5 Message Passing and Neurobiology

> Basically there are two types of animals: animals, and animals that have no brains; they are called plants. They don't need a nervous system because they don't move actively, they don't pull up their roots and run in a forest fire! Anything that moves actively requires a nervous system; otherwise it would come to a quick death.
>
> —Rodolfo Llinas

5.1 Introduction

In chapter 4, we saw the form that variational inference takes for two kinds of generative model. In this chapter, we focus on the *process theories* that arise from these inferential dynamics: theories that explain how the brain may implement variational inference. Central to this implementation of Bayesian belief updating is the notion of Bayesian message passing, which encompasses belief propagation and variational message passing, among other schemes. The idea subtending these schemes is that everything does not directly depend on everything else. Instead, each variable in a generative model depends on relatively few other variables. Similarly, the brain exhibits a sparse connectivity structure, wherein the activity of any neuron depends only on those neurons with which it shares synapses. This chapter focuses on the way we can map the sparse message passing associated with variational inference to the sparse connectivity structure of biological computation.

Let us take a step back from the technical material of chapter 4 and turn our attention to the process theories accompanying Active Inference. It is important to draw a distinction between a principle (i.e., the minimization of free energy) and a process theory about how this principle may be

implemented in a certain kind of system, such as the brain. The latter lets us develop hypotheses that are answerable to empirical data. To address the ways in which Active Inference may manifest in the brain, we equate the message passing we saw at the end of chapter 4 with synaptic communication and the dynamics of gradient descent with neuronal activity. The dual aim of this chapter is to introduce readers with a technical background to the neurobiology of Active Inference and to highlight to biologists the relevance of theory to practical neuroscience. We emphasize that this chapter is not intended as the final word on process theories for Active Inference (Pezzulo, Rigoli, and Friston 2015, 2018; Friston and Buzsaki 2016; Friston and Herreros 2016; Friston, FitzGerald et al. 2017; Friston, Parr et al. 2017; Parr and Friston 2018b; Parr, Markovic et al. 2019); it is simply the interpretation that seems most consistent with currently available evidence. Nor is our aim to endorse a specific process theory but to illustrate how the ideas formulated in chapters 1–4 may be put to work in formulating hypotheses answerable to neurobiological measurements.

This chapter is organized as follows. First, in section 5.2, we consider the role of a cortical microcircuit. This is a functional unit comprising several neural populations connected to one another. The pattern of connectivity is replicated over many cortical regions. We highlight the relationship between this stereotyped circuit and the message passing architectures of figures 4.4 and 4.6—themselves recapitulated over hierarchical levels. In section 5.3 we move to effector systems and the formulation of motor control under Active Inference. This deals with the way in which the motor cortex tunes spinal and brain stem reflex arcs to generate purposeful behavior. Section 5.4 touches on ideas relating to subcortical structures like the thalamus and basal ganglia—which have important roles in planning and decision-making. We then consider, in section 5.5, modulation of synaptic efficacy, including the role of neurotransmitters in precision optimization and of plastic changes in learning. Finally, in section 5.6 we return to the theme of hierarchy and the relationship between decision-making and movement generation.

5.2 Microcircuits and Messages

In chapter 4, we saw that the belief-update equations mediating variational inference may be interpreted in terms of a (neuronal) network. The inference schemes presented—for continuous and categorical models—each give rise

to a stereotyped circuitry whose structure is repeated in hierarchical generative models. Similarly, the architecture of the cerebral cortex has a stereotyped structure (Shipp 2007). The neocortex is divided into six layers (or laminae), numbered from superficial (close the brain's surface) to deep (closer to the subcortical white matter). Each layer is characterized by the presence of specific cell types and patterns of connectivity (Zeki and Shipp 1988, Felleman and Van Essen 1991, Callaway and Wiser 2009); this connectivity is summarized in the schematic of a single cortical column in figure 5.1.

A cortical column in one region of the brain connects to columns in other regions and to subcortical structures. Cortical regions are often depicted in a hierarchy that (loosely speaking) assigns those regions closest to sensory input or motor output to the lowest rungs of the hierarchy. As we move further away from these regions—for example, from primary to secondary visual cortex—we ascend the hierarchy. This notion of hierarchy is licensed by the laminar-specific connectivity structure illustrated in figure 5.1. Ascending projections (i.e., connections) from lower cortical areas or sensory (primary) thalamic nuclei tend to target the spiny stellate cells in layer IV. Lower cortical areas give rise to ascending connections from their superficial pyramidal cells (layers II and III). Descending projections from higher cortical areas target both superficial and deep layers of lower areas, with origins in the deep (layer VI) pyramidal cells. In addition, deep pyramidal cells (of Betz) in layer V project to various other targets, including subcortical nuclei—like the basal ganglia and secondary thalamic nuclei—and spinal motor neurons.

