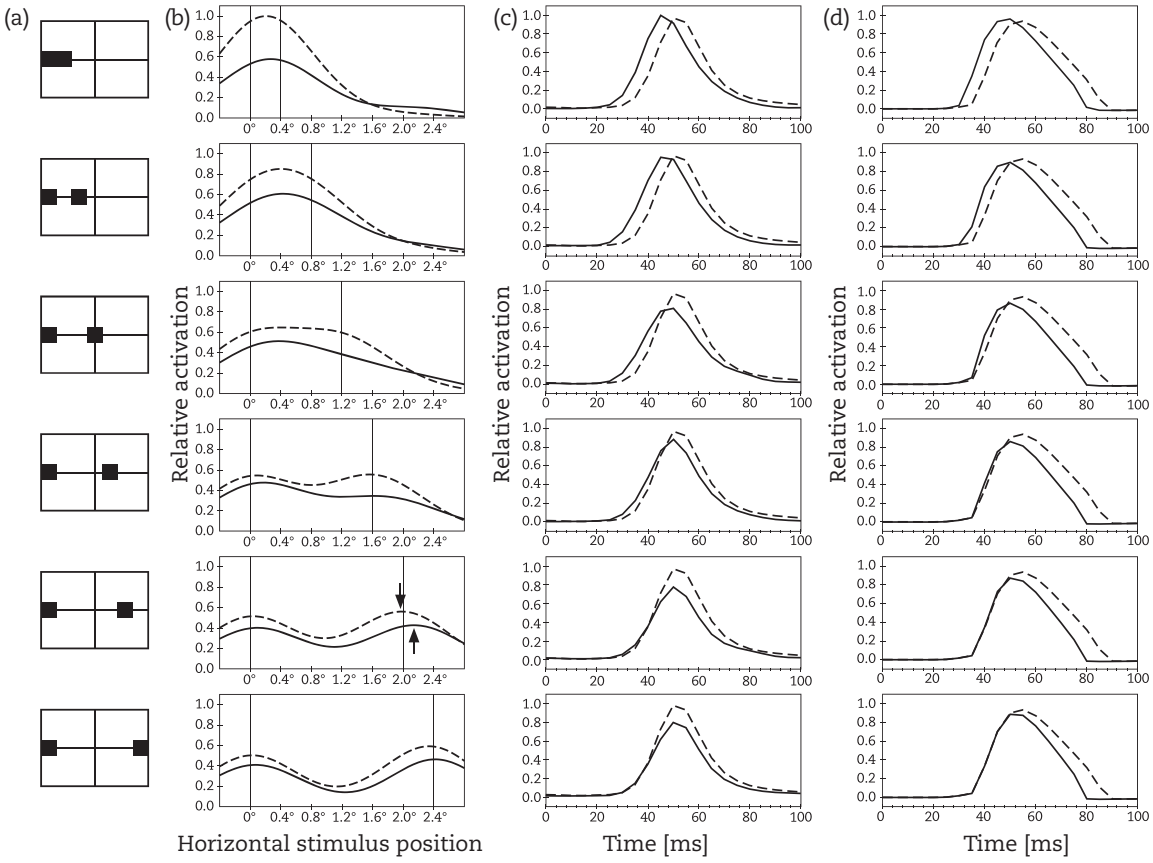

FIGURE 3.11: One-dimensional DPAs constructed from neural activity in cat visual cortex using the optimal linear estimator (OLE) method. (*a*) Composite stimulus patterns, with fixed location of one (nasally located) stimulus and six different stimulus distances. (*b*) DPAs constructed for the composite stimuli (solid lines) compared to the linear superposition of DPAs constructed for the elementary stimuli when presented separately (dashed lines). DPAs for the composite stimuli show consistently lower activation levels than the superposition. In addition, for larger stimulus distances, a repulsion of the two activation peaks from each other can be observed (highlighted by arrows for one peak). This effect is more pronounced in the later phase of the response (not shown). (*c*) Time course of total DPA activation in the region of the fixed nasally located stimulus for composite stimuli presentation (solid line) and presentation of this stimulus alone (dashed line). Total activation for each time was computed by integrating over the DPA in a 0.8° wide band around the stimulus position. For low-stimulus distance, the rise of activation starts earlier and higher activation values are reached for the composite stimulus presentation than for the elementary stimulus alone. This effect disappears for larger stimulus separations, and only a decrease of activation in the later phase of the response remains. (*d*) Activation time course in a DF model for the same stimulus conditions, scaled to maximal activation in each condition analogous to the DPA results. Adapted from Jancke et al., 1999.

described earlier, and additional analyses were performed with an alternative method for DPA construction, described in Box 3.4.

The rationale for comparing DPAs for elementary and composite stimuli in this study is the following: If there are no interactions between individual stimuli in the population code representation, it should be possible to fully predict the neural response to a pair of stimuli from the neural response that each of the stimuli evokes by itself. In the simplest case of a linear neural response behavior, the activity pattern evoked by two simultaneously presented stimuli should be the linear superposition (i.e., the point-wise sum) of the activity patterns evoked by the two stimuli individually. Deviations from the linear superposition indicate the presence of interactions, and their timing and dependence on stimulus locations can reveal temporal and spatial properties of the interactions. Note that this reasoning does not imply that the interactions only appear in the case of composite stimuli. They likewise affect neural activity patterns for the elementary stimuli, but only the comparison between the two conditions allows us to distinguish between pure feed-forward activation and interaction effects.

The authors first compared DPAs constructed from average firing rates over the whole time course of the stimulus presentation. They consistently found that activation in response to the composite stimulus was significantly weaker than what linear superposition of the two elementary stimuli would predict (see 3.11b and 3.12). This effect is especially

BOX 3.4 OLE METHOD FOR CONSTRUCTING DPAS

An alternative approach exists for the construction of DPAs, in which the tuning curves are not determined directly from measured neural firing rates but are derived through optimal linear estimation (OLE) from expected, idealized activation distributions for the reference conditions. This approach was adapted from an analogous method for the computation of population vectors (Salinas & Abbott, 1994). It has been applied to analyze both the visual representations in cat primary visual cortex (Jancke et al., 1999) and the movement preparation in macaque motor cortex (Erlhagen, Bastian, Jancke, Riehle, & Schöner, 1999), and the results have been consistent with those of the direct method. We will describe it here for the latter application.

The central assumption for the OLE approach is that in the reference conditions the DPA should represent a certain feature value (like reach direction) in a fixed functional form, called the target DPA. We use an exponential of the cosine function, centered on the target direction x_k of the reach movement, as target DPA:

$$\hat{u}_k(x) = A \cdot \exp\left[\sigma \cos(x - x_k) - 1\right] - B$$

Here, σ is a width parameter for the activation peak in the target distributions (approximately 45°), and the parameters A and B are chosen such that the activation values range approximately from 0 to 1.

