16

Competing populations and decision making

We make multiple decisions in daily life. Should I cut across a busy street or take the safer pedestrian underground path which causes a 2-minute detour? Should I say "Hello" to the person I see on the other side of the street or move on? Should I spend money on a simple and cheap bicycle which is less likely to be stolen, or on a faster, shiny, expensive one? Which college should I choose after high school? Should I continue after college for graduate studies and get a PhD? Some of these are small decisions of minor relevance, but there are also important decisions that can influence the course of life for several years.

Decisions are most easily analyzed in the context of games. Small children already learn in board games that they need to decide between several possibilities. A typical example is shown in Fig. 16.1a. Would you advise a child to take the safe long path to the left, or the shorter one with the risk of being reset to "Start"? What would you decide?

The situation depicted in the board game presents a choice between a safe and a risky option. It is typical for decision problems that are empirically studied in the field of neuroeconomics (Platt and Huettel, 2008; Rangel *et al*., 2008; Glimcher *et al*., 2008). Suppose that you have a choice between winning 100 dollars with 100% probability or 200

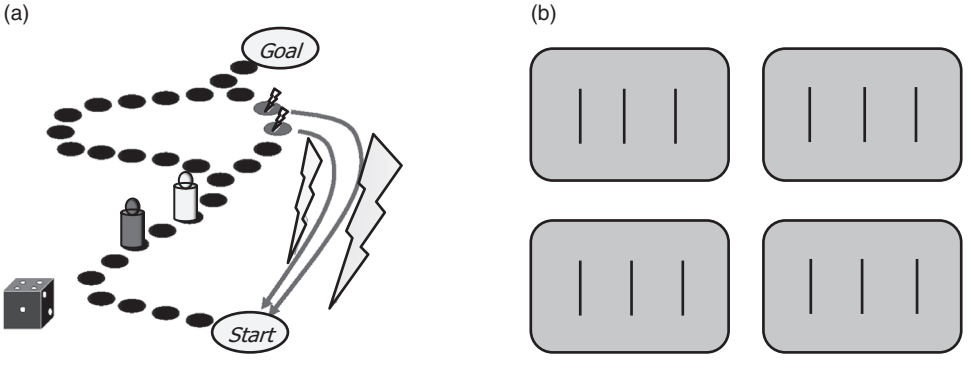

Fig. 16.1 Decision processes. (a) In a board game, your die shows the digit 4, and you have to move the white token. The right path is shorter, but more risky, because the token has to restart if it ends on one of the flashed fields. How would you decide? (b) Perceptual decision making. Three vertical bars are presented on a gray screen (four examples are shown): Is the central bar shifted left or right compared to a perfectly symmetrical arrangement?

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

dollars with 50% probability, which option would you choose? Suppose you just received 200 dollars, but you now have the unfortunate choice between losing half of it (100 dollars) with 100% probability or even all of it (200 dollars) with 50% probability, which option would you choose? If the brain activity of a human subject is imaged, while he answers these or similar monetary questions, the areas of highest activity associated with value, risk, and loss can be identified, at the coarse resolution of brain regions (Platt and Huettel, 2008; Rangel *et al*., 2008).

In this chapter we work on a more microscopic level, i.e., that of neuronal activity during decision making (Gold and Shadlen, 2007). Decision making requires (i) a suitable representation of inputs and potential outcomes as well as of the values attached to the options; (ii) a selection process that picks one of the options; and (iii) potentially also some feedback that enables learning so as to achieve improved performance over several trials (Rangel *et al*., 2008). Decision making involves different brain systems and has conscious as well as unconscious aspects (Sanfey and Chang, 2008). Here, we focus on the dynamic *selection* between different options in the context of perceptual decision making. There are three reasons for this focus. First, measurements of neuronal activity of single neurons or groups of neurons are available that indicate a correlation of neural activity with the choice made during a decision. Second, these experimental measurements can be linked to neuronal models of decision making. And, finally, the moment of the final selection between different choices lies at the heart of decision making.

In Section 16.1, we review some of the classic recordings of neural activity during decision making in monkeys (Gold and Shadlen, 2007). In Section 16.2, a model is presented that describes the process of decision making as a competition between neuronal populations that share the same pool of inhibitory neurons (Wang, 2002). The mathematical analysis of the dynamics in such a model of competition is outlined in Section 16.3. Alternative descriptions of decision making are presented in Section 16.4. We close the chapter by situating models of decision making in the larger context of fundamental questions related to the notion of "free will" (Section 16.5).

16.1 Perceptual decision making

Many perceptual phenomena can be formulated as a problem of decision making. In a typical experiment of visual psychophysics, a subject observes a short flash of three vertical black bars on a gray background (Fig. 16.1b). Is the middle bar shifted to the left or to the right compared to a symmetric arrangement of the three bars where it is exactly in the center? If the shift is very small, or if the bars are presented with low contrast on a noisy screen, the question is difficult to answer. The subject who holds a button in each hand, indicates his decision (left or right) by pressing the corresponding button. In other words, he reports his perception as a decision.

In what follows, we focus on an experimental paradigm with visual random dot motion stimuli used for the study of perceptual decision making in monkeys (Salzman *et al*., 1990; Roitman and Shadlen, 2002; Gold and Shadlen, 2007). The stimulus consists of a random

pattern of moving dots, where most, but not necessarily all, of the dots move coherently in the same direction; see Fig. 16.2. Typically, two different directions of motion are used, for example upward or downward. The monkey has been trained to indicate the perceived motion direction by saccadic eye movements to one of two targets see Fig. 16.2b.

Note that, in contrast to the examples given at the beginning of this chapter, problems of perceptual decision making typically have no direct monetary value or risk associated with them. Normally we do not care whether a bar is shifted to the left or to the right, or whether dots move upward or downward. Nevertheless, a correct perceptual decision might be life-saving if a moving stripe pattern in the bush is correctly recognized as an approaching tiger as opposed to a movement of the leaves in the wind.

16.1.1 Perception of motion

Neurons in the middle temporal visual area (MT, also called V5) are activated by large-scale motion stimuli. The receptive field of an MT neuron, i.e., the region of visual space that is sensitive to motion stimuli, is considerably larger than that of a neuron in the primary visual cortex; see Chapter 12. Different neurons in MT respond to different directions of motion, but just as in other parts of the visual cortex, area MT has a columnar structure so that clusters of neighboring neurons share receptive fields with a similar preferred direction of motion (Albright *et al*., 1984).

At the beginning of a typical recording session with an extracellular electrode in MT (Salzman *et al*., 1990), the location of the receptive field and the preferred direction of motion of a single neuron or cluster of neighboring neurons is determined by varying the movement angle and the location of the random dot stimulus. Once the receptive properties of the local MT neurons have been determined, only two different classes of stimuli are used, i.e., dots moving coherently in the preferred direction of the recorded neuron, and dots moving coherently in the opposite direction.

After each presentation of a random dot motion pattern, two targets are switched on, one at a location in the direction of stimulus motion, the other one on the opposite side. The monkey is trained to indicate the movement direction of the stimulus by a saccadic eye movement to the corresponding target. After training, the perceptual decision between a dot movement in the cell's preferred direction (P) or the null direction (N) is reliably performed by the monkey if a noise-free stimulus is used where all dots move in the same direction. However, the task becomes more difficult if only a small fraction of dots move coherently in one of the two directions while the rest of the dots move in a random direction. The behavioral performance can be assessed with the psychometric function which represents the percentage of saccades to the target P as a function of coherence, where coherence indicates the fraction of coherently moving dots (Fig. 16.2b).