The middle schematic in figure 5.1 shows one possible mapping from the network for predictive coding (figure 4.4) to the laminar anatomy of the cortex (Friston, Parr, and de Vries 2017). This is a little complicated to interpret, but the key points are as follows. The ascending input to layer IV spiny stellate cells is associated with the prediction error for hidden causes ($\tilde{\varepsilon}_v^{(i)}$). The ascending output from layer III superficial pyramidal cells represents the same prediction error for the next layer of the hierarchy ($\tilde{\varepsilon}_v^{(i+1)}$). Descending input represents the prediction ($\tilde{g}^{(i+1)}$) from the higher level, while descending output is the prediction for the lower level ($\tilde{g}^{(i)}$). At the lowest level, we see descending predictions coming from layer V, consistent with the output to spinal motor neurons shown on the left. We will return to this in section 5.3. Recall from chapter 4 that the role of hidden causes is to link together hierarchical levels of a model that operates over multiple different timescales. This contrasts with the hidden states, whose role is

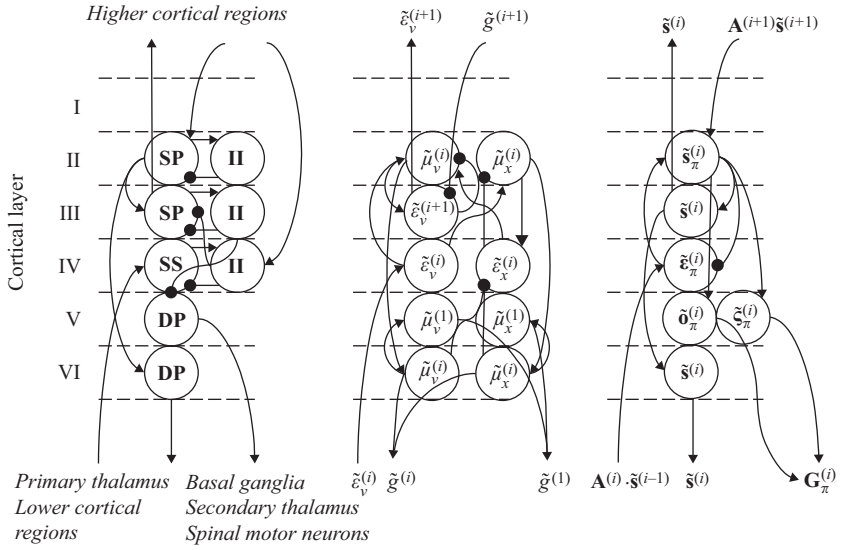


Figure 5.1

Canonical cortical microcircuits illustrating the relationship between inferential message passing and the architecture of the cerebral cortex. *Left:* Simplified schematic based on a synthesis of Miller 2003; Haeusler and Maass 2007; Shipp 2007, 2016; and Bastos et al. 2012 (refer to these papers for a summary of the neuroanatomical observations from which this synthesis is derived). Round arrowheads indicate inhibition; normal arrowheads indicate excitatory connections. The neural populations divide coarsely into superficial pyramidal (SP), deep pyramidal (DP), spiny stellate (SS), and inhibitory interneurons (II). *Middle:* Message passing that underwrites hierarchical predictive coding. *Right:* Message passing needed to solve a partially observable Markov decision process (POMDP).

in the dynamics at a specific timescale—consistent with their role in the intrinsic (within-column) connectivity in figure 5.1.

Asymmetry in message passing is important, as it offers empirical predictions about the difference between ascending and descending activity. One of these predictions is that we might expect these messages to be communicated by neural activity at different temporal frequencies. The reason for this is that the operations required to compute a prediction error from an expectation are nonlinear (Friston 2019b). This nonlinearity is due to the computation of predictions using nonlinear functions (g) that tend to increase the frequency of a signal (e.g., a doubling of frequency on squaring a sine wave). A prediction arising from this is that ascending

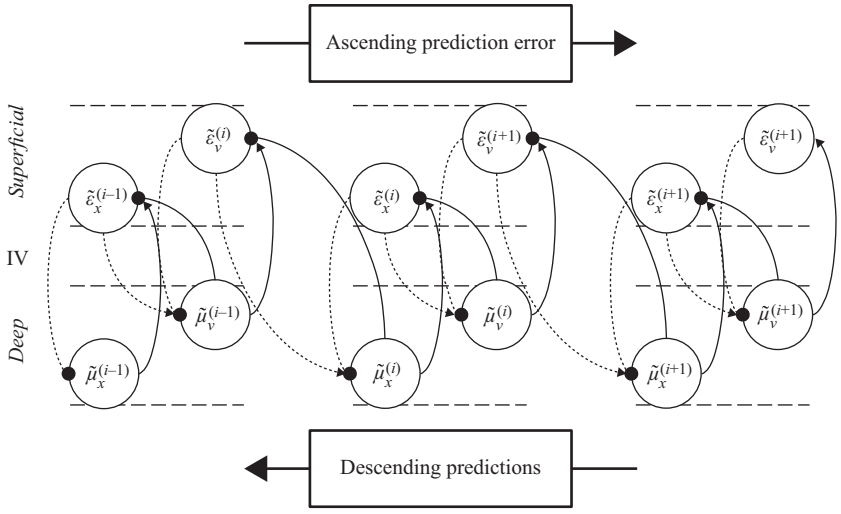


Figure 5.2
Simplified version of the predictive coding scheme shown in the middle of figure 5.1, unpacked to show the message passing between three cortical regions. Emphasis is on the asymmetry in message passing, with dotted lines showing ascending messages (prediction errors) and solid lines showing descending messages (predictions). Figure 5.1 is a finer-grained version of this schematic, including the intermediate neurons in the polysynaptic connections shown here. In this figure, coarse-graining the laminar specificity and dividing the cortex into Superficial and Deep relative to layer IV recovers a predictive coding scheme that will be familiar to many readers.