Now, we effectively ask: What does each tuning curve have to look like such that the sum of all tuning curves, weighted with the neural firing rates for each reference condition, yields the desired target DPA in every reference condition? To do this, we first choose a discrete sampling of the feature space for the DPA (which is independent of the number of neurons or reference conditions). The goal is then to find a tuning curve f_i for each neuron i such that for each reference condition and at every sampling point x_l, the weighted sum of all tuning curves approximates the target DPA $\hat{u}$:

$$\hat{u}_k(x_l) \approx \sum_i r_i(x_k, t_{rtp}) f_i(x_l)$$

As before, the reaction time periods t_{rtp} of all trials are used as reference conditions, and accordingly, the average firing rate of each neuron for the reach direction x_k during this time period, $r_i(x_k, t_{rtp})$, is used as weight for the tuning curve.

With this goal, we can formulate a concrete optimization problem. We want to find a set of tuning curves f_i that minimizes the mean quadratic error E, which measures the deviation of the weighted sum of tuning curves from the target DPA:

$$E = \frac{1}{n_k n_l} \sum_k \sum_l \left(\hat{u}_k(x_l) - \sum_i r_i(x_k, t_{rtp}) f_i(x_l) \right)^2$$

Here, n_k is the number of conditions (the six reach directions), and n_l is the number of sampling points x_l (36 for this example). This optimization problem can be solved with standard mathematical methods and yields tuning curves for each neuron without requiring any previous knowledge about their properties. The DPA for any other condition a and time period t can then be computed from the tuning curves in the same way as in the direct method:

$$u(x) = \sum_i r_i(a,t) f_i(x)$$

An interesting property of this method is that the derived tuning curves for neurons are not normalized and may vary in shape. That means that some neurons may contribute more strongly to the DPAs, while others may be nearly ignored. This is often appropriate, since neurons even from the same cortical area do not necessarily contribute equally to represent a certain feature value in a population code. Furthermore, the final normalization of the DPA used in the direct method can be omitted, since the optimization implicitly adjusts the strengths of the tuning curves to compensate for sampling effects.

Although the OLE method will find a set of tuning curves of any desired resolution and for any target distributions, the quality of the fit and the significance of the result will depend on the available data. If only a small number of neurons was recorded, or the response properties of the neurons show little variance between each other, the resulting fit of the target DPAs will likely be poor. On the other hand, if only few reference conditions are used, it becomes easier to achieve a satisfactory fit, but the DPAs obtained from these tuning curves may not reliably reflect neural activation patterns under test conditions.

pronounced for small stimulus separations, but it is still apparent in the largest stimulus distance of 2.4°. At this distance, there is little overlap between activity distributions for the two elementary stimuli, and the DPA for the composite stimulus shows a pronounced bimodal pattern. The levels of activation at the two stimulus locations in the composite DPA are even lower than the activation levels observed for each elementary stimulus alone, which rules out the possibility that the observed reduction is an effect of saturation. This indicates that there are pronounced inhibitory interactions that shape the activity distribution in the visual cortex.

To estimate the temporal properties of the interactions, the authors constructed DPAs for smaller time windows and analyzed changes over the time course of the stimulus presentation (Figure 3.11c). The analysis focused on the emergence of activation at the location of the most nasally presented stimulus, which was shared between all composite stimuli. Despite the overall pattern of reduced activation described previously, they found that during the early part of the response there was evidence for excitatory interactions in the composite stimuli. When two stimuli were simultaneously presented in close proximity, the activation level for the nasally positioned stimulus was not only increased compared to the single elementary stimulus but was even higher than that predicted from the superposition of the two elementary stimuli. Compared to the single-stimulus presentation, the activation increased and reached its maximum earlier, but then also decreased faster and was lower during the late phase of the response. For larger distances between stimuli, the signs of early excitatory interactions disappeared, and there was only an overall suppression of the activation.

Finally, Jancke and colleagues found a spatial signature of interactions in the representation of visual stimuli. For larger stimulus distances (1.6°

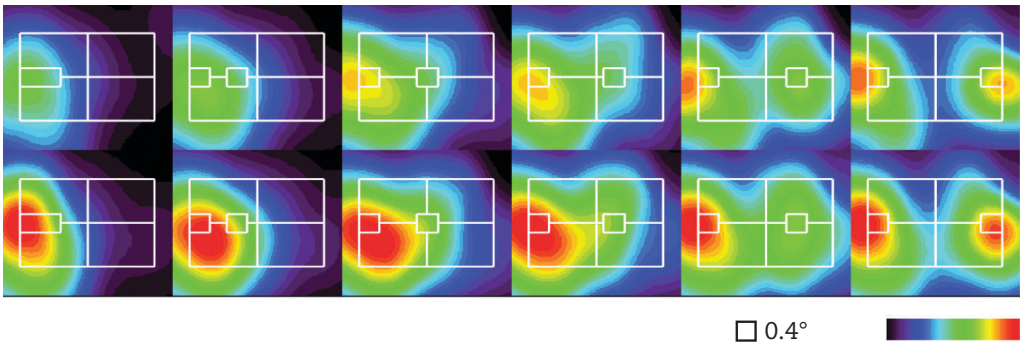

FIGURE 3.12: Two-dimensional DPAs computed from neural firing rates for the composite stimuli in the test conditions (*top*), compared to the linear superposition of DPAs derived for the two elementary stimuli of each condition (*bottom*). Neural firing rates were averaged over the time interval from 30 to 80 ms after stimulus onset for these DPAs, and activation values were normalized to the maximum activation in the superposition for the plots (calibration bar at the bottom right). For all composite stimuli, a significant reduction of neural activation in response to the composite stimuli as compared to the superposition of the elementary stimuli was observed. Adapted from Jancke et al., 1999.

and greater), the DPA for the composite stimuli showed a bimodal pattern, with two activation peaks located approximately over the stimulus locations (3.11b and 3.12). However, when the exact positions of these peaks were compared to those that appear in the superposition of DPAs for the elementary stimuli, a systematic deviation was observed: Especially during the late phase of the response, the activation peaks shift outward and away from each other, with an increase in the distance between them of up to 0.3° (highlighted by arrows for one peak in Figure 3.11b, although the effect is less pronounced in the this plot since it averages over the whole response time period).

Modeling Interaction Effects with Dynamic Fields

The experimentally observed differences between the activation distributions for elementary and composite stimuli in the study by Jancke and colleagues can be explained by patterns of lateral interaction that act on the activation distribution. From neurophysiological research we know that neurons that form a population do not just passively reflect the input they receive and convey it to the next area in the cortical processing hierarchy; these neurons also act on each other by means of synaptic connections. These connections are called *lateral connections* because they target the same population from which they originate and thus do not link different levels in the cortical processing hierarchy. In primary visual cortex, excitatory lateral interactions have been identified between orientation selective cells. These connections link primarily neurons with similar preferred orientations (that are close to each other in feature space), and interaction strength declines with increasing disparity between preferred orientations (Ts'o, Gilbert, & Wiesel, 1986). Both excitatory and inhibitory lateral interactions have been described in the motor cortex. Interactions between cells encoding reach movements with similar directions are excitatory, whereas cells that code for dissimilar directions are coupled inhibitorily (Georgopoulos, Taira, & Lukashin, 1993).