An electrode in MT can be used not only to record neural activity, but also to stimulate a cluster of neurons in the neighborhood of the electrode. Since neighboring neurons have similar preferred directions of motion, current injection into the electrode can bias the perception of the monkey in favor of the neurons' preferred direction, even if the random

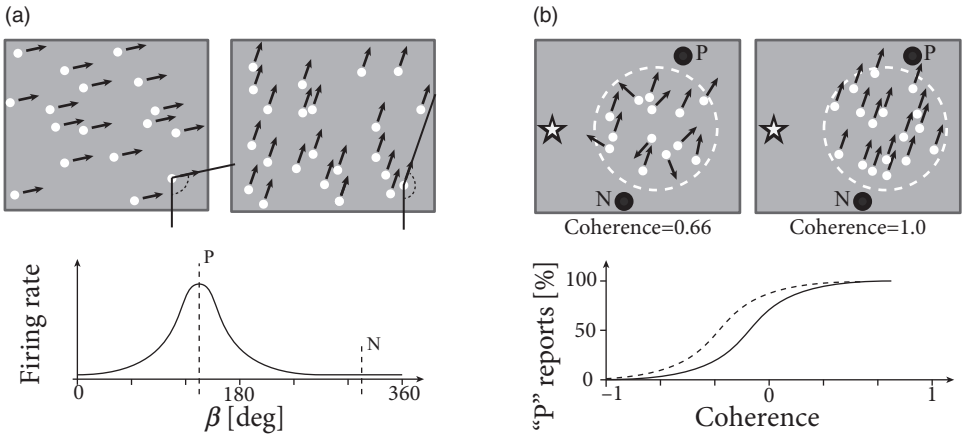

Fig. 16.2 Random dot stimuli and perception of motion. (a) Top: A pattern of random dots moving in direction β is presented on the screen. Different motion directions are tested. Bottom: The firing rate response of a neuron in area MT depends on the direction of motion β of the random dot stimulus. The preferred direction is marked "P," the null direction "N" (schematic figure). (b) Top: In the first phase of each trial, the monkey fixates on the star while a moving random dot stimulus is presented inside the receptive field (dashed circle) of a neuron. After visual stimulation is switched off, the monkey indicates by eye movements to one of the two targets (solid black circles, marked P and N) whether the perceived motion is in the direction "P" or "N." Bottom: The percentage of "P" reports (vertical axis) is plotted as a function of the coherence (horizontal axis) of the stimulus (solid line). If, during presentation of a random dot motion pattern, the MT column of neurons with preferred direction "P" is electrically stimulated, the percentage of times the monkey reports a perception of "P" is increased (dashed line). Coherence of 1 indicates that all points move in the P direction, while coherence of 0.66 indicates that one third of the points move in a random direction. Coherence of -1 indicates coherent motion in the "N" direction; schematically redrawn after Salzman *et al*. (1990).

dot pattern has no or only a small amount of coherence (Fig. 16.2b). This indicates that the perceptual decision of the monkey relies on the motion information represented in the activity of MT neurons (Salzman *et al*., 1990).

While the monkey's perceptual decision is influenced by the manipulation of MT neurons, this result does not imply that the decision itself is made in MT. It is likely to be made at a later stage, in an area that uses the information of MT neurons.

16.1.2 Where is the decision taken?

The short answer is: we do not know. However, an interesting observation has been made in the lateral intra-parietal (LIP) area during experiments of perceptual decision making with moving random dot stimuli (Roitman and Shadlen, 2002).

Before discussing the experiment, we need to present a few facts about the properties of LIP neurons. Area LIP is located in the visual processing stream between the primary visual cortex and the frontal eye field region involved in control of saccadic eye move-

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

ments. Neurons in area LIP respond during the *preparation* of saccadic eye movements. Different neurons in LIP have different receptive fields. The location of the receptive field corresponds to a potential target region of eye movements. In other words, a LIP neuron responds just *before* a saccadic eye movement into its receptive field occurs.

As in the previous subsection, monkeys in the experiment of Roitman and Shadlen are trained to indicate the direction of a moving dot pattern by saccadic eye movements to one of two visual targets. The first target is located in the receptive field of a LIP neuron. Therefore, the recorded neuron is expected to respond whenever the monkey prepares a movement to the first target. The second target is located in the opposite direction. The task is designed such that a random dot stimulus moving in the direction of the first target indicates that the monkey should make an eye movement toward it; the correct response to a stimulus moving in the opposite direction is a saccade to the second target (Fig. 16.3a). The difficulty of the task can be varied by changing the fraction of coherently moving

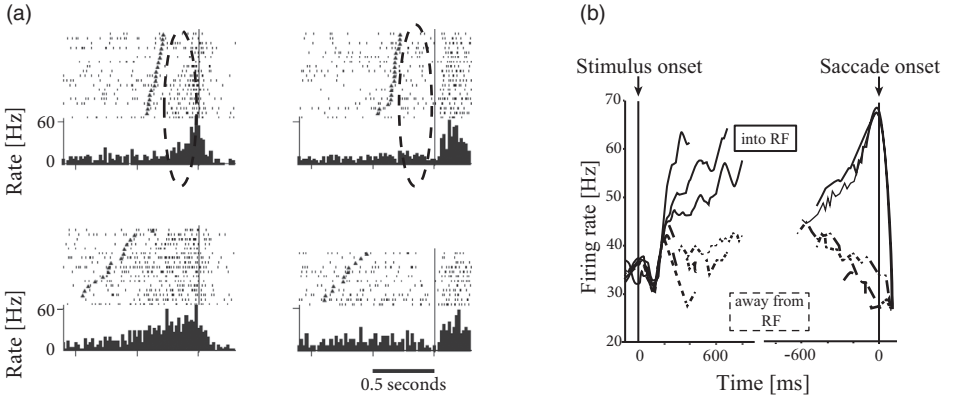

Fig. 16.3 (a) Neurons in the lateral intra-parietal (LIP) area have receptive fields (RF) that represent potential targets of a saccade. A LIP neuron responds strongly just before the saccade, if the saccadic movement is *into* its RF (left, dashed line surrounds region of interest), and is suppressed if the movement is in the opposite direction, away from its RF (right). Monkeys observed random dot motion stimuli and had been trained to report the direction of the stimulus by saccadic eye movements either "into" or "away from" the RF of the recorded neuron. For histograms and spike raster, trials were aligned to saccade onset (sac, vertical line). Filled triangles indicate onset of motion stimulus. Responses were faster for stimuli with larger coherence (top row, coherence = 51.2%) than small coherence (bottom row, coherence = 6.4%), and stronger for movements into the RF (left column) than away from the RF (right column). (b) Firing rate response of LIP neurons (averaged over 54 neurons) aligned to stimulus onset (left part of graph) or saccade onset (right part of graph). The stronger the coherence (thick solid line: coherence = 51.2%, other solid lines: 12.8% and 3.2%) of a random dot motion stimulus initiating a saccade "into" the RF, the faster the rise of the initial response of LIP neurons (left). However, whatever the coherence, the LIP neurons always reach the same firing rate, at the moment when a saccade into the RF starts (right). The neurons are suppressed, if the monkey chooses the opposite saccadic target ("away from the RF," dashed lines, left and right). Adapted from Roitman and Shadlen (2002).

dots. The behavioral reaction time of the monkey was measured as a function of stimulus coherence. At the same time, the activity of neurons in LIP was recorded.

Roitman and Shadlen found that, during the presentation of the moving dot stimulus, the activity of LIP neurons increased. The rate of increase after stimulus onset was higher for stimuli with a large fraction of coherent points than for stimuli with little or no coherence. Importantly, when the responses were averaged and aligned to the onset of the *saccade*, LIP neurons always reached the same level of activity just before a saccade *into* their receptive field (Fig. 16.3b).

These findings are consistent with the idea that the decision to perform a saccade occurs at the moment when LIP neurons reach a threshold value. For stimuli with a high degree of coherence, the activity increases more rapidly, the threshold is reached earlier, and reaction times are shorter than for stimuli with a low degree of coherence. Therefore, Roitman and Shadlen suggest that "a threshold level of LIP activity appears to mark the completion of the decision process" (Roitman and Shadlen, 2002).

16.2 Competition through common inhibition

The essential features of the experiments of Roitman and Shadlen (2002) can be described by a simple model of decision making where neuronal populations compete with each other through shared inhibition.