messages—originating from error units—may be measurable in higher frequency bands than descending messages—originating from expectation units (see figure 5.2). This is consistent with measured spectral asymmetries, wherein ascending connections are typically associated with gamma frequencies and descending connections with alpha or beta bands (Arnal and Giraud 2012; Bastos, Litvak et al. 2015).

The schematic on the right in figure 5.1 shows an interpretation of the message passing for a POMDP model as a cortical microcircuit. This has a structure similar to the predictive coding architecture, with expectations ($\mathbf{s}$) represented in superficial and deep pyramidal cells and propagated up and down cortical hierarchies. In addition, error units ($\boldsymbol{\varepsilon}$) in layer IV are in receipt of ascending signals. In contrast to predictive coding–style architectures, the messages passed between regions are mixtures of expectations as opposed to errors (Friston, Rosch et al. 2017; Parr, Markovic et al. 2019). However, the

overall structure of minimizing an error (i.e., free energy gradient) by updating expectations is preserved. This message passing distinguishes between expectations conditioned on a policy (subscript π) and those averaged under policies. To translate from the former to the latter, we also need posterior beliefs about policies ($\boldsymbol{\pi}$). We will return to this in section 5.4, but for now it is worth highlighting the consistency of this message passing with the targeting of superficial cortical layers by subcortical structures that could compute these beliefs.

In figure 5.1, note the absence of a one-to-one mapping between the architecture on the left and the message passing schemes in the middle and right. For example, there appears to be a discrepancy between the middle and left: the descending input in the column on the left arrives at layers II and IV, but that in the middle graphic targets layer III. This highlights that the connections implied by message passing schemes may not manifest as single synapses. The descending inhibitory connection targeting layer III could be mediated by the combination of an excitatory projection to layer IV inhibitory interneurons, and the projection of these interneurons to layer III. This disynaptic pathway resolves the apparent discrepancy between the two architectures.

In sections 5.3 and 5.4, we deal with the layer V neurons' role in movement and planning, corresponding to their spinal (or brain stem) and subcortical projections, respectively. In sections 5.5 and 5.6, we deal with the ways neural message passing is modulated over time and then return to the relationship between the microcircuits for categorical and continuous inference.

5.3 Motor Commands

The schematic on the left in figure 5.1 shows that layer V of the cortex projects to spinal pyramidal neurons and that this can be interpreted as a prediction (Adams, Shipp, and Friston 2013). This is unpacked in figure 5.3, which shows the spinal components of this circuit; we see a prediction based on the expectations encoded by Betz cells in layer V of the primary motor cortex. This is subtracted from the incoming proprioceptive input to the dorsal horn of the spinal cord, resulting in a proprioceptive prediction error. This error drives muscle activity that results in its own suppression—as proprioceptive data change to comply with predictions. The idea of a motor command as a prediction is central to Active Inference, as it highlights the

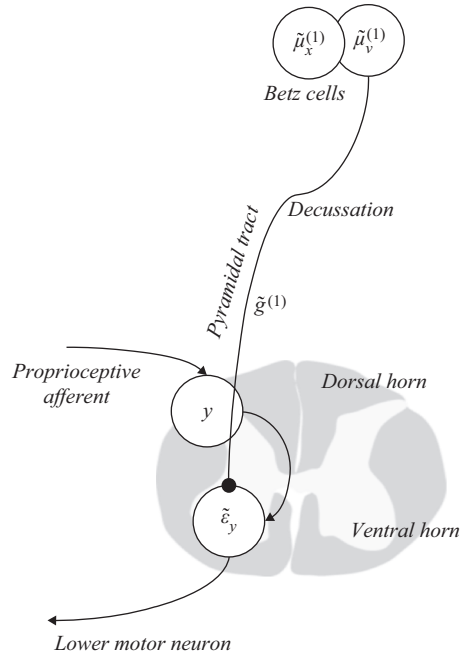


Figure 5.3
Neuroanatomy associated with Active Inference in modulating spinal motor reflexes. Starting from the Betz cells (upper motor neurons) in layer V of the motor cortex, the pyramidal tract carries predictions of proprioceptive input under the movement entailed by motor cortical expectations. The tract decussates (crosses over) and synapses—sometimes polysynaptically—on lower motor neurons in the ventral horn of the spinal cord. Subtracting the predictions from proprioceptive afferent signals arriving at the dorsal horn of the spinal cord results in an error that says how much muscle contraction would be required to meet the prediction. Lower motor neurons then cause this muscle contraction (or relaxation), ensuring that the resulting proprioceptive data match descending predictions.

duality of action and perception. The action part is the minimization of any discrepancy between predictions and sensory data by changing the data. This says that the only thing we should need to generate a movement is a prediction of the sensory consequences anticipated if that movement were to be executed. The fact that proprioceptive prediction errors can always be resolved by reflexes (as opposed to belief updating) offers a possible explanation for the paucity of granular cells in layer IV of the primary motor cortex (Shipp, Adams, and Friston 2013).