This pattern of lateral interactions—mutual excitation over short distances in feature space, and mutual inhibition over longer distances—is the same that is typically used in DFs to promote the formation of stabilized local activation peaks. Jancke and colleagues set up a DF model to demonstrate that interaction effects of this type can indeed account for their experimental observations, and to estimate the quantitative properties of the interactions necessary for this. The model was fit to match the activation time course in the DPA (Figure 3.11d). The overall suppression in the case of composite stimuli can be reproduced and explained in the model by the presence of inhibitory interactions through which strong activation at one location in a field can decrease activation levels elsewhere along feature space. The observation that suppression effects occur even for the largest stimulus distances in the experiment is consistent with the assumption that these inhibitory interactions act over long ranges. The early increase of activation for the composite stimuli relative to the superposition case is reproduced through the lateral excitatory interactions in the DF. This increase of activation was only observed for small stimulus

distances, which fits the assumption in the model that lateral excitation is limited to a smaller range with respect to feature space.

The experiment also showed differences in time course between these interaction effects. The signatures of excitatory interactions appeared earlier but were no longer apparent later during stimulus presentation, indicating a pattern of early excitation and later inhibition in the population response. The DF model as discussed so far cannot account for this, but Jancke and colleagues employed an extension of the basic model that is described in detail in the next section. This extension separates the field into an excitatory and an inhibitory layer to reflect that inhibitory interactions in biological neural populations are conveyed by inhibitory interneurons. This modification creates a behavior of the model in which inhibition appears with a delayed onset but then cancels out the effects of lateral excitation, thus reproducing the experimental observations.

Finally, through the combination of excitatory and inhibitory interactions, the model can also account for the observed repulsion effect between activation peaks in the DPA. One further assumption is required here, namely, that the strength of lateral inhibition decreases at longer distances. In the DF model, this is typically realized by using a broad Gaussian function as the inhibitory interaction kernel. If now two activation peaks exist within moderate distance from each other, the inhibition is particularly strong in the region between them. This region is relatively close to both peaks and therefore receives strong inhibition from both of them. The region on the opposite side of each peak only receives strong inhibition from one active region. As a result, since each peak receives more inhibition on one side than on the other, the two peaks drift slightly apart. The repulsion can be especially pronounced if the inhibition is combined with short-range excitation, which acts to keep the size of each peak stable while still allowing shifts in position. We will return to this effect and explicate it in greater detail in Chapter 6, where we discuss behavioral results from humans in a visual working memory task.

The results discussed in this section demonstrate how DF models can be employed to explain experimental observations at the level of population activity. Moreover, they provide empirical support for the biological plausibility of the typical interaction patterns used in behavioral and robotic DF models. In continuation of this work, a more quantitative investigation of interactions in the primary visual cortex using a DF model was presented by Markounikau and colleagues (2010), based on neural data obtained through voltage-sensitive dye imaging.

Two-Layer Dynamic Fields

The extension of the basic DF model used to capture interaction patterns in the work of Jancke and colleagues is the two-layer field. In this form of DF model, separate layers are used to describe activation of excitatory and inhibitory subpopulations. This extension reflects more closely the properties of biological neurons and is often useful to capture detailed activation time courses of real neural data.

The neurophysiological motivation for two-layer fields is a basic property of biological neurons, described by Dale's law. *Dale's law*, in a modern formulation (Eccles, 1976), states that neurons emit the same set of neurotransmitters at all their synapses. This has been found to be true with very few exceptions. Dale's law in particular implies that the effect that the firing of one neuron has on postsynaptic neurons can be either excitatory or inhibitory, but not both. In neural populations in the cortex, excitatory neurons, like pyramidal cells, can have long-ranging axons and are responsible for conveying information between cortical areas. Excitatory interactions can be conveyed by direct synaptic connections between these excitatory neurons. Inhibitory neurons typically project more locally and convey indirect inhibitory interactions between the excitatory neurons. For instance, a pyramidal cell may have synaptic projections to a group of inhibitory interneurons and excite them. The activated interneurons then project to other pyramidal cells and inhibit their activity.

This connectivity has some consequences for the activation time course in neural populations. In particular, it introduces a delay for inhibitory interactions. When an external stimulus arrives, it can directly activate the excitatory neurons. The inhibition that limits the growth of activation and mediates competition within the population only appears after the inhibitory neurons have been sufficiently activated to start firing. They may be excited either directly by the external stimulus or by the excitatory neurons within the population itself. In the latter case—which we assume in the DF model—an additional delay is created, since the inhibitory neurons only receive input after the

excitatory ones have started firing. The delayed onset of inhibition means that an external stimulus may produce an initial overshoot of excitation, which then decreases as it is balanced by rising inhibition. This gives rise to a phasic-tonic response behavior in the excitatory neurons (although it is not the only cause of this pattern).

In the DF model, this connectivity and the resulting effects on the activation time course can be replicated by introducing separate layers for the excitatory and inhibitory subpopulations (Figure 3.13; see Box 3.5 for the formal description). The basic structure for the two-layer field is as follows: The two layers, excitatory and inhibitory, are defined over the same feature space and are both governed by differential equations similar to those used in one-layer DFs. In the version considered here, only the excitatory layer receives direct external input. Excitatory interactions are implemented through connections of the excitatory layer onto itself, described by an interaction kernel (e.g., a Gaussian function). In addition, the excitatory layer also projects to and excites the inhibitory layer. These projections are topological; that is, a projection from any point along the feature space on the excitatory layer acts most strongly onto the same point in feature space on the inhibitory layer. The inhibitory layer, in turn, projects back to the excitatory layer in an inhibitory fashion (that is, it creates a negative input in that layer's field equation). Within the inhibitory layer, there are typically no lateral interactions.

The projections between the two layers can be described by interaction kernels, just like the lateral interactions. Note that the effective spread of inhibition is determined by properties of both the projection from the excitatory to the inhibitory layer and of the reverse projection. Let us assume, for instance, that all three projections in the two-layer field (from excitatory to excitatory, excitatory to inhibitory, and inhibitory to excitatory) are described by Gaussian kernels of the same width. Then the effective range of inhibition in the excitatory layer will be wider than the range of lateral excitation, because the inhibition is spread by two kernels instead of just one. In practice, the two-layer field is sometimes set up in such a way that the projection from the excitatory to the inhibitory field is purely local (point-to-point, without an interaction kernel). The kernel for the reverse projection is then made wider to produce the overall pattern of local excitation and surround inhibition. This is a simplification done to reduce the computational load and the number of parameters. It is not meant to reflect any neurophysiological property of the inhibitory neurons or the neural connectivity pattern.