We consider a network of spiking neurons (Fig. 16.4) consisting of two excitatory populations interacting with a common pool in inhibitory neurons (Y. Wang *et al.*, 2002). Within the two excitatory populations neurons are randomly connected with connection weight w_{EE}. Connections to and from the inhibitory populations have weights w_{IE} and w_{EI}, respectively. Neuronal parameters and connection weights are adjusted such that, in the absence of external input, all neurons exhibit spontaneous activity at low firing rates. In other words, the network is in a state of asynchronous irregular firing.

Stimulation corresponds to a positive mean input into one or both groups of excitatory neurons. For example, for a description of the experiments of Roitman and Shadlen discussed in the previous section, we can identify input into population 1 as indicating

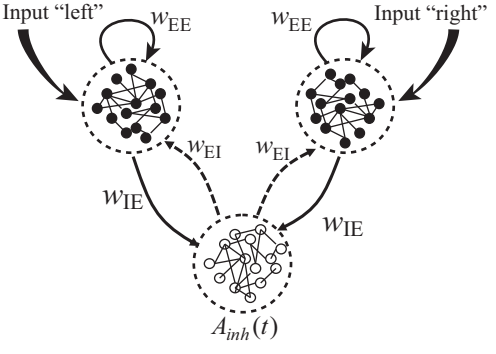

Fig. 16.4 Competition between neuronal pools. Two populations of excitatory neurons interact with a common pool of inhibitory neurons. Input signals indicating movement to the left are fed into population 1 with activity $A_{\mathrm{E},1}(t)$. Each population of excitatory neurons makes excitatory connections of strength w_{EE} onto itself. The inhibitory population receives input of strength w_{IE} from the two excitatory populations and sends back inhibition of strength w_{EI}.

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

coherent motion of the random dot pattern to the left whereas input into population 2 indicates motion to the right (Fig. 16.4). Since the stimulus in the experiments has a random component (e.g., the fraction of coherent dots is less than 100%), the input into each population is described as a mean plus some noise.

If the pattern has a high degree of coherence and moves to the left, the mean input to population 1 is high. This induces a high activity $A_{E,1}$ which in turn excites the inhibitory population which transmits inhibition to both excitatory pools. However, only the stimulated pool can overcome the inhibition so that the activity of the other excitatory population is suppressed. Since, at most one of the two populations can be active at the same time, the two populations are said to "compete" with each other. The competition is induced by the shared inhibition. If the external stimulus favors one of the two populations, the population receiving the stronger stimulus "wins" the competition. In the absence of external stimulation, or for a weak unbiased stimulus, both populations exhibit low activity.

To highlight the dynamics of competition, let us now focus on a strong, but unbiased stimulus. Here, unbiased means that, after stimulus onset, both excitatory populations receive an input of the same mean, but with a different realization of the noise (Fig. 16.5a). Immediately after the onset of stimulation, both excitatory populations increase their firing rates. Soon afterward, however, one of the activities grows further at the expense of the other one, which is suppressed. The population which develops a high activity is called the "winner" of the competition. In the next section, we will show mathematically how the shared inhibition induces a competition between the two excitatory populations.

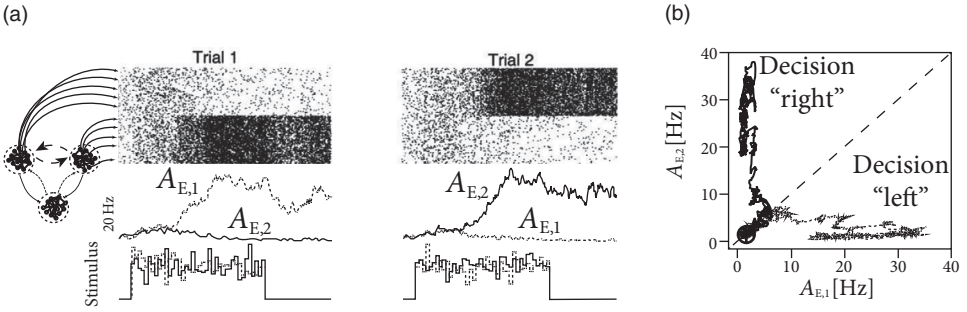

Fig. 16.5 Competition between neuronal pools. (a) Top: Spiking activity of two populations in trial 1 (left) and trial 2 (right). Dots denote spikes. Average across the population gives the population activities $A_{E,1}(t)$ and $A_{E,2}(t)$. During presentation of an unbiased stimulus (e.g., equal number of points moving to the left and to the right), one of the excitatory population develops a high population activity, while the other one is suppressed, indicating a spontaneous decision to the left (trial 1) or to the right (trial 2). Bottom: An unbiased stimulus corresponds to an input to the left and right populations of equal mean, but different realizations of noise. (b) The dynamics of population activities $A_{E,1}(t), A_{E,2}(t)$ can be visualized in the phase plane. In the absence of stimulation, the activity of both excitatory populations exhibits a low firing rate of less than 5 Hz (circle). Upon stimulation, the dynamics converge either to the region corresponding to "decision left" (characterized by high values of $A_{E,1}$) or to "decision right" (characterized by high values of $A_{E,2}$). Adapted from Wang *et al*. (2002).

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

16.3 Dynamics of decision making

In this section, we present a mathematical analysis of decision making in models of interacting populations. We start in Section 16.3.1 with the rate equations for a model with three populations, two excitatory ones which interact with a common inhibitory population. In Section 16.3.2, the rate model with three populations is reduced to a simplified system described by two differential equations. The fixed points of the two-dimensional dynamical system are analyzed in the phase plane (Section 16.3.3) for several situations relevant to experiments on decision making. Finally, in Section 16.3.4 the formalism of competition through shared inhibition is generalized to the case of K competing populations.

16.3.1 Model with three populations

In order to analyze the model of Fig. 16.4, we use the rate equations of Chapter 15 and formulate for each of the three interacting populations a differential equation for the input potential. Let

$$A_{\mathrm{E},k} = g_{\mathrm{E}}(h_{E,k}) \tag{16.1}$$

denote the population activity of an excitatory population k driven by an input potential $h_{E,k}$. Similarly, $A_{\mathrm{inh}} = g_{\mathrm{inh}}(h_{\mathrm{inh}})$ is the activity of the inhibitory population under the influence of the input potential h_{inh}. Here g_{E} and g_{inh} are the gain functions of excitatory and inhibitory neurons, respectively. The input potentials evolve according to

$$\tau_{\mathrm{E}} \frac{\mathrm{d}h_{E,1}}{\mathrm{d}t} = -h_{E,1} + w_{\mathrm{EE}}\, g_{\mathrm{E}}(h_{E,1}) + w_{\mathrm{EI}}\, g_{\mathrm{inh}}(h_{\mathrm{inh}}) + R I_1\,, \tag{16.2}$$

$$\tau_{\mathrm{E}} \frac{\mathrm{d}h_{E,2}}{\mathrm{d}t} = -h_{E,2} + w_{\mathrm{EE}}\, g_{\mathrm{E}}(h_{E,2}) + w_{\mathrm{EI}}\, g_{\mathrm{inh}}(h_{\mathrm{inh}}) + R I_2\,, \tag{16.3}$$

$$\tau_{\mathrm{inh}} \frac{\mathrm{d}h_{\mathrm{inh}}}{\mathrm{d}t} = -h_{\mathrm{inh}} + w_{\mathrm{IE}}\, g_{\mathrm{E}}(h_{E,1}) + w_{\mathrm{IE}}\, g_{\mathrm{E}}(h_{E,2})\,; \tag{16.4}$$

see Eqs. (15.3) and (15.1). Here w_{EE} denotes the strength of recurrent coupling within each of the excitatory populations and w_{EI} the coupling from the inhibitory to the excitatory population of neurons. Inhibitory neurons are driven by the input from excitatory populations via connections of strength w_{IE}. We assume that inhibitory neurons have no self-coupling, but feed their activity A_{inh} back to both excitatory populations with a negative coupling coefficient, $w_{\mathrm{EI}} < 0$. Note that the two excitatory populations are completely equivalent, i.e., they contain neurons of the same type and the same coupling strength. However, the two populations receive separate inputs, I_1 and I_2, respectively. We call an input "biased" (i.e., favoring one of the two options represented by the excitatory populations) if $I_1 \neq I_2$. We emphasize that the only interaction between the two excitatory populations is indirect via the shared inhibitory population.