An important aspect of this sort of motor control is the notion of sensory attenuation (Brown et al. 2013). Consider the problem of initiating a new movement. To do this, we need to be able to predict that we are moving. However, until we are moving, the proprioceptive data to hand contradicts this hypothesis and could prompt its revision. Therefore, we need a way to preclude sensory data from updating our expectations, so that we can entertain the (initially false) belief that we are moving so that this belief can be realized through action (cf. ideomotor phenomena). The implication is that we need to be able to attend away from proprioceptive data by turning down their gain. Technically, this gain is given by the precision (inverse variance) with which these data are predicted. To attenuate this, we need descending motor tracts to predict not just the data but the precision of—or confidence placed in—those data, decreasing this precision to initiate movement. This is known as sensory attenuation and can be thought of as the complement to sensory attention, equipping us with the capacity to ignore certain prediction errors, such as those generated by saccadic eye movements (here sensory attenuation is known as saccadic suppression). Failures to attenuate are thought to be central to a range of neurological and psychiatric syndromes, including passivity phenomena (Pareés et al. 2014) and—at its most extreme—catatonic states in schizophrenia and the failure to initiate movements in Parkinson's disease.

5.4 Subcortical Structures

In addition to its projections to the spinal cord, cortical layer V targets several other structures. Among these is the striatum (Shipp 2007, Wall et al. 2013)—a structure deep within the cerebrum comprising the caudate nucleus and putamen. The striatum is the input nucleus of a complex network of structures known as the basal ganglia. Medium spiny neurons are the functional units of the striatum, taking input from the cortex and projecting to other nuclei of the basal ganglia. These divide into two types—those that express D1 dopamine receptors and those that express D2 receptors—where dopamine enhances activity of the former and attenuates it for the latter. The former cells are the origin of the direct pathway through the basal ganglia, connected by a single inhibitory synapse to the output nuclei (the internal globus pallidus and substantia nigra pars reticulata). The D2 cells give rise to the indirect pathway, a slightly more complex course with two

inhibitory and one excitatory synapse. The striatum inhibits the external globus pallidus, which itself inhibits the subthalamic nucleus (STN). The STN projects to the basal ganglia outputs, meaning that the output nuclei are inhibited by the direct pathway and disinhibited by the indirect pathway. As these nuclei are themselves inhibitory, the net result of activating D1-expressing striatal neurons is disinhibition of behavior, which would be suppressed by D2-expressing neurons (Freeze et al. 2013).

Given that we have associated proprioceptive predictions with the projections to the spinal cord, which messages should we associate with the layer V projections to the striatum? Inspection of figure 5.1 offers a possible solution. Predicted outcomes ($\mathbf{o}$) and the differences between preferred and predicted outcomes ($\boldsymbol{\varsigma}$) are shown in this layer, combining to compute the expected free energy ($\mathbf{G}$) of a policy. Computing the last of these in the striatum is consistent with the notion that the basal ganglia are involved in planning—that is, evaluating alternative courses of action. Figure 5.4 shows one possible mapping of the message passing for policy evaluation onto the anatomy of the basal ganglia.

The key thing to draw from figure 5.4 is that, as described in chapter 4, posterior probabilities over policies ($\boldsymbol{\pi}$)—shown here in the internal globus pallidus—are computed on the basis of their expected free energy. This pattern follows the direct pathway from layer V of the cortex through the striatum to the basal ganglia output nuclei. However, there are a few additional subtleties. In the hierarchical scheme shown on the left in figure 4.4, we see that expectations about states at higher levels can influence beliefs about policies at lower levels. Figure 5.4 shows this on the left, where the expected observations under high-level states are used to form empirical priors ($\mathbf{E}$) that influence policy selection independently of the expected free energy. We will return to this in chapter 7, but the main idea is that we have beliefs about how we act in particular contexts. These prior expectations tend to bias policy evaluation when we find ourselves in the same context again—much like habit formation. In this sense, the influences of $\mathbf{E}$ and $\mathbf{G}$ can be seen as habitual and goal-directed drives, respectively. In reinforcement learning (Lee et al. 2014), these are sometimes referred to as "model-free" and "model-based" systems.[1] Associating these with the direct and indirect pathways of the basal ganglia has an interesting consequence: it implies that dopamine modulates the balance of the two. Remember that dopamine tends to promote the direct pathway and execution of specific

policies (Moss and Bolam 2008)—presumably those associated with the lowest expected free energy. In contrast, low dopamine might be expected to favor context-sensitive priors in the indirect pathway, whose role is to suppress implausible policies in a given context. In a sense, striatal dopamine can be thought of as modulating the balance between inferring what to do and what not to do (Parr 2020).

