

7 Active Inference in Discrete Time

What I cannot create, I do not understand.

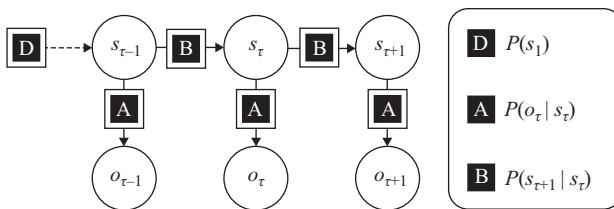
—Richard Feynman

7.1 Introduction

So far, we have discussed the principles of Active Inference at a relatively abstract level. This chapter deals with specific examples—and how they may be specified in a practical setting. We focus on models of categorical variables in discrete time. Through a series of examples, building in complexity, we illustrate models of perceptual processing, decision-making, information seeking, learning, and hierarchical inference. These examples are chosen to highlight as simply as possible emergent properties—including measurable physiology and behavior—of Active Inference schemes.

7.2 Perceptual Processing

We begin by considering perceptual processing and the inversion of the sort of discrete-time models introduced in chapter 4. Later in this chapter, we build to a full partially observable Markov decision process (POMDP). However, we start with a special case of a POMDP in which we can ignore choices and behavior: a hidden Markov model (HMM), which may be used for perceptual inference of a sequential and categorical sort (see figure 7.1). To motivate this, we will appeal to a simple example. Imagine listening to a performance of a short piece of music. The sequence of notes that are written in the score may be thought of as hidden (unobserved) states, while the sequence of notes we actually hear are the (observable) outcomes. If the performer is a professional musician, the correspondence between the

**Figure 7.1**

This hidden Markov model uses the same notation introduced in chapter 4 to express a sequence of states (s) that evolve through time. At each time, they give rise to an observable outcome (o). The state at one time depends only on the state at the previous time (with this dependency expressed in \mathbf{B}). The first state in the sequence has prior probability \mathbf{D} . The generation of outcomes from states depends on the likelihood distribution (\mathbf{A}). This specification of an HMM is generic, with specific generative models depending on specific choices for \mathbf{A} , \mathbf{B} , and \mathbf{D} .

hidden states and the outcomes may be very close. However, if an amateur, there may be an additional degree of stochasticity in the (likelihood) mapping from the note that should be played to that which is heard. In this scenario, it may still be possible to infer which note should have been heard, given prior beliefs about the probability that each note is preceded or succeeded by another.

The example of listening to the amateur musician may be formalized in the following way. First, we decide on how reliably our musician actually plays the note (outcome) she intends to (hidden state). We can express this through the \mathbf{A} -matrix, whose elements indicate the probability of an outcome (rows) given a state (columns). In our toy example, we set this as follows:

$$\mathbf{A} = \frac{1}{10} \begin{bmatrix} 7 & 1 & 1 & 1 \\ 1 & 7 & 1 & 1 \\ 1 & 1 & 7 & 1 \\ 1 & 1 & 1 & 7 \end{bmatrix} \quad (7.1)$$

This says that 70 percent of the time, our musician hits her intended note. We then specify the transition probabilities in the \mathbf{B} -matrix, which account for the probability of the next state (rows) given the current state (columns):

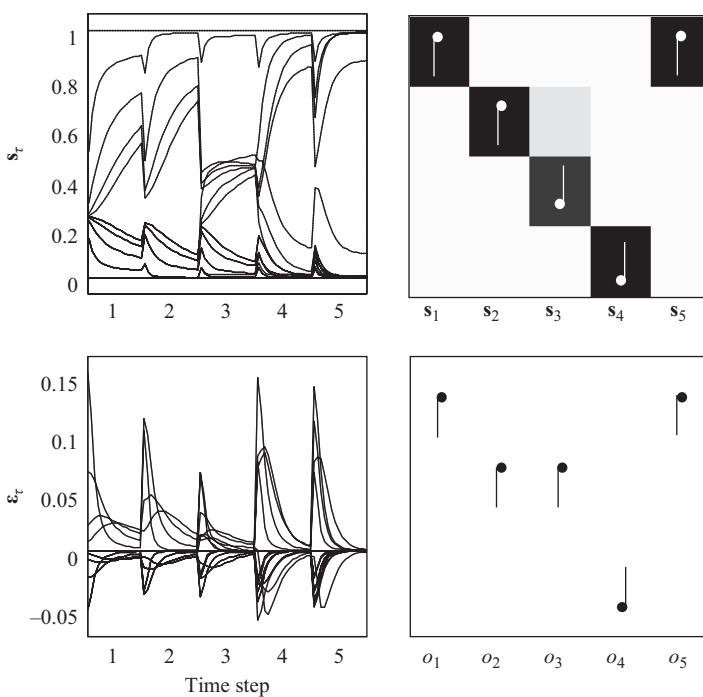
$$\mathbf{B} = \frac{1}{100} \begin{bmatrix} 1 & 1 & 1 & 97 \\ 97 & 1 & 1 & 1 \\ 1 & 97 & 1 & 1 \\ 1 & 1 & 97 & 1 \end{bmatrix} \quad (7.2)$$