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

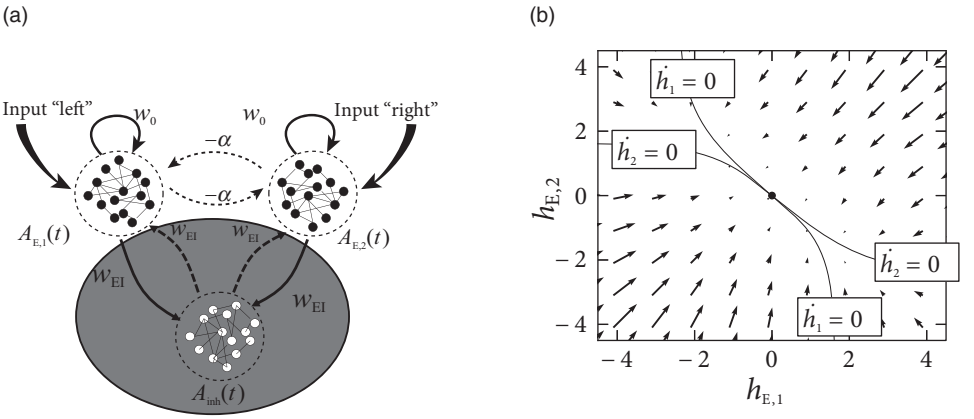

Fig. 16.6 Effective inhibition. (a) Two populations of excitatory neurons interact with a common pool of inhibitory neurons. The inhibitory population is replaced by an effective inhibitory coupling of strength α between the two excitatory populations. In addition, both populations of excitatory neurons make excitatory connections of strength $w_0 = w_{EE} - \alpha$ onto itself. (b) Phase plane analysis in the absence of stimulation. The nullclines $dh_{E,1}/dt = 0$ and $dh_{E,2}/dt = 0$ are shown as a function of $h_{E,1}$ (horizontal axis) and $h_{E,2}$ (vertical axis). There is a single crossing point corresponding to a stable fixed point close to $h_{E,1} = h_{E,2} = 0$. Arrows indicate the flow toward the fixed point.

16.3.2 Effective inhibition

The system of three differential equations (16.2)–(16.4) is still relatively complicated. However, from Chapter 4 we know that for a two-dimensional system of equations we can use the powerful mathematical tools of phase plane analysis. This is the main reason why we now reduce the three equations to two.

To do so, we make two assumptions. First, we assume that the membrane time constant of inhibition is shorter than that of excitation, $\tau_{inh} \ll \tau_E$. Formally, we consider the limit of a separation of time scales $\tau_{inh}/\tau_E \to 0$. Therefore we can treat the dynamics of h_{inh} in Eq. (16.4) as instantaneous, so that the inhibitory potential is always at its fixed point

$$h_{inh} = w_{IE}\left[g_E(h_{E,1}) + g_E(h_{E,2})\right]. \tag{16.5}$$

Is this assumption justified? Inhibitory neurons do indeed fire at higher firing rates than excitatory ones and are in this sense "faster." However, this observation on its own does not imply that the membrane time constants of excitatory and inhibitory neurons, respectively, would differ by a factor of 10 or more; in fact, they don't. Nevertheless, a focus on the raw membrane time constant is also too limited in scope, since we should also take into account synaptic processes. Excitatory synapses typically have an NMDA component with time constants in the range of a hundred milliseconds or more, whereas inhibitory synapses are fast. We recall from Chapter 15 that the rate equations that we use here are in any case highly simplified and do not fully reflect the potentially much richer dynamics of neuronal populations.

Intuitively, the assumption of a separation of time scales implies that inhibition reacts faster to a change in the input than excitation. In the following we simply assume the

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

separation of time scales between inhibition and excitation, because it enables a significant simplification of the mathematical treatment. Essentially, it means that the variable $h_{\rm inh}$ can be removed from the system of three equations (16.2)–(16.4). Thus we drop Eq. (16.4) and replace in Eqs. (16.2) and (16.3) the input potential $h_{\rm inh}$ by the right-hand side of Eq. (16.5).

The second assumption is not absolutely necessary, but it makes the remaining two equations more transparent. The assumption concerns the shape of the gain function of inhibitory neurons. We require a linear gain function and set

$$g_{\rm inh}(h_{\rm inh}) = \gamma h_{\rm inh}\,, \tag{16.6}$$

with a slope factor $\gamma > 0$. If we insert Eqs. (16.5) and (16.6) into (16.2) and (16.3) we arrive at

$$\tau_{\rm E}\,\frac{{\rm d}h_{E,1}}{{\rm d}t} = -\,h_{E,1} + (w_{\rm EE} - \alpha)\,g_{\rm E}(h_{E,1}) - \alpha\,g_{\rm E}(h_{E,2}) + RI_1\,, \tag{16.7}$$

$$\tau_{\rm E}\,\frac{{\rm d}h_{E,2}}{{\rm d}t} = -\,h_{E,2} + (w_{\rm EE} - \alpha)\,g_{\rm E}(h_{E,2}) - \alpha\,g_{\rm E}(h_{E,1}) + RI_2\,, \tag{16.8}$$

where we have introduced a parameter $\alpha = -\gamma w_{\rm EI} w_{\rm IE} > 0$. Thus, the model of three populations has been replaced by a model with two excitatory populations that interact with an *effective inhibitory* coupling of strength α; see Fig. 16.6a. Even though neurons make either excitatory or inhibitory synapses, never both ("Dale's law"), the above derivation shows that, under appropriate assumptions, there is a mathematically *equivalent* description where explicit inhibition by inhibitory neurons is replaced by *effective inhibition* between excitatory neurons. The effective inhibitory coupling allows us to discuss competition between neuronal groups in a transparent manner.

16.3.3 Phase plane analysis

The advantage of the reduced system with two differential equations (16.7) and (16.8) and effective inhibition is that it can be studied using phase plane analysis; see Figs. 16.6b and 16.7.

In the absence of stimulation, there exists only a single fixed point $h_{E,1} = h_{E,2} \approx 0$, corresponding to a small level of spontaneous activity (Fig. 16.6b).

If a stimulus $I_1 > 0$ favors the first population, the fixed point moves to an asymmetric position where population 1 exhibits much stronger activity $A_{E,1} = g(h_{E,1})$ than population 2 (Fig. 16.7a). Note that, at the fixed point, $h_{E,2} \ll 0$. In other words, the effective interaction between the two populations causes a strong inhibitory input potential to population 2. This is a characteristic feature of a competitive network. If one of the populations exhibits a strong activity, it inhibits activity of the others so that only the activity of a single winning population "survives." This principle can also be applied to more than two interacting populations, as we shall see in Section 16.3.4.