The above is consistent with perturbations of the dopaminergic system; its depletion in severe Parkinson's disease causes akinesia—a failure to enact specific policies—while exogenous dopamine agonists promote impulsive behaviors (Frank 2005; Galea et al. 2012; Friston, Schwartenbeck et al. 2014). In addition, it is consistent with conceptual models of basal ganglia function. For example, Nambu (2004) suggests that the direct pathway mediates a fast and focused inhibition of the internal globus pallidus, followed by a broad and slow excitation, which causes excitation and inhibition of the targets of the basal ganglia, respectively. This is thought to ensure a "centre-surround" pattern that facilitates motor programs with a high specificity, which is consistent with the fast processes computing the expected free energy facilitating action and the broader contextualization of the slower pathway communicating empirical priors.

The final observation to make about figure 5.4 is that there are two levels of a cortical hierarchy (superscripted) contributing to the same basal ganglia circuit. This suggests temporally slower regions in targeting of indirect pathway neurons, but both fast and slow influences over the direct pathway. As we ascend cortical hierarchies, neurons tend to represent slower dynamics. For example, we may expect frontal cortical regions to represent longer timescales than parietal regions. This is consistent with the anatomical distribution of cortical inputs to the basal ganglia pathways (Wall et al. 2013). This temporal coarse graining in the indirect pathway is complemented by spatial coarseness, with the direct pathway medium spiny neurons exhibiting larger dendritic arbors (Gertler et al. 2008), enabling finer tuning. Therefore, the anatomy of figure 5.4 is endorsed by evidence from both clinical pathology (e.g., Parkinsonism) and cellular morphology.

In addition to the basal ganglia, many other important subcortical structures contribute to neuronal message passing. In the next section, we will discuss those from which neuromodulatory systems originate; we will conclude this section by briefly touching on the thalamus. We will not be able to do this highly complex structure full justice; however, we can

outline some basic principles. The thalamus is often divided into primary and secondary nuclei. Figure 5.1 shows that primary thalamic nuclei can play the same role as lower cortical regions, in the sense that they target layer IV of the cortex and receive input from layer VI deep pyramidal cells (Thomson 2010, Olsen et al. 2012). An example is the lateral geniculate nucleus in the visual system, often thought of as a relay between the eye and the visual cortex. Like those neurons representing prediction errors, this receives both sensory information from the eye and backward projections from the cortex, which could be construed as predictions. Second order thalamic nuclei include the mediodorsal nucleus and the pulvinar, which interact with frontal and posterior cortices, respectively. These may have a role in predicting second order statistics (i.e., precision and variance) of sensory or higher order inputs and have been associated with figure-ground discrimination tasks (Kanai et al. 2015). Simplistically, this suggests that the division of the thalamus into primary and secondary nuclei may be a manifestation of the division into first and second order statistics.

5.5 Neuromodulation and Learning

Structural neuroanatomy is important, but it only gives us part of the picture of neural processing because the presence of a connection does not tell us much about the way it is used. As an example, consider the role of the substantia nigra shown in figure 5.4. The modulatory effect this has on striatal connectivity underwrites very different outputs from the basal ganglia, depending on the amount of dopamine released. Fast modulation of synaptic efficacy of this form can be contrasted with the slower but more persistent changes that occur with learning. In this section, we focus on these two ways in which synaptic efficacy can change.

Precision is an important concept in understanding neuromodulation (Feldman and Friston 2010). We touched on this in chapter 4 and in our discussion of sensory attenuation above, where we saw that it acts as a multiplicative weight on the prediction errors. More broadly, precision is a measure of confidence in a probability distribution. The relationship between the two is simple. If we have very precise beliefs about how data are generated from hidden states, then our beliefs about hidden states can be updated by observing data more than if we are not confident in those