This says that there is a 97 percent probability of the first note being followed by the second, the second by the third, and so on. If we know that the sequence always begins with the first note, we set the prior probability:

$$\mathbf{D} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (7.3)$$

Together, equations 7.1–7.3 completely specify the HMM generative model shown in figure 7.1. In other words, they provide a description of our beliefs about how the music we hear is generated by our amateur musician. Using equation 4.12 and substituting in our generative model, we can simulate the dynamics of the Bayesian belief updating induced by a sequence of outcomes. This is shown in figure 7.2. Note the increase in confidence shown in the upper-left plot as more data are accumulated over time, except for the third time step, where an unexpected outcome has occurred. This outcome could be explained in two ways. First, it may be that the intended note really was an unusual note under our prior beliefs in equation 7.2. This is made less likely by the rarity of such transitions under the **B**-matrix of this model. The alternative, more plausible explanation is that the musician played the wrong note by mistake. As shown in the third column of the upper-right plot, this is the explanation that our simulated listener settles on. However, a nonzero probability is assigned to the possibility that it was the right note after all. The capacity to report this sort of uncertainty is a key feature of the Bayesian perspective afforded by Active Inference.

The model shown here may be made more sophisticated in many ways, but perhaps the simplest relies on the factorization of the state-space (Mirza et al. 2016). An example might be the pitch and dynamics of the note (with a similar distinction in the outcomes). In a visual inference task, the factorization may be into what and where, which has a great deal of currency in neurobiology (Ungerleider and Haxby 1994). In subsequent sections, we will appeal to this sort of factorization to separate those states that can be influenced by the creature in question from those that cannot. For further reading on this sort of model (without actions in play) and the kinds of neuronal message passing scheme that might be used to invert it through minimizing free energy, see Parr, Markovic et al. (2019).

**Figure 7.2**

These simulated perceptual inference plots illustrate the belief-updating process in an example trial based on the generative model outlined in the main text. *Upper-left:* Beliefs (posterior probabilities) about each note in the sequence at each time step. *Upper-right:* As the numerical values of these beliefs are difficult to track the beliefs at the end of the sequence, having heard each note (i.e., retrospective beliefs) are shown. Each column shows (retrospective) beliefs about the hidden states at a given time step. Each row represents an alternative hypothesis for that hidden state. The darker the shading, the more probable that note is considered to have been (with black indicating a probability of one and white a probability of zero). *Lower-left:* (Negative) free energy gradients (i.e., prediction errors) over time. The rate of change of the beliefs in the upper-left plot is determined by the value of these errors at each time step. *Lower-right:* Sequence of musical notes presented to our synthetic agent (i.e., the observations he receives during time steps 1 to 5). Note that while at the third time step (o_3) the listener heard the second note (third column of the lower-right plot), he infers the third note with higher probability (third column of the upper-right plot).

7.3 Decision-Making and Planning as Inference

The HMM used above illustrates a very simple form of categorical inference based on a sequence of outcomes. However, the sort of (sessile) creature that this describes is rather uninteresting. Autonomous creatures are clearly more than passive recipients of sensory data. Instead, they actively change their environment and engage in a bidirectional exchange with their sensorium. This speaks to the importance of converting an HMM into a POMDP, whereby we must infer not only how our environment is changing but also how our chosen course of action changes it and which course of action to choose.