A particularly interesting situation arises with a strong but unbiased stimulus, as we have

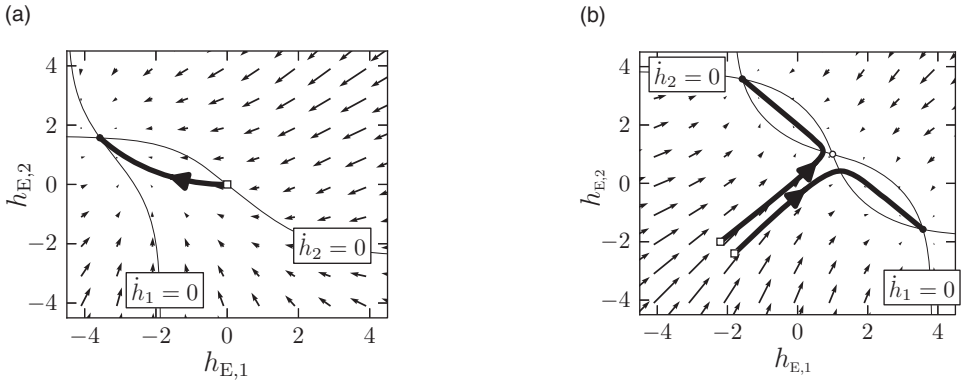

Fig. 16.7 Phase plane analysis of the competition model during stimulation. (a) A strong stimulus $I_1 > 0 = I_2$ gives rise to a single stable fixed point, corresponding to high firing rates $A_{E,1}$ of the first population. This indicates a choice "left." The nullclines $dh_{E,1}/dt = 0$ (solid line) and $dh_{E,2}/dt = 0$ (solid line) are shown as a function of $h_{E,1}$ (horizontal axis) and $h_{E,2}$ (vertical axis). Arrows indicate the flow toward the fixed point. A sample trajectory is indicated (thick line). (b) Phase plane analysis with strong, but ambiguous stimulation $I_1 = I_2 > 0$. The symmetric fixed point is unstable, and the flow converges to one of the two stable fixed points. Two sample trajectories are shown.

already seen in the simulations of Fig. 16.5. The phase plane analysis of Fig. 16.7b shows that, with a strong unbiased stimulus $I_1 = I_2 \gg 0$, three fixed points exist. The symmetric fixed point $h_{E,1} = h_{E,2}$ is a saddle point and therefore unstable. The two other fixed points occur at equivalent positions symmetrically to the left and right of the diagonal. These are the fixed points that enforce a decision "left" or "right."

It depends on the initial conditions, or on tiny fluctuations in the noise of the input, whether the system ends up in the left or right fixed point. If, before the onset of the unbiased strong stimulation, the system was at the stable resting point close to $h_{E,1} = h_{E,2} \approx 0$, then the dynamics is first attracted toward the saddle point, before it bends over to either the left or right stable fixed point (Fig. 16.7b). Thus, the phase plane analysis of the two-dimensional system correctly reflects the dynamics observed in the simulations of the model with populations of hundreds of spiking neurons (Fig. 16.5b).

16.3.4 Formal winner-take-all networks

The arguments that were developed above for the case of a binary choice between two options can be generalized to a situation with K possible outcomes. Each outcome is represented by one population of excitatory neurons. Analogous to the arguments in Fig. 16.6a, we work with an effective inhibition of strength $\alpha > 0$ between the K pools of neurons and with a self-interaction of strength w_0 within each pool of neurons.

The activity of population k is then

$$A_k(t) = g(h_k(t)) \tag{16.9}$$

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

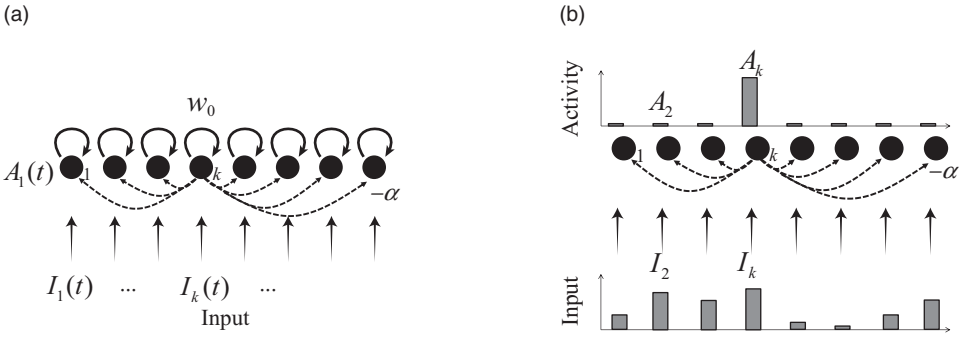

Fig. 16.8 Formal winner-take-all network. (a) Network architecture: each artificial neuron $1 \leq k \leq K$ receives an input I_k. Each neuron has a positive feedback of magnitude w_0 onto itself but inhibits with strength α all other neurons. (b) In a pattern of fixed inputs $I_k > 0$, $1 \leq k \leq K$ switched on at time t_0, the network converges to a state where only a single "winner" neuron is active, i.e., the one which receives the strongest input.

with input potential

$$\tau \frac{\mathrm{d}h_k}{\mathrm{d}t} = -h_k + w_0\, g(h_k) - \alpha \sum_{j \neq k} g_{\mathrm{E}}(h_j) + R I_k, \tag{16.10}$$

where the sum runs over all neurons $1 \leq j \leq K$, except neuron k. Note that we assume here a network of interacting *populations*, but it is common to draw the network as an interaction between formal units. Despite the fact that, in our interpretation, each unit represents a whole population, the units are often called "artificial neurons;" see Fig. 16.8a. Winner-take-all networks are a standard topic of artificial neural networks (Hertz *et al.*, 1991; Kohonen, 1984; Haykin, 1994).

For a suitable choice of coupling parameters w_0 and α the network implements a competition between artificial neurons, as highlighted in the following example.

Example: Competition

Consider a network of formal neurons described by activities $A_k = [1 + \tanh(h - \theta)]\, A_{\max}/2$. We work in unit-free variables and set $A_{\max} = 1$ and $\theta = 5$. Thus, for an input potential $h = 0$ the activity is nearly zero while for $h = 10$ it is close to 1. The input potential, given by Eq. (16.10), contains contributions from external input as well as contributions from recurrent interactions within the network.

Suppose that for all times $t < t_0$ the external input vanishes, $I_k = 0$ for all k. Thus, at time t_0 the input potential h_k and the activity A_k are negligible for all units k. Therefore the interactions within the network are negligible as well.

At time t_0 the input is switched on to a new fixed value I_k which is different for each neuron; see Fig. 16.8b. The activity of the neuron k which receives the strongest input grows more rapidly than that of the others so that its activity also increases more rapidly.

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

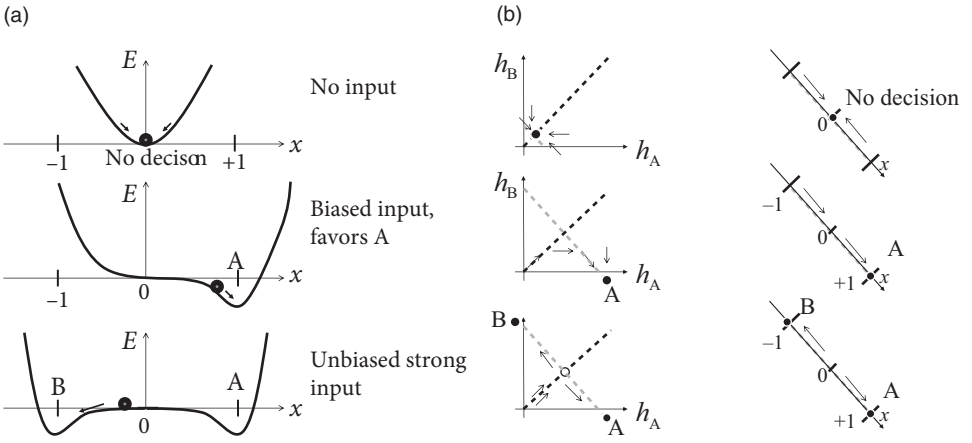

Fig. 16.9 Energy picture of decision making. (a) Decisions correspond to a ball rolling down an energy landscape, plotted as a function of a formal decision variable x. A value of $x = 0$ indicates that no decision is taken (top, no input), whereas a value of $x = \pm 1$ reflects a decision for options A or B, respectively. An input favoring option A deforms and tilts the energy landscape so that the minimum of the energy is in the neighborhood of $x = 1$ (middle). A strong but unbiased input creates two energy minima corresponding to options A and B. Only one of the two options can be taken by the rolling ball (bottom). (b) The corresponding phase plane diagram of the input potentials h_A (horizontal axis) and h_B (vertical axis). Fixed points are indicated as solid (stable) or empty (saddle) circles. The decision variable x moves along the axis (dashed gray line) perpendicular to the diagonal (dashed black line) and is replotted again as a one-dimensional flow on the right-hand side of the figure.