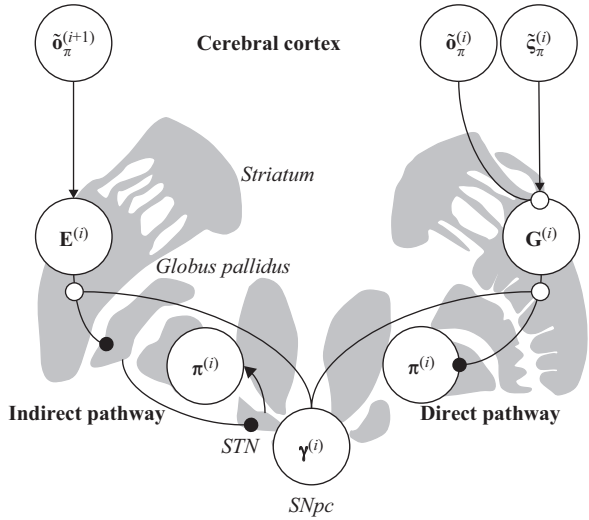


Figure 5.4
Direct and indirect pathways of the message passing for policy selection associated with Active Inference through the basal ganglia using a POMDP generative model. Pathways from the cerebral cortex culminate in estimation of policies. The direct pathway (*right*) goes from cortex to striatum to internal globus pallidus. The indirect pathway (*left*) goes from cortex to striatum to external globus pallidus to subthalamic nucleus (STN) to internal globus pallidus. Both pathways exist bilaterally; in addition, the substantia nigra pars compacta (SNpc) is shown modulating the balance between the two. (Note: This is a simplification of basal ganglia connectivity.)

beliefs. When belief updates manifest as changes in neuronal firing, a more precise likelihood distribution manifests as an increased neural response to a given sensory stimulus. This is essential for cognitive functions from attention (Parr and Friston 2019a) to multisensory integration (Limanowski and Friston 2019).

This perspective on synaptic gain control tells us something simple but important. If precision is an attribute of some distribution in a generative model, then there should be different precisions associated with different distributions. This is intuitively sensible, as we can be more or less confident in the reliability of our sensations, in how things will dynamically evolve, and even in how we might act (Parr and Friston 2017b). We have seen the last of these in the context of dopaminergic signaling in the basal ganglia. The greater the associated precision, the more confident we are that our policies will minimize expected free energy.

If one of the roles of the dopaminergic system—originating in the substantia nigra pars compacta and the ventral tegmental area of the midbrain—is to signal confidence in what to do, then do other neuromodulatory systems play similar roles? Table 5.1 summarizes the evidence associating precisions with neuromodulatory systems—with the symbols sometimes used for these precisions. Specifically, the cholinergic system arising from the basal nucleus of Meynert appears to signal the precision of some likelihood distributions. The noradrenergic system, from the locus coeruleus, seems to play a role in signaling the precision of transitions over time. The serotonergic system seems less clear but may relate to the precision of prior preferences.

Why is it useful to be able to associate these precisions with neuromodulatory systems? The answer is threefold: it lets us explain observed biology, form hypotheses, and develop noninvasive methods to measure precision. We will highlight one example of each of these. First, regarding explanations of observed biology, empirical measurements of dopamine signals famously look like "reward prediction errors" (Schultz 1997)—with animals' dopamine increasing on receiving unexpected fruit juice or seeing a cue signaling imminent fruit juice. Active Inference offers an alternative explanation of these findings (Schwartenbeck, FitzGerald, Mathys, Dolan, and Friston 2015). Achieving a reward (or fulfilling our preferences) or encountering a cue indicative of a future reward enhances our confidence that we are pursuing a policy that minimizes expected free energy. This increase in confidence manifests as a spike in dopamine.

Second, regarding formation of hypotheses, an example concerns the decrease in cholinergic signaling associated with Lewy body dementia (Parr, Benrimoh et al. 2018)—a condition that leads to complex visual hallucinations. One plausible explanation for this is that accumulation of pathology in higher visual cortices prompts a mismatch between the predictions from these areas and the activity in primary visual cortices. Such a mismatch downgrades confidence in the associated likelihood distributions and causes loss of cholinergic signaling. The consequence of this loss of precision is a failure to update beliefs on the basis of sensory data, meaning perception loses the constraints afforded by sensation. This could explain the development of hallucinatory percepts in this condition.

Third, regarding noninvasive measurement of precision parameters, an example is the identification of computational phenotypes. There are a number of peripheral manifestations of central neurochemical activity,

Table 5.1
Putative roles of neurotransmitters in Active Inference

Neurotransmitter	Precision	Evidence
Acetylcholine	Likelihood (ζ)	• Presence of presynaptic receptors on thalamocortical afferents (Sahin et al. 1992, Lavine et al. 1997) • Modulation of gain of visually evoked responses (Gil et al. 1997, Disney et al. 2007) • Changes in effective connectivity with pharmacological manipulations (Moran et al. 2013) • Modeling of behavioral responses under pharmacological manipulation (Vossel et al. 2014, Marshall et al. 2016)
Noradrenaline	Transitions (ω)	• Maintenance of persistent prefrontal (delay-period) activity (requiring precise transition probabilities) depends on noradrenaline (Arnsten and Li 2005, Zhang et al. 2013) • Pupillary responses to surprising (i.e., imprecise) sequences (Nassar et al. 2012, Lavín et al. 2013, Liao et al. 2016, Krishnamurthy et al. 2017, Vincent et al. 2019) • Modeling of behavioral responses under pharmacological manipulation (Marshall et al. 2016)
Dopamine	Policies (γ)	• Expressed postsynaptically on striatal medium spiny neurons (Freund et al. 1984, Yager et al. 2015) • Computational fMRI reveals midbrain activity with changes in precision (Schwartenbeck, FitzGerald, Mathys, Dolan, and Friston 2015) • Modeling of behavioral responses under pharmacological manipulation (Marshall et al. 2016)
Serotonin	Preferences or interoceptive likelihood (χ)	• Receptors expressed on layer V pyramidal cells (Aghajanian and Marek 1999, Lambe et al. 2000, Elliott et al. 2018) in medial prefrontal cortex • Medial prefrontal cortical regions heavily implicated in interoceptive processing and autonomic regulation (Marek et al. 2013, Mukherjee et al. 2016)