The two-layer field shows a delayed onset of inhibition according to the same mechanism described earlier for the biological neural system. In particular, if an external input is applied to the system, it drives the activation in the excitatory layer, while the inhibitory layer initially remains unchanged. When the activation of the excitatory layer reaches the threshold of the output function, the interactions start to come into effect. The lateral interactions within the excitatory layer drive activation further up locally, and at the same time the activation of the inhibitory layer is increased.

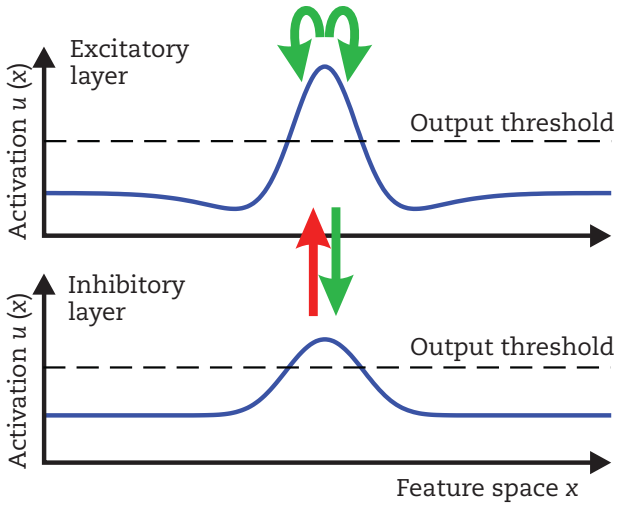

FIGURE 3.13: Architecture of two-layer field. The excitatory layer (*top*) projects onto itself and onto the inhibitory layer (*bottom*; green arrows). The inhibitory layer projects back onto the excitatory layer (red arrow). All projections are spread out and smoothed by Gaussian interaction kernels.

BOX 3.5 TWO-LAYER DYNAMIC FIELD

A two-layer field consists of an excitatory and an inhibitory activation distribution over the same feature space x, each governed by a differential equation. We designate the activation variable for the excitatory layer with the letter u, the one for the inhibitory layer with v. The basic structure for the two-layer field contains three projections: an excitatory projection from layer u to itself, a second excitatory projection from layer u to layer v, and an inhibitory projection back from layer v to layer u. Each of them is specified by an interaction kernel k that describes the connection weight as a function of distance in feature space. The three kernel functions are k_{uu}, k_{vu}, and k_{uv}. Here, the first letter in the index always designates the target of the projection; the second, its origin. The field equations are then:

$$\tau_u \dot{u}(x,t) = -u(x,t) + h_u + s(x,t) + \int k_{uu}(x-x')\, g\big(u(x',t)\big)\, dx' - \int k_{uv}(x-x')\, g\big(v(x',t)\big)\, dx'$$

$$\tau_v \dot{v}(x,t) = -v(x,t) + h_v + \int k_{vu}(x-x')\, g\big(u(x',t)\big)\, dx'$$

The output function g is again a sigmoid (logistic) function as in the one-layer system. The interaction kernels are typically Gaussian functions of the form:

$$k_{uu}(x-x') = c_{uu} \cdot \exp\left(-\frac{(x-x')^2}{2\sigma_{uu}^2}\right)$$

The parameter c_{uu} specifies the strength of the projection, the parameter σ_{uu} the width of the Gaussian kernel. The inhibitory kernel may include an additional constant term to produce global inhibition.

In this formulation, the effective width of inhibition is determined by both the kernels k_{uv} and k_{vu}. It is sometimes desirable to simplify this by omitting one of the kernels and using a simpler point-to-point connection for the projection from the layer u to layer v. This yields the dynamical system

$$\tau_u \dot{u}(x,t) = -u(x,t) + h_u + s(x,t) + \int k_{uu}(x-x')\, g\big(u(x',t)\big)\, dx' - \int k_{uv}(x-x')\, g\big(v(x',t)\big)\, dx'$$

$$\tau_v \dot{v}(x,t) = -v(x,t) + h_v + c_{vu}\, g\big(u(x,t)\big)$$

If only global inhibition is required in a model, this architecture can be further simplified by replacing the continuous inhibitory layer by a single inhibitory node. This node receives input from the whole excitatory layer and projects homogeneous inhibition back to it:

$$\tau_u \dot{u}(x,t) = -u(x,t) + h_u + s(x,t) + \int k_{uu}(x-x')\, g\big(u(x',t)\big)\, dx' - c_{uv}\, g\big(v(t)\big)$$

$$\tau_v \dot{v}(t) = -v(t) + h_v + c_{vu} \int g\big(u(x,t)\big)$$

Note that this formulation with a single inhibitory node shows a somewhat different behavior than the form with a continuous layer and purely global inhibition: In a continuous layer, the total output can increase very gradually as an activation peak becomes wider. When only a single node is used, the total output is always the sigmoid of the single activation variable. It can be useful to choose a sigmoid function with a very shallow slope here to allow a more gradual increase of the inhibition.

However, at this point there are still no inhibitory interactions active; activation in the inhibitory layer is only beginning to rise and is still near the resting level. Only after some time, when the input from the excitatory layer has driven the activation in the inhibitory layer beyond the output threshold as well, does the inhibitory projection take effect. Until this happens, activation in the excitatory layer can rise under the influence of external input and self-excitation, without being controlled by inhibition. This can often result in an overshoot of excitation, with activation levels significantly higher than would be reached with instantaneous inhibition. This property of two-layer fields will be explored in the exercises for this chapter.

For moderate interaction strengths, the system will generally settle into a stable state after the initial overshoot, with balanced activation in the excitatory and inhibitory layers. However, the system is prone to some level of oscillation: Whenever the activation in one layer changes in a way that affects its output, it takes some time for the other layer to reach the new attractor state determined by the changed input. During this delay, the output of this other field still reflects its old state, not the new attractor it is moving to. For instance, when activation in the inhibitory layer is quickly rising after the initial overshoot of excitation, the inhibitory input this produces sets a new attractor for the excitatory layer—likely one that is much lower than the current activation level. But the excitatory layer doesn't move to this attractor instantaneously; instead, it keeps producing a strong output for some time, that keeps driving activation in the inhibitory layer. The result is now an overshoot of inhibition.

For certain configurations, the two-layer field can act as a stable oscillator that maintains the pattern of alternating excitatory and inhibitory overshooting indefinitely. Under most conditions, prolonged oscillations are undesirable. One way to reduce them, even in the presence of strong interactions between the two layers, is to use different time constants τ for the two layers' dynamics. For instance, if the time constant of the inhibitory layer is much lower (and its dynamics therefore faster) than that of the excitatory layer, its activation will quickly reach the new attractor state after its input changes. This gives the excitatory layer little time to overshoot and therefore strongly reduces oscillations.