The strong activity of neuron k inhibits the development of activity in the other neurons so that, in the end, the neuron with the strongest input wins the competition and its activity is the only one to survive.

16.4 Alternative decision models

In the previous two sections, we discussed decision models as the competitive interaction between two or more populations of excitatory neurons. In this section we present two different models of decision making, i.e., the energy picture in Section 16.4.1 and the drift-diffusion model in Section 16.4.2. Both models are phenomenological concepts to describe decision making. However, both models are also related to the phase diagram of the model of two neuronal populations, encountered in Section 16.3.

16.4.1 The energy picture

Binary decisions can be visualized as a ball in a hilly energy landscape. Once the ball rolls in a certain direction, a decision starts to form. The decision is finalized when the ball

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

approaches the energy minimum; see Fig. 16.9a. The decision variable x reflects a decision for option A if it approaches a fixed point in the neighborhood of $x \approx 1$, and a decision for option B for $x \approx -1$. If the variable x is trapped in a minimum close to $x = 0$, no decision is taken.

The dynamics of the decision process can be formulated as gradient descent

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\eta \frac{\mathrm{d}E}{\mathrm{d}x} \tag{16.11}$$

with a positive constant η. In other words, in a short time step Δt, the decision variable moves by an amount $-\eta \Delta t \, \mathrm{d}E/\mathrm{d}x$. Thus, if the slope is positive, the movement is toward the left. As a result, the movement is always downhill, so that the energy decreases along the trajectory $x(t)$. We can calculate the change of the energy along the trajectory:

$$\frac{\mathrm{d}E(x(t))}{\mathrm{d}t} = \frac{\mathrm{d}E}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} = -\eta \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)^2 \leq 0\,. \tag{16.12}$$

Therefore the energy plays the role of a Liapunov function of the system, i.e., a quantity that cannot increase along the trajectory of a dynamical system .

Interestingly, the energy picture can be related to the phase plane analysis of the two-dimensional model that we encountered earlier in Figs. 16.6b and 16.7. The diagonal of the phase plane plays the role of the boundary between the options A and B while the variable x indicates the projection onto an axis orthogonal to the diagonal. Position $x = 0$ is the unbiased, undecided position on the diagonal; see Fig. 16.9. In the case of strong unbiased input, the one-dimensional flow diagram of the variable x presents a reasonable summary of the flow pattern in the two-dimensional system, because the saddle point in the phase plane is attractive along the diagonal and is reached rapidly while the flow in the perpendicular direction is much slower (Wang *et al*., 2002; Bogacz *et al*., 2006; Wong and Wang, 2006).

The above arguments regarding the Liapunov function of the network can be made more precise and formulated as a general theorem (Cohen and Grossberg, 1983; Hopfield, 1984). We consider an arbitrary network of K neuronal populations $1 \leq j \leq K$ with population rate $A_j = g(h_j) \geq 0$ where g is a gain function with derivative $g' > 0$ and h follows the dynamics

$$\tau \frac{\mathrm{d}h_j}{\mathrm{d}t} = -h_j + RI_j + \sum_k w_{jk} g(h_k) \tag{16.13}$$

with fixed inputs I_j. If the coupling is *symmetric*, i.e., $w_{ij} = w_{ji}$, then the energy

$$E = -\sum_i \sum_j w_{ij} A_i A_j - \sum_i A_i R I_i + \sum_i \int_0^{A_i} g^{-1}(a)\,\mathrm{d}a \tag{16.14}$$

is a Liapunov function of the dynamics.

The proof follows by taking the derivative. We exploit the fact that $w_{ij} = w_{ji}$ and apply

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

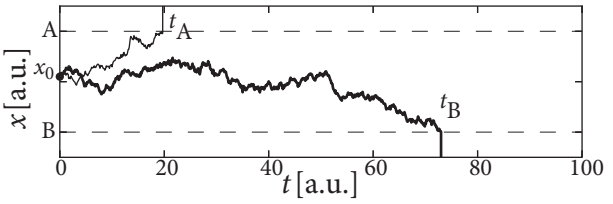

Fig. 16.10 Drift-diffusion model. The decision variable $x(t)$ starts from x_0 and undergoes a biased random walk toward one of the two thresholds, marked by horizontal dashed lines. In the first trial (thin solid line), the choice is option A and the reaction time indicated as t_A. In another trial (thick solid line) the trajectory hits, after a time t_B the threshold for option B.

the chain rule $dA_i/dt = g'(h_i)dh_i/dt$ so as to find

$$\begin{aligned}\frac{dE}{dt} &= -\sum_i \left[\sum_j w_{ij}A_j\right] g'(h_i)\frac{dh_i}{dt} - \sum_i R I_i g'(h_i)\frac{dh_i}{dt} + \sum_i g^{-1}(A_i)\, g'(h_i)\frac{dh_i}{dt} \\ &= -\tau \sum_i g'(h_i)\left[\frac{dh_i}{dt}\right]^2 \le 0\,. \end{aligned} \tag{16.15}$$

In the second line we have used Eq. (16.13). Furthermore, since the neuronal gain function stays below a biologically sustainable firing rate $g(x) \le A^{\max}$, the energy is bounded from below. Therefore the flow of a symmetric network of interacting populations will always converge to one of the stable fixed points corresponding to an energy minimum, unless the initial condition is chosen to lie on an unstable fixed point, in which case the dynamics stays there until it is perturbed by some input.

Example: Binary decision network revisited

The binary decision network of Eqs. (16.7) and (16.8) with effective inhibition α and recurrent interactions w_0 consists of two populations. Interactions are symmetric since $w_{12} = w_{21} = -\alpha$. Therefore the energy function

$$E = -w_0[A_1^2 + A_2^2] + 2\alpha A_1 A_2 - R\,[I_1 A_1 + I_2 A_2] + \int_0^{A_1} g^{-1}(x)dx + \int_0^{A_2} g^{-1}(y)dy \tag{16.16}$$

is a Liapunov function of the dynamics defined in Eqs. (16.7) and (16.8). Since the dynamics is bounded, there must be stable fixed points.

16.4.2 Drift-diffusion model

The drift-diffusion model is a phenomenological model to describe choice preferences and distributions of reaction times in binary decision making tasks (Ratcliff and Rouder, 1998). At each trial of a decision experiment, a decision variable x is initialized at time t_0 at a value

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

$x(t_0) = x_0$. Thereafter, the decision variable evolves according to

$$\frac{dx}{dt} = (I_A - I_B) + \sigma\xi(t) \tag{16.17}$$

where $\xi(t)$ is Gaussian white noise of unit mean and variance σ^2. An input $I_A > I_B$ causes a "drift" of the variable x toward positive values while the noise ξ leads to a "diffusion"-like motion of the trajectory; hence the name "drift-diffusion model."

The reaction time is the time at which the variable x reaches one of two thresholds, Θ_A or Θ_B, respectively (Fig. 16.10). For example t_B, defined by $t_B = \min\{t|x(t) = \Theta_B\}$, is the reaction time in a trial where the choice falls on option B.

Parameters of the phenomenological drift-diffusion model are the values of thresholds Θ_A and Θ_B, and the strength of the input $I_A - I_B$ compared to that of the noise. The initial condition x_0 can be identical in all trials or chosen in each trial independently from a small interval that reflects uncontrolled variations in the bias of the subject. The time t_0 is typically the moment when the subject receives the choice stimulus, but it is also possible to start the drift-diffusion process a few milliseconds later so as to account for the propagation delay from the sensors to the brain (Ratcliff and Rouder, 1998; Ratcliff and McKoon, 2008).