Source: Parr and Friston 2018.

including the relationship between spontaneous blink rate and dopamine (Karson 1983) and between pupillary size and noradrenaline (Koss 1986). Recent work exploring the latter (Vincent et al. 2019) has demonstrated a relationship between the transition precision expected to be inferred by an ideal Bayesian observer and the dynamics of pupillary constriction and dilatation. The implication is that we could probe someone's implicit generative model (i.e., empirical prior beliefs) through peripheral measurements of this sort.

While fast changes in precision are important, this is a crude way of optimizing effective connectivity; it leads to an increase or decrease in the gain of a signal, but nothing more subtle. If we want to change the way the signal is interpreted, we need to rely on learning. We will return to this in detail in chapter 7. However, the basic idea is that we hold beliefs not just about states of the world but also about the fixed (or slowly varying) parameters that determine the dependencies between variables (Friston, FitzGerald et al. 2016). The substrate of these beliefs is the efficacy of synaptic connections between the neural populations representing time-varying variables (like hidden states or outcomes). When we observe an outcome that we believe was generated by a given state, we can update beliefs about the parameter connecting the two, reflecting an increase in the probability of them co-occurring in the future. In other words, we get a strengthening of the synapses between the two populations of neurons. The result is Hebb's famous edict (paraphrased): "Cells that fire together, wire together."

An important feature of figure 5.1 is that, in both predictive coding and marginal message passing schemes, the connections entering and leaving a cortical column relate to likelihood distributions. In contrast, transition probabilities and continuous dynamics depend on connections within a microcircuit. This suggests that learning dynamics should lead to changes in intrinsic connectivity, while learning observation models should modify extrinsic connectivity. Using techniques like dynamic causal modeling—which allow for evaluation of effective connectivity measures from neuroimaging data—it is possible to put these hypotheses to the test (Tsvetanov et al. 2016, Zhou et al. 2018). This highlights the role of process theories of this sort: they let us go beyond abstract theorizing to form specific testable hypotheses.

5.6 Continuous and Discrete Hierarchies

Finally, it is worth highlighting the move from continuous representations at low levels of a neural hierarchy to categorical variables at higher levels. The point is that the discrete and continuous message passing schemes we have considered likely coexist in the brain because we are able to hold beliefs of a categorical sort (e.g., in identifying what an object is or who a person is) in addition to being able to interface with continuously varying sensory receptors and effectors (e.g., muscle length or visual luminance contrast). This is reflected in neurophysiology, where some neurons are selective to specific stimuli and others vary in proportion to the intensity of a stimulus.

An interesting observation is that our interface with the world around us is in the continuous domain, the implication of which is that the lowest level of any hierarchy in the brain must be continuous. Having said this, we saw in figure 5.4 that policy selection in the basal ganglia may be framed as a discrete process, selecting between alternative movements. This tells us that we can think of movements as a composition of discrete trajectories into purposeful action. Where the lowest level might deal with the requisite changes in muscle length, descending input is based on decisions about which movement to make. From the perspective of a generative model, this means associating alternative (discrete) hypotheses about the world with the (continuous) dynamics entailed by those hypotheses. In chapter 8, we will return to the question of how to put these together from a computational perspective. Here, we simply note that the further we move from sensory receptors, the more we tend to find discretized representations in neural systems. Indeed, the very existence of classical receptive fields in neurophysiology could be interpreted as a probabilistic representation that the world is in some particular regime of a perceptual state-space—a state-space that is tiled by receptive fields and consequently partitioned into lots of little categories. Figure 5.5 brings together these schemes and acts as a summary of the ideas set out in this chapter.

5.7 Summary

This chapter has sought to outline the points of connection between the message passing schemes implied by the generative models of chapter 4 and the neurobiology of inference, action, and planning. What do we gain by