Besides the stable oscillatory states, the two-layer dynamic field supports the same instabilities and stable states as those of the single-layer field. It can form stabilized peaks of activation (with associated peaks in both layers) in response to localized input by going through a detection instability. Peaks can become self-sustained in the absence of input if the interactions within the field are sufficiently strong. If the inhibitory projection is sufficiently broad, it can mediate competition between distant peaks and, together with the excitatory interactions, produce selection decisions if two or more localized inputs are applied. For these reasons, one-layer and two-layer fields may often be used interchangeably when the focus is on more macroscopic properties of DFs. The advantage of the two-layer model is that it can produce more realistic results with respect to detailed activation time courses. This is demonstrated with concrete examples in the next section.

Fitting Neural Data for Movement Preparation with Dynamic Field Models

To model the activation time course for movement preparation, Bastian and colleagues (1998) employed a simplified form of the general two-layer architecture. Since the model requires no localized inhibitory projections but only global inhibition, the continuous inhibitory layer was replaced by a single dynamic node (see Box 3.5). This node receives positive input from the whole excitatory field and projects homogeneous inhibition back to it. This corresponds to a population of inhibitory neurons that have homogeneous connectivity to all excitatory neurons involved in the formation of the movement plan, independent of their preferred direction. The model then consists of an excitatory layer that spans the space of possible reach directions (from 0° to 360°) and the single inhibitory node. External input is then applied to the excitatory layer to reflect the stimulus settings in the experimental study. The first input reflects the pre-cues, consisting of either one, two, or three Gaussians, each centered on one of the six possible reach directions (always contiguous to each other in the case of multiple pre-cues). Then, the stimulus indicating the definite reach goal is modeled as a single, stronger Gaussian input appearing after a fixed delay.

The activation time course in the model presented by Bastian et al. reproduces the key observations in the DPA analysis of experimental data (Figure 3.10d–e). There is an initial steep rise

of activation following the presentation of the pre-cue, then a transient decrease during the delay period. Activation then rises again more strongly in response to the definite reach stimulus and falls to resting level at the end of the trial. The activation induced by the pre-cue is always centered on the midpoint of the pre-cued locations and retains its shape over the delay period. The activation profile becomes broader as the number of pre-cued locations is increased, but also flatter due to the normalizing effects of inhibitory interactions. If the definite reach direction indicated by the second stimulus is not at the center of the pre-cue profile, the peak of activation is shifted to the correct location by the second stimulus input.

Some comments on the process of fitting the DF model are warranted here. While some of the qualitative effects observed in the experiment can be reproduced directly through the field mechanics without any specific tuning, obtaining a reasonable fit of the activation time course requires a careful choice of parameters in the DF model. Unlike in more neurophysiological models, the parameters here are generally not constrained by anatomical or physiological properties of biological neurons, such as actual membrane potentials or ranges and patterns of synaptic connectivity. The DPA provides a functional description of the population representation that captures activation with respect to behaviorally or perceptually relevant feature spaces. Accordingly, the parameters of the DF model have to reflect the functional properties of the population activity and not the characteristics of single neurons or connections. The model fitting involves determining properties of the external input for the field model, interaction patterns, and timing parameters. The width of the interactions should reflect the width of typical activation patterns for simple stimuli as well as ranges of any explicitly tested interaction effects. The required parameters for the interaction strengths can to some degree be estimated from the stability of the population activation under changing inputs and the strength of normalization effects. To obtain quantitative fits of experimental data, extensive tuning of model parameters is often necessary. This is performed through repeated adjustments of model parameters and comparisons of simulation results and experimental data (either manually or using some form of optimization algorithm). Analytical solutions for these optimization problems are generally not available, except for the very simplest DF systems (Amari, 1977).

Relationship Between DPAs, Dynamic Fields, and Neural Populations

Before we conclude this chapter, we would like to contrast the different concepts addressed here. We introduced the method of DPA as an analysis method for neural data. The DPA performs a transformation from firing rates of individual neurons into a continuous activation distribution. This allows a specific view onto neural activity, focusing in particular on what is represented in a neural population with respect to perceptual or behavioral variables. The DPA method does not generate any novel data, and it does not by itself explain how certain activation patterns come about or how they change over time. While we can generate activation time courses with the DPA method, as shown in the previous examples, these only describe what was measured by electrophysiological recordings and do not make any statement about what drives the changes in activation. What the DPA method can do, however, is give us some understanding of how neural processing relates to behavior and by what mechanisms it might be governed.

In contrast, DF models are actually generative models. Through a set of differential equations, they specify concrete rules according to which activation patterns change over time. With a DF model, one can try out arbitrary stimulus settings and time courses and see what activation patterns they produce. And for every point along feature space and at every moment in time, one can pinpoint what factors influence the activation level. The link between DPAs and DF models is that both employ the same form of representation, namely a continuous activation distribution over a metric feature space. Through this shared format, the DPA also links DF models with biological neural data. While DF models inherently make predictions about activation patterns, the DPA method makes it possible to interpret these predictions in terms of biological neural response patterns and to test them with empirical data.

There are some caveats to keep in mind when considering this link between DF models and biological neural populations. First, many DF models—especially when formulated at a behavioral level—do not specify the region of the brain in which the neural activation patterns should be observable. In particular, a single DF in a model can

generally not be assumed to correspond directly to a specific cortical or subcortical region. Since DFs are typically defined based on functional considerations, they may describe activation patterns that are in fact distributed over several areas in the brain (for an example of this using a DF-based approach to fMRI, see Buss, Magnotta, Schöner, Huppert, & Spencer, 2014). Conversely, the activation patterns described in two different DFs may be intermixed in the biological system in a single area.

This last point touches on another important aspect of the relationship between DFs and biological neural representations: DF models do not generally attempt to describe full activation patterns of a specific neural population, but only the activation with respect to a certain feature or parameter space as is relevant for a task. For instance, one may define a field over the space of edge orientation that models a certain aspect of processing in the early visual cortex, ignoring the sensitivity of these cortical regions for other features, such as color, spatial frequency, and stimulus position. We have encountered an analogous limitation for the DPA method when applied to the visual cortex: Since activation distributions are computed from experimentally observed tuning curves of neurons, they can only be determined with respect to those parameter spaces for which tuning curves are measured (through systematic variation of the stimulus parameters). In that example, only the spatial tuning curves were determined, while dependence of the firing rates on other visual features was not tested. Consequently, the resulting activation distributions are defined only over visual space and do not yield any information about the sensitivity of this cortical region to other features.