Example: Drift-diffusion model versus neuronal models

In the original formulation, the drift-diffusion model was used as a "black box," i.e., a phenomenological model with parameters that can be fitted to match the distribution of reaction times and choice preferences to behavioral experiments. Interestingly, however, variants of one-dimensional drift-diffusion models can be derived from the models of neural populations with competitive interaction that we have discussed in earlier sections of this chapter (Bogacz *et al.*, 2006; Wong and Wang, 2006; Roxin and Ledberg, 2008). The essential idea can be best explained in the energy picture; see Fig. 16.9. We assume a small amount of noise. In the absence of input, the decision variable jitters around the stable fixed point $x = 0$. Its momentary value $x \approx 0$ serves as an initial condition, once the input is switched on. Suppose the input is strong but unbiased. Two new valleys form around $x \approx \pm 1$. However, in the neighborhood of $x = 0$ the landscape is flat, so that noise leads to a diffusive motion of the trajectory. A biased input $I_A > I_B$ tilts the energy landscape to the right which causes a corresponding drift term in the diffusion process. The location where the slope of the valley becomes steep can be associated with the threshold Θ_A in the diffusion model.

16.5 Human decisions, determinism, and free will

In the previous section, we compared the process of decision making to a ball in an energy landscape. However, as outlined in the introduction to this chapter, decision

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

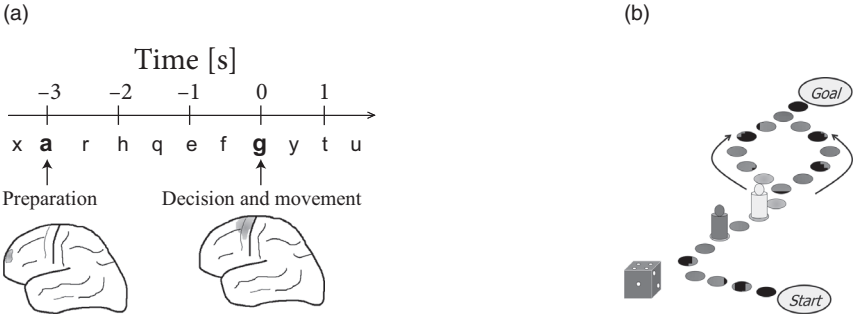

Fig. 16.11 Decision processes and the notion of free will. (a) In a modern variant of the Libet experiment (Libet, 1985), a subject lies in the fMRI-scanner while watching a rapid sequence of letters (Soon *et al*., 2008) that appear at a rate of 2 Hz (top; horizontal axis shows time in seconds). The subject spontaneously decides to move his left or right index finger and reports the letter he saw at the moment when he felt the "urge to move" interpreted as the moment when he took a decision (right; letter "g" defines time zero). However, already a few seconds earlier (e.g., when the letter "a" was shown, left) the brain activity in frontal areas of the subject has a high correlation with his final decision; schematic figure, adapted from Soon *et al*. (2008) and Haggard (2008). (b) The decision to move the left or right index finger is completely irrelevant, similar to a game where a decision between two equivalent choices has to be made.

making incorporates a broad set of phenomena and processes (Rangel *et al*., 2008; Sanfey and Chang, 2008). Here we sketch a link from the simplified model from decision making to the bigger picture.

Adult humans in a state of normal health feel that they are in control of their actions: "The street is too busy, therefore I decide to take the safer underground pathway"; "Because there have been many accidents at this crossing, I decide to break early and be particularly careful"; "I would rather prepare for the exams than go to the movies." We all know examples of consciously controlling our decisions and actions.

Voluntary control of actions can be understood in opposition to pure reflexes (Haggard, 2008). If the doctor hits the right spot on your knee, your foot moves without you intending it. If an object approaches your eyes from the front, you automatically move your head. There are also cases where reflexes have been learned from experience. For example, during your first driving lessons you had to consciously control your foot in order to step on the brakes when a red traffic light appeared in front of your car. After years of experience, you start to break even before you become aware of a conscious decision. Similarly, if you are a good tennis player you will respond to a serve with an arm movement that was trained so often that it has become as fast as a reflex. Nevertheless, you could decide to take back control and try to inhibit your automatic response, if for some reason you want to disturb your opponent. The feeling of voluntary control is what makes you feel responsible for the things you do.

The movement of our arms and legs is controlled by muscles which in turn receive action potentials from the brain via the spinal cord. The human cortex contains several areas

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

that are involved in voluntary actions. The question of where and how the brain controls our decisions and represents our will has triggered the research field of "neuroscience of volition" (Haggard, 2008).

16.5.1 The Libet experiment

The classic experiment in the research field of human volition was performed by Libet (1985). In this experiment, subjects decide on their own when to move their right hand. After each trial, subjects report when they felt the "urge to move," with respect to a rapidly rotating hand of a clock. The reported "urge to move" is in fact about 200 ms earlier than the actual movement. Most interestingly, however, electrical brain activity measured by EEG recordings indicates that the brain exhibits signals of preparatory activity several hundred milliseconds *before* the reported "urge to move." Thus, if we agree to interpret the felt "urge to move" as the conscious decision to move the hand, then we must also accept the fact that the brain has unconsciously prepared our decision.

A modern and debated variant of the Libet experiment is shown in Fig. 16.11a. The main difference to the original Libet experiment (where the decision was limited to "move" or "not move") is that subjects now hold two buttons, one in the left and the other in the right hand (Soon *et al*., 2008). Subjects are free to decide when to move and press either of the two buttons. While subjects perform the experiment, they watch a stream of letters at a rate of two letters per second. At the end of each trial, they indicate at which letter they had felt the "urge to move." The reported letter serves as a timing reference for the subsequent analysis.

During the experiment, brain activity was recorded through functional magnetic resonance imaging (fMRI). Using statistical pattern classification techniques, the authors aimed at predicting the final response outcome (left or right) based on the activity patterns in localized brain areas. If brain activity contained no cue about the final decision, the prediction would always be 50%. However, the authors found that activity patterns in fronto-polar cortex 5 seconds before the reported "urge to move" allowed them to predict the final choice (left or right) with a precision of 55–60% (Soon *et al*., 2008) which is above chance but far from a reliable prediction.

16.5.2 Relevant and irrelevant decisions – a critique

What, if anything, can we learn about decision making and volition from these and similar experiments? In a naive interpretation, the results seem to suggest that the brain has taken its own decision a long time before the subject becomes aware of it. As Wolfgang Prinz puts it: "We don't do what we want, but we want what we do" (Prinz, 2004).

There is little doubt that our actions, plans, and wishes are represented in the brain. Our childhood memories are stored in our brain; our knowledge of the world is memorized in the brain; our values and priorities acquired through education, reading, understanding,

trial and error, or simply through being embedded in our culture, must also be stored in the brain. Thus, a large fraction, if not all, of what we consider our conscious personality is located in the brain.

Most actions where we care about our decision are relevant choices. The decision of whether to take the risky shortcut across a busy street or the safer underground pathway depends on what we have experienced in the past. Similarly, the decision in the board game of Fig. 16.1a depends on the player's attitude toward risk, which has been formed by previous experiences in similar situations. However, the decision task in the scientific experiment of Libet (1985) or Soon *et al*. (2008) is a completely irrelevant one. Subjects don't really care whether they move the left or right finger. The decision has nothing to do with life-long experience or attitude toward risk. In cases like this one, any decision is arbitrary and therefore easily influenced by noise. Think of the board game of Fig. 16.1a and compare it with the situation in Fig. 16.11b. While the first one asks for a decision between a risky and a safe path, the second one poses an irrelevant choice. In the latter case, we might, just for the sake of advancing the game, decide to go left, based on the whim of the moment, but we know that right would do just as well.