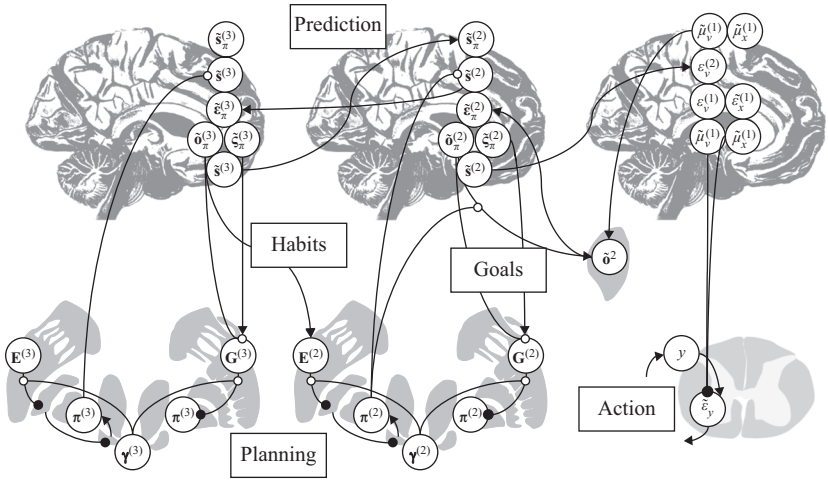


Figure 5.5

Anatomy of inference (based on Friston, Parr, and de Vries 2017) connects schematics from figures 5.1–5.4, providing a summary of the ideas in this chapter. Two hierarchical loops through the cortex and basal ganglia highlight the distinction between habits—based on input from higher levels—and the more context-sensitive, goal-directed (explorative and exploitative) behavior resulting from expected free energy minimization. Note the influence of inferences about policies on **s**, implementing the Bayesian model averaging under policies referred to in the main text. This projection from basal ganglia to cortex may be mediated by intermediate structures, such as the thalamus. On the right, the categorical POMDP-based messages are relayed into a continuous predictive-coding network, involved in generation of action. Each categorical state is associated with an alternative prediction of continuous variables and contributes to a prediction error. The message in the opposite direction computes the posterior probability of the associated categorical outcome (**o**), which depends on priors based on the policy-dependent outcome ($\mathbf{o}_\pi$), beliefs about the policy ($\boldsymbol{\pi}$), and the likelihood of the continuous trajectory that may be computed from posterior expectations (μ) and variances (not shown) at the continuous level. More connections could be included here; for instance, in addition to habits (**E**), the selection of goals (**C** from chapter 4) is itself likely to depend on higher hierarchical levels, leading to hierarchical control of motivation (see Pezzulo, Rigoli, and Friston 2018 for details).

relating message passing to neuronal communication? It allows us to make empirical predictions based on the generative model that we hypothesize the brain is inverting. This may take the form of an evoked response—the change in potential that is measurable at the scalp on presenting the brain with a sensory stimulus—whose time course will depend on the amount of belief updating induced by that stimulus. Alternatively, computational neuroimaging methods can be used to associate simulated inferences with those brain regions exhibiting similar temporal dynamics (Schwartenbeck, FitzGerald, Mathys, Dolan, and Friston 2015). Making this association is important in understanding pathology—and therapeutics—for computational (i.e., neurological and psychiatric) disorders, allowing for expression of functional pathologies in terms of their biology.

Finally, it is worth acknowledging that much of the brain has been conspicuously absent in this chapter—partly for reasons of space but also because neuroscience is a work in progress. There are many opportunities to extend (or even replace) the account given in this chapter. To some degree we can extrapolate from what we have seen here. For example, parts of the amygdala are cytoarchitecturally equivalent to basal ganglia nuclei. Does this mean there is a class of policies evaluated by the amygdala? Could this structure be to autonomic policies what the basal ganglia are to those in the skeletomotor domain? Might other structures (like the pulvinar) play similar roles for other (e.g., mental) classes of policy? How should we understand cortical architectures that differ from the six-layered structure in figure 5.1? The cerebellum and the hippocampal formation each exhibit distinct but stereotyped microcircuitry (Wesson and Wilson 2011). Should we see these as anatomical rearrangements of the same Bayesian message passing schemes, or do they deal with different aspects of a generative model (Pezzulo, Kemere, and van der Meer 2017; Stoianov et al. 2020)? We raise these questions not to offer any answers but to highlight some of the exciting avenues of future research in theoretical neurobiology. Active Inference and its associated process theories offer a rigorous formal and conceptual framework in which to address these questions.

This is a section of doi:10.7551/mitpress/12441.001.0001

Active Inference

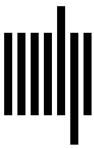
The Free Energy Principle in Mind, Brain, and Behavior

By: Thomas Parr, Giovanni Pezzulo, Karl J. Friston

Citation:

Active Inference: The Free Energy Principle in Mind, Brain, and Behavior
By: Thomas Parr, Giovanni Pezzulo, Karl J. Friston
DOI: 10.7551/mitpress/12441.001.0001
ISBN (electronic): 9780262369978
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data is available.

Names: Parr, Thomas, 1993– author. | Pezzulo, Giovanni, author. | Friston, K. J. (Karl J.), author.
Title: Active inference : the free energy principle in mind, brain, and behavior / Thomas Parr, Giovanni Pezzulo, and Karl J. Friston.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021023032 | ISBN 9780262045353 (hardcover)
Subjects: LCSH: Perception. | Inference. | Neurobiology. | Human behavior models. | Knowledge, Theory of. | Bayesian statistical decision theory.
Classification: LCC BF311 .P31366 2022 | DDC 153—dc23
LC record available at https://lccn.loc.gov/2021023032