It is important to keep this limitation to certain feature spaces in mind when interpreting the results of a DPA or when matching DF models to cortical regions. Choosing an inappropriate feature space in a DPA analysis, for instance, can lead to misleading results if the sensitivity of a neural population for that feature is only incidental and not functionally significant. On the other hand, abstraction from complex neural responses reflecting different features and behavioral parameters to only a few selected feature spaces can be helpful for forming concise models.

CONCLUSION

In this chapter, we have shown that models based on DFT can account for neural population data in quantitative detail. This firmly establishes that DFT is grounded in neurophysiology and supports the hypothesis—central to DFT—that population activation is the privileged level of description at which neural process accounts of perception, action, and cognition can be achieved. This hypothesis is aligned with a growing consensus in neuroscience that population activity provides the best prediction of behavior (Cohen & Newsome, 2009).

In DFT, peaks of activation in dynamics fields are units of representation; their locations in the field are estimates of the sensory, cognitive, or motor parameters that a DF represents. The peak location corresponds to the population vector of neurophysiology. Peaks localized in DFs are not necessarily localized within a cortical area. Whether that happens or not is a question of the topographical organization of the parametric map in the area. In the absence of topographical order, neurons tuned to similar values along the dimension of the field may be spatially distributed within the area, as happens for motor cortex. The construction of a distribution of population activation, or DPA, frees mapping of neural activity in the brain onto DFs of the constraints of topography. In the end, what is functionally significant is the connectivity of neurons in the brain, not how they are physically arranged. The DPA is constructed over a behaviorally relevant dimension (e.g., a sensory or motor space) in which perceptual or motor states are points. Each neuron contributes its entire tuning curve or receptive field profile. Neurons are thus "smeared out" across the DPA. A field location is not represented by an individual neuron. So when a peak of activation contributes to the specification of a behavior, it is really an entire subpopulation of neurons that makes that contribution.

In DFT, peaks of activation are stabilized by neural interaction. Signatures of such interaction are observed in neural data, which we reviewed. The fact that peaks are attractor states is critical for the entire framework of DFT. We will see in the rest of the book that the stability property of peak solutions is at the basis of how DFT generates cognitive function. Decisions arise as peaks emerge from instabilities of non-peak solutions. Decisions are stabilized because the peaks that instantiate decisions are stable states. We will see how DFT architectures work because individual activation fields function robustly, even as they are richly coupled with other fields. We will see how sustained peaks form the basis for working memory. Comparisons, selection decisions, coordinate transforms, or any

transformation of representational states requires stability of the units of representation.

In neural terms, stabilizing peaks of activation is costly. Local excitation and global inhibition require neural connectivity within the population that is sufficiently strong to potentially overrule incoming signals (e.g., in a selection decision). Using a neural population in an activation field to represent a single feature dimension is an expensive solution. This becomes dramatically obvious when the spaces to be represented become high dimensional, say, have 10, 20, or 50 dimensions. Why would the CNS use all this computational machinery just to represent points in an admittedly high-dimensional space? This will be discussed in Chapters 5 and 8, where we will look at the binding problem, in which dimensions such as color, texture, and orientation are represented in individual fields, each combining the feature dimension with visual space. Binding is achieved by linking activation peaks across such fields through shared spatial dimensions.

The mathematics of the DFT framework builds on modeling that was performed in the 1970s at a more biophysically detailed level of description to capture the dynamics of neural activity in small cortical populations (Wilson & Cowan, 1972). Recognizing that the cytoarchitectonics of cortical layers are relatively homogeneous along the cortical surface, with strongly overlapping dendritic trees and a reproducible structure of neural networks, these authors proposed a neural field dynamics, in which the cortical surface was described as an excitable continuum (Wilson & Cowan, 1973). In one way, this was a precursor to the ongoing quest to identify fundamental functional circuits at this level of description (e.g., Binzegger, Douglas, & Martin, 2004). On the other hand, the authors realized that the neural dynamics of their models gave rise to activation patterns that were not mere transformations of their inputs, but autonomously generated patterns of activation. In hindsight, it is curious that the self-excited activation patterns and neural oscillation observed in these models had relatively little impact on the field of cortical neurophysiology. This was the decade after Hubel and Wiesel's (1959, 1968) breakthrough discovery of the functional architecture of the cortex, which shaped the thinking of neurophysiologists through the concepts of tuning curves, receptive fields, and cortical maps. These concepts are, at first approximation, reflections of the forward connectivity from the sensory surface to the cortical layer. So most empirical questions were then focused on that forward connectivity. Intracortical interaction was thought to merely modulate such forward maps (Ferster & Miller, 2000). The activation patterns generated by strong interaction in the neural dynamic models was associated with phenomena outside the regular function of cortex, such as hallucinations or epilepsy (Ermentrout, 1998).

There is a modern literature on the dynamics of neural fields which studies, in the spirit of applied mathematics, the class of solutions and dynamic phenomena that are possible within different types of mathematical models formulated on the basis of biophysical and neuroanatomical principles (Coombes, beim Graben, Potthast, & Wright 2014). This literature is useful to modelers working within the framework of DFT, as it provides exemplary mathematical models that are well understood and can serve as concrete mathematical formalizations of conceptual accounts. Amari's analysis of the dynamics of one- and two-layer neural fields (Amari, 1977), on which most of the models in this book are based, was a trailblazer of this type of approach. By identifying the different attractor states and their bifurcations, Amari's work enabled us to map units of representation, peaks, and subthreshold activation patterns onto different attractor regimes of his neural field dynamics.

There is also a literature of modeling populations of neurons at a more biophysically detailed level of description, often the level of spiking neurons. Only recently have these models begun to connect to cognitive function or behavior, primarily in the domain of perceptual decision-making and working memory. In some cases, modelers working at the spiking level simply reproduce the dynamic phenomena observed at the population level and compare them qualitatively with single-neuron tuning curves or firing patterns (Wei, Wang, & Wang, 2012; Wong & Wang, 2006). Ultimately, the goal is to establish how mechanisms at the level of synaptic, membrane, or other single-cell mechanisms relate to cognitive function (Durstewitz, Kelc, & Güntürkün, 1999). It is often found that population activity modeled at the spiking level is congruent with population activity modeled as space-time continuous dynamics (Deco, Jirsa, Robinson, Breakspear, & Friston, 2008). In fact, the neural dynamics of population activation can be viewed as a macroscopic approximation of the more microscopic description, an approximation

that runs under the label "mean-field theory" (Trappenberg, 2010). Recently, systematic efforts have been made to mathematically derive neural dynamics at the population level from the dynamics of populations of spiking neurons (Faugeras, Touboul, & Cessac, 2009).