Interestingly, even in the irrelevant situation of the experiment of Soon *et al*. (2008), the predictive power of the brain activity five seconds before the conscious "urge to move" is only in the range of 60%. Moreover, in a different experimental design, the subject could "veto" at the last moment a previously prepared movement suggesting the possibility of voluntary inhibition of actions that are only weakly predicted (Brass and Haggard, 2007). Finally, there is also the problem of whether we can really identify a reported "urge to move" with a precise moment of decision. If we take the picture of the ball in the energy landscape, the ball starts to roll in a certain direction while still remaining in the flat region. But this does not yet imply a final decision, because novel input could tilt the energy landscape in the opposite direction.

16.6 Summary

Decisions are prepared and made in the brain so that numerous physiological correlates of decision making can be found in the human and monkey cortex. The fields of cognitive neuroscience associated with these questions are called "neuroeconomics" and "neuroscience of volition."

An influential computational model describes decision making as the competition of several populations of excitatory neurons which share a common pool of inhibitory neurons. Under suitable conditions, the explicit model of inhibitory neurons can be replaced by an effective inhibitory coupling between excitatory populations. In a rate model, the competitive interactions between two excitatory populations can be understood using phase plane analysis. Equivalently, the decision process can be described as downward motion in an energy landscape which plays the role of a Liapunov function. The energy picture is valid for any rate model where all units of the network are coupled by symmetric interactions.

The drift-diffusion model, which has been used in the past as a black-box model for

reaction time distributions and choice preferences, can, under appropriate assumptions, be related to a rate model of competitively interacting populations of neurons.

Literature

There are several accessible introductions to the problem of decision making in neuroeconomics (Platt and Huettel, 2008; Rangel *et al*., 2008; Glimcher *et al*., 2008). The neurophysiological correlates of decision making are reviewed in Gold and Shadlen (2007), Romo and Salinas (2003) and Deco *et al*. (2009, 2010).

The competitive model of decision making that we presented in this chapter is discussed in Y. Wang *et al*. (2002) and Wong and Wang (2006), but competitive interaction through inhibition is a much older topic in the field of computational neuroscience and artificial neural networks (Grossberg, 1969; Kohonen, 1984; Hertz *et al*., 1991; Haykin, 1994). Competitive models of spiking neurons with shared inhibition have also been applied to other tasks of perceptual decision making, (see e.g., Machens *et al*., 2005).

Energy as a Liapunov function for rate models of neurons has been introduced by Cohen and Grossberg (1983). In the context of associative memories (to be discussed in the next chapter) energy functions have been used for binary neuron models by Hopfield (1982) and for rate models by Hopfield (1984).

Drift-diffusion models have been reviewed by Ratcliff and Rouder (1998) and Ratcliff and McKoon (2008). The relation of drift-diffusion models to neuronal decision models has been discussed by Bogacz *et al*. (2006) and Wong and Wang (2006) and has been worked out in the general case by Roxin and Ledberg (2008).

A highly recommended overview of neuroscience around the questions of volition is given by Haggard (2008), who reviews both the original Libet experiment (Libet, 1985) and its modern variants. The fMRI study of Soon *et al*. (2008) is also accessible to the non-specialized reader.

Exercises

1. **Phase plane analysis of a binary decision process**. *Consider the following system (in unit-free variables)*

$$\frac{\mathrm{d}h_{E,1}}{\mathrm{d}t} = -h_{E,1} + (w_{\mathrm{EE}} - \alpha)\, g_{\mathrm{E}}(h_{E,1}) - \alpha\, g_{\mathrm{E}}(h_{E,2}) + h_1^{\mathrm{ext}}, \tag{16.18}$$

$$\frac{\mathrm{d}h_{E,2}}{\mathrm{d}t} = -h_{E,2} + (w_{\mathrm{EE}} - \alpha)\, g_{\mathrm{E}}(h_{E,2}) - \alpha\, g_{\mathrm{E}}(h_{E,1}) + h_2^{\mathrm{ext}}, \tag{16.19}$$

where $\alpha = 1$ *and* $w_{\mathrm{EE}} = 1.5$. *The function* $g(h)$ *is piecewise linear:* $g(h) = 0$ *for* $h < -0.2$*;* $g(h) = 0.1 + 0.5h$ *for* $-0.2 \le h \le 0.2$*;* $g(h) = h$ *for* $0.2 < h < 0.8$ $g(h) = 0.4 + 0.5h$ *for* $0.8 \le h \le 1.2$*; and* $g(h) = 1$ *for* $h > 1.2$.
(a) Draw the two nullclines ($\mathrm{d}h_1/\mathrm{d}t = 0$ *and* $\mathrm{d}h_2/\mathrm{d}t = 0$*) in the phase plane with horizontal axis* h_1 *and vertical axis* h_2 *for the case* $h_1^{\mathrm{ext}} = h_2^{\mathrm{ext}} = 0.8$.
(b) Add flow arrows on the nullclines.
(c) Set $h_1^{\mathrm{ext}} = h_2^{\mathrm{ext}} = b$ *and study the fixed point on the diagonal* $h_1 = h_2 = h^*$. *Find an expression*

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press

for $h^*(b)$ under the assumption that the fixed point is in the region where $g(h) = h$. Analyze the stability of this fixed point.
(d) We now drop the assumption that the fixed point is in the region where $g(h) = h$. Consider an arbitrary sufficiently smooth function $g(h)$ as well as arbitrary couplings α and w_{EE}, and give a formula for the fixed point on the diagonal.
(e) Assume now that $\alpha = 0.75$ and $w_{\mathrm{EE}} = 1.5$. Linearize about the fixed point in (d) and calculate the two eigenvalues.
Hint: Introduce a parameter $\beta = 0.75 g'(h^*)$.
(f) Show that the fixed point is stable for $g'(h^*) = 0$ and unstable for $g'(h^*) = 1$. At which value of β does it change stability?
(g) Describe in words your findings. What happens with a weak or a strong unbiased input to the decision model?

2. **Winner-take-all in artificial neural networks**. *Consider a network of formal neurons described by activities $A_k = (h_k - 1)$ for $1 \le h_k \le 2$, $A_k = 0$ for $h_k < 1$, and $A_k = 1$ for $h_k > 2$. We write $A_k = g_{\mathrm{E}}(h_k)$.*
The update happens in discrete time according to

$$h_k(t + \Delta t) = w_0\, g(h_k(t)) - \alpha \sum_{j \neq k} g_{\mathrm{E}}(h_j(t)) + h_k^{\mathrm{ext}}(t)\,. \tag{16.20}$$

The external input vanishes for $t \le 0$. For $t > 0$ the input to unit k is $h_k^{\mathrm{ext}} = (0.5)^k + 1.0$.
(a) Set $w_0 = 2$ and $\alpha = 1$. Follow the evolution of the activities for three time steps.
(b) What happens if you change α? What happens if you keep $\alpha = 1$ but decrease w_0?
(c) Derive sufficient conditions so that the only fixed point is $A_k = \delta_{k,1}$, i.e., only the unit with the strongest input is active. Assume that the maximal external input to the maximally excited neuron is $h_k^{\mathrm{ext}} \le 2$.

3. **Energy picture**. *Consider the energy function*

$$E(x) = [1 - (I_A + I_B)]\, x^2 + \frac{1}{4} x^4 + (I_A - I_B)\, x \tag{16.21}$$

where I_A and I_B are inputs in support of options A *and* B*, respectively.*
(a) Draw qualitatively the energy landscape in the absence of input, $I_{\mathrm{A}} = I_{\mathrm{B}} = 0$.
(b) Draw qualitatively the energy landscape for $I_B = 0$ when I_A takes one of the three values $\{0.5, 1.0, 1.5\}$.
(c) Draw the energy landscape for $I_A = I_B = c$ when c varies in the range $[0.5, 1.5]$.
(d) Determine the flow $\Delta x = -\Delta t\, \eta\, \mathrm{d}E/\mathrm{d}x$ for a small positive parameter η for all the relevant cases from (a)–(c).
(e) Compare your results with Fig. 16.9.

https://doi.org/10.1017/CBO9781107447615.020 Published online by Cambridge University Press