A vast literature exists for neural network models that are primarily characterized by the forward connectivity from sensory systems to cortical representations. Most connectionist modeling is in this fold, but so is modeling that is closely tied to cortical functional architecture (e.g., Riesenhuber & Poggio, 2000). Attempts have been made to derive the structure of cortical architecture from abstract principles (Wiskott & Sejnowski, 2002). Such models ultimately project onto a "decision" layer, within which the perceptual information from the sensory surface is in some sense optimally encoded. On that decision layer, additional computations must be made to then actually perform the decision. For instance, a classifier may learn to associate the output of a feed-forward network with particular object classes (Riesenhuber & Poggio, 2000). A possible ultimate vision of the DFT framework could be that such complex forward neural networks would replace the simple input–output mappings used in most DFT models to provide localized input along the dimensions that activation fields represent. This presupposes that the forward connectivity is organized so that functional neighborhoods emerge in which neighboring sites on the decision layer represent neighboring choices. Self-organized feature maps (Kohonen, 1982; Sirosh & Miikkulainen, 1994) are the candidate structures for how such a mapping could come about. In Part 3 of the book we will look at learning forward projections, although this topic needs to be explored further than covered in this book. A primary difficulty is the strong reduction in dimensionality that a mapping onto self-organized feature maps implies. Chapter 5 argues that the sensory array may typically have 10,000 or more inherent dimensions—that is, that the patterns of sensory stimulation may change in 10,000 or more different ways. Forward neural networks from the sensory surface may strongly compress this number of dimensions, because stimuli coming from the real world do not vary independently in all these dimensions. Even so, the outcome of such compression for a neural representations of visual objects, for instance, still leaves hundreds of relevant dimensions (Kurková et al., 2008). As mentioned earlier, each field considered in this book represents only a handful of dimensions, at best. The theoretical reason for this limitation lies in the stabilization of peaks by neural interaction. The neural connectivity of local excitation and global inhibition becomes increasingly costly with increasing number of dimensions. This is an as-yet open issue that requires more study and deeper understanding.

One radical alternative is to give up the stability requirement altogether. Some researchers have argued that neural computation can do without stable states, being instead based on transients (Maass, Natschl, & Markram, 2002). This idea has recently been linked to the notion of vector symbolic architectures (VSAs), first pursued by Smolensky (1990) to extend connectionism to higher cognition, and now implemented in spiking neural network models (Eliasmith, 2013). In VSAs, neural patterns are used to encode high-dimensional information. For instance, an activation vector built from 1000 neurons is thought to encode 1000 dimensions with the activation level of each neuron encoding one dimension. Such high-dimensional vectors tend to be uncorrelated just by the geometry of high-dimensional space; there are a lot more ways vectors can be orthogonal to each other than for them to be parallel to each other. This makes it possible to superpose vectors, combine them, and to again extract components from them, all typical operations of information processing. The idea is then that the computations of conventional information processing can be realized in VSA by passing activation patterns along a processing chain in a sequence of transient neural states. One open question is how such a system may interface with sensory-motor processes, for which stability is clearly a necessity. More generally, the interface of VSAs to both sensory and motor information requires a form of recoding, in which sensory information is encoded by creating a high-dimensional neural pattern vector and motor commands are then generated by decoding them from high-dimensional neural pattern vectors. Such interfaces make it difficult for cognitive processes to remain linked to online sensory information and ongoing motor action. They also make it difficult to generate sequences of mental operations that are aligned with their physical acting-out in the world (we will study this in detail in Chapter 14). Finally, there is to date no behavioral or neural evidence for such a divide between the sensory and motor domains and an information-processing domain.

Can stability be retained as a property of neural processing while still representing high-dimensional information? One possibility is to tailor the neural connectivity to specifically stabilize particular, complex patterns of neural activity. This is what the Hopfield network does (Hopfield, 1982, 1984). The idea is that, to encode high-dimensional patterns, the network learns both the forward connectivity to induce the pattern and the interaction connectivity to stabilize the pattern. Exactly how a Hopfield network could perform the functions of DFT is not clear at this time. In particular, it is not easy to conceive of something like detection instability in a Hopfield network. Such a network is always in some stable pattern of activation. It isn't clear that it has an "off" state, where it represents the absence of any particular pattern, and can then transition to an "on" state, where it may initiate an action or mental operation. In Chapter 12 we will explore how far the DFT framework goes toward capturing the learning of object representations using only low-dimensional feature representations. Moving toward more complex, higher-dimensional representational states is one of the research frontiers of DFT.

But first we need to return to the tight link of the low-dimensional DFs to the sensory and motor domains and their coupling to behavioral dynamics, in the next chapter.

REFERENCES

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics, 27*(2), 77–87.

Bastian, A., Riehle, A., Erlhagen, W., & Schöner, G. (1998). Prior information preshapes the population representation of movement direction in motor cortex. *Neuroreport, 9*(2), 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*(7), 2047–2058.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature, 436*(7054), 1161–1165.

Binzegger, T., Douglas, R. J., & Martin, K. A. C. (2004). A quantitative map of the circuit of cat primary visual cortex. *Journal of Neuroscience, 24*(39), 8441–8453.

Britten, K. H., & Newsome, W. T. (1998). Tuning bandwidths for near-threshold stimuli in area MT. *Journal of Neurophysiology, 80*(2), 762–770.

Buss, A. T., Magnotta, V., Schöner, G., Huppert, T. J., & Spencer, J. P. (2014). Testing bridge theories of brain function using theory-driven fMRI. Manuscript submitted for publication.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron, 45*(5), 801–814.

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.

Conway, B. R., & Tsao, D. Y. (2009). Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proceedings of the National Academy of Sciences U.S.A., 106*(42), 18034–18039.

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

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Computational Biology, 4*(8), e1000092.

Durstewitz, D., Kelc, M., & Güntürkün, O. (1999). A neurocomputational theory of the dopaminergic modulation of working memory functions. *Journal of Neuroscience, 19*, 2807–2822.

Eccles, J. (1976). From electrical to chemical transmission in the central nervous system. *Notes and Records of the Royal Society of London, 30*(2), 219–230.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York: Oxford University Press.

Erickson, R. (1974). Parallel "population" neural coding in feature extraction. In F. Schmitt & F. Worden (Eds.), *The Neurosciences. Third Study Program* (pp. 155–169). Cambridge, MA: MIT Press.

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.

Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics, 61*, 353–430.

Faugeras, O., Touboul, J., & Cessac, B. (2009). A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. *Frontiers in Computational Neuroscience, 3*, 1–28.

Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Reviews of Neuroscience, 23*, 441–471.

Fitzgerald, P. J. (2006). Receptive field properties of the macaque second somatosensory

cortex: Representation of orientation on different finger pads. *Journal of Neuroscience, 26*(24), 6473–6484.

Georgopoulos, A. P. (1995). Motor cortex and cognitive processing. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 507–517). Cambridge, MA: MIT Press.

Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience, 2*(11), 1527–1537.

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 the direction of movement by a neuronal population. *Journal of Neuroscience, 8*(8), 2928–2937.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science, 233*(4771), 1416–1419.

Georgopoulos, A. P., Taira, M., & Lukashin, A. (1993). Cognitive neurophysiology of the motor cortex. *Science, 260*(5104), 47–52.

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.

Harris, L. R., & Jenkin, M. R. M. (1997). Computational and psychophysical mechanisms of visual coding. In M. R. M. Jenkin & L. R. Harris (Eds.), *Computational and psychophysical mechanisms of visual coding* (pp. 1–19). Cambridge, UK: Cambridge University Press.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences U.S.A., 79*, 2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences U.S.A., 81*, 3088–3092.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*(1), 215–243.

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*(20), 9016–9028.

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*(6), 1187–1211.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59–69.

Kurková, V., Neruda, R., Koutník, J., Franzius, M., Wilbert, N., & Wiskott, L. (2008). Invariant object recognition with slow feature analysis. In *Artificial Neural Networks—ICANN 2008* (Vol. 5163, pp. 961–970). Berlin: Springer-Verlag.

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature, 332*(6162), 357–360.

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation, 14*(11), 2531–2560.

Markounikau, V., Igel, C., Grinvald, A., & Jancke, D. (2010). A dynamic neural field model of mesoscopic cortical activity captured with voltage-sensitive dye imaging. *PLoS Computational Biology, 6*(9), e1000919.

Nichols, M. J., & Newsome, W. T. (2002). Middle temporal visual area microstimulation influences veridical judgments of motion direction. *Journal of Neuroscience, 22*(21), 9530–9540.

Orban, G. A. (1984). *Neuronal operations in the visual cortex*. Berlin: Springer-Verlag.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology, 86*(5), 2505–2519.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience, 5*(12), 1332–1338.

Riesenhuber, M., & Poggio, T. (2000). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Salinas, E., & Abbott, L. F. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience, 1*, 89–107.

Schwartz, A. B., Kettner, R. E., & Georgopoulos, A. P. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement. *Journal of Neuroscience, 8*(8), 2913–2927.

Sherrington, C. S. (1906). *The integrative action of the nervous system*. New Haven, CT: Yale University Press.

Sirosh, J., & Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics, 71*, 65–78.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures

in connectionist systems. *Artificial Intelligence, 46*, 159–216.

Tchumatchenko, T., Malyshev, A., Wolf, F., & Volgushev, M. (2011). Ultrafast population encoding by cortical neurons. *Journal of Neuroscience, 31*(34), 12171–12179.

Trappenberg, T. P. (2010). *Fundamentals of computational neuroscience* (2nd ed.). Oxford, UK: Oxford University Press.

Treue, S., Hol, K., & Rauber, H. J. (2000). Seeing multiple directions of motion–physiology and psychophysics. *Nature Neuroscience, 3*(3), 270–276.

Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience, 6*(4), 1160–1170.

Wei, Z., Wang, X.-J., & Wang, D.-H. (2012). From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *Journal of Neuroscience, 32*, 11228–11240.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal, 12*, 1–24.

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

Wiskott, L., & Sejnowski, T. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation, 14*(4), 715–770.

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience, 26*(4), 1314–1328.

EXERCISES FOR CHAPTER 3

The interactive simulator provided in `launcher-TwoLayerField_preset` implements the extended field equation with two layers. The graphical user interface (GUI) shows the activation of two fields (or layers) as blue plots in two separate sets of axes, with the excitatory field u at the top and the inhibitory field v below. Field input and output are plotted in the same way as in the simulator for the one-layer field.

The sliders allow you to control field parameters (resting level and noise strength), interaction parameters, and input settings. Interactions in the architecture include self-excitation in field u with strength c_{uu}, excitation from field u to field v with strength c_{vu}, as well as local and global inhibition from field v to field u with strengths c_{uv}^{loc} and c_{uv}^{glob}, respectively.

Exercise 1: Detection Instability

Starting from the settings no `interactions` (default), try to set up the interactions in the two-layer field to produce a detection instability. Increase interaction strengths c_{uu}, c_{vu}, and c_{uv}^{loc}. When the activation level in u is driven beyond 0 by external input, a peak should form in both u (with activation higher than input) and v, and surround inhibition should be projected from v back to u. Notice how inhibition is only produced when there is supra-threshold activation in field v. Otherwise, the strength of the inhibitory projection, c_{uv}^{loc}, is irrelevant. If you have trouble finding appropriate parameters, you can select the predefined setting "stabilized" from the dropdown menu.

Test whether there is a bistable regime by applying an input to form a peak and then decreasing the input strength back to a level that initially did not induce a peak. The peak should remain stable when the input is decreased moderately, and only collapse once input is diminished more strongly. Once the excitatory peak in u disappears, the peak in v collapses as well.

Exercise 2: Self-Sustained Activation Peaks

Increase the interaction strengths to produce self-sustained peaks that remain stable even when the input is removed completely. If you are having trouble finding appropriate parameters, select the setting "memory" from the dropdown menu. You should be able to have multiple memory peaks in the field at the same time.

Exercise 3: Selection

Set the strength of the local inhibition, c_{uv}^{loc}, to zero and increase the strength of the global inhibition, c_{uv}^{glob}, so that you get self-stabilized peaks again (or choose the predefined setting "selection"). You should now be able to create a selection behavior: Set up two inputs at different locations, both of the same amplitude, so that they are just sufficient to drive the field activation beyond the output threshold. Press the Reset button to set the field activation back to the resting level and let it evolve in response to the input. You should get an activation peak at one input location and none at the other. Try varying the stimulus amplitudes to see how this influences selection behavior.

You can also try to change the parameter to get a single-peak memory behavior.

Exercise 4: Oscillations

You may already have seen some oscillatory behavior of the two-layer field during the previous

exercises. To explore this in detail, open the parameter panel and set the time constants τ of both fields to 20. Then set $c_{uu} = c_{vu} = c_{uv}^{\text{loc}} = 15$ and $c_{uv}^{\text{glob}} = 0$. Now apply a single localized input to induce a peak. You should be able to observe an overshoot of excitation after going through the detection instability: Activation in u rises strongly in the beginning, but then decreases again as inhibition starts to build up. You can use the Reset button to observe the time course of the peak formation multiple times.

If you now increase interaction strengths even further, you can create perpetual oscillations in the two-layer field.

Exercise 5: Repulsion Effect

Try to create the repulsion effect that was observed and modeled for the composite stimuli in the work on visual representations. Select the setting stabilized and create two self-stabilized peaks through local inputs. Keep shifting the peaks closer to each other by slowly changing the input positions. While the activation peaks are centered on the inputs when they are distant from each other, you can observe an outward deviation of the peak center from the input center when you move the inputs closer together (if you move them very close, the peaks will merge). You can experiment with the same effect for memory peaks.