Figure 7.3 shows a POMDP generative model. This is the same as that introduced in chapter 4, where the details of inference in this sort of model are unpacked. Note the similarity of this structure to the HMM in figure 7.1

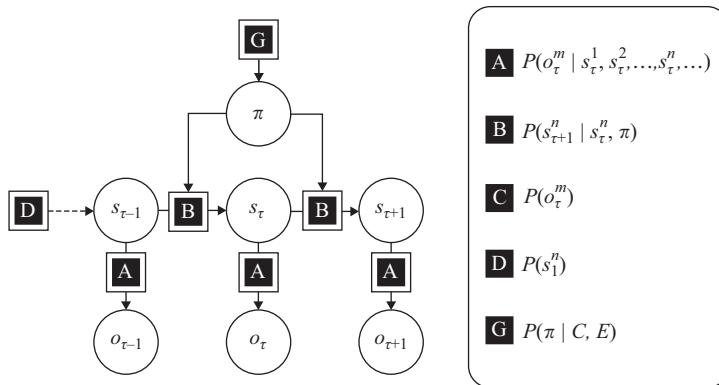


Figure 7.3

POMDP from figure 4.3, unpacking the probability distributions in terms of hidden state factors and outcome modalities. (Figure 7.1 is a special case of this structure.) Three points of note: First, the factorization of the hidden states now means that the distribution encoded by A has (potentially) many state factors in its conditioning set and can no longer be encoded by a matrix. Instead, this becomes a tensor object, in which each index corresponds to a state factor. Second, the separation of the outcomes into different modalities means there will be a separate A tensor for each modality. Third, while C and E appear in the panel on the right, they do not appear in the factor graph on the left because they only get into the generative model via prior beliefs about policies. For an alternative perspective on this, see Parr and Friston (2018d) and van de Laar and de Vries (2019).

and the addition of an extra variable (π), on which the transition probabilities (**B**) are conditioned. This means we can entertain alternative hypotheses about the dynamics of states. These hypotheses may be interpreted as plans that a creature may select between. This perspective equates policy evaluation with model comparison and says that a policy is simply an explanatory variable for an observed sequence of (self-generated) sensations.

The model in figure 7.3 differs subtly from that introduced in chapter 4: it allows factorization of states (superscript n) and of outcomes (superscript m). The utility of this is obvious when we consider the factorization of the visual world into where an object is and what it is. Clearly, it would be extremely inefficient (and incur a high complexity cost) to represent every possible combination of location and identity, when identity is (normally) invariant to location and vice versa. A similar argument may be used for factorization of time from identity and location (Friston and Buzsaki 2016). The benefit of introducing this factorization at this stage is that we can separate those states of the world over which a creature has control from those that it does not. While the transition probabilities governing the former will be different under each policy, the latter will be invariant to this.

With these preliminaries in place, we now outline a simple example of a task (Friston, FitzGerald et al. 2017) that requires planning and illustrates some of the key aspects of active inference using POMDPs. This involves a rat in a T-maze containing an aversive stimulus in one arm, an attractive stimulus in another, and a cue that indicates the location of the two stimuli in the final arm. This setup means that the rat can behave in (broadly) two ways. It could choose to go straight to one of the two arms that might contain the attractive stimulus, risking the aversive stimulus. Alternatively, it could choose to seek out the informative cue and then go to the arm most likely to contain the attractive stimulus.

This choice speaks to the classical exploration-exploitation dilemma in psychology: a dilemma that is resolved under Active Inference. The resolution stems from the minimization of expected free energy mandated by prior beliefs about policies. To review this briefly (see chapter 4 for details), the most probable policies (for a creature who minimizes its variational free energy) are those that lead to the lowest expected free energy. The expected free energy has the following form:

$$G(\pi) = \underbrace{\mathbb{E}_{Q(\tilde{s}|\pi)}[H[P(\tilde{o}|\tilde{s})]] - H[Q(\tilde{o}|\pi)]}_{\text{Negative epistemic value } (-\mathcal{I}(\pi))} - \underbrace{\mathbb{E}_{Q(\tilde{o}|\pi)}[\ln P(\tilde{o}|C)]}_{\text{Pragmatic value}} \quad (7.4)$$

This decomposition of the expected free energy into epistemic and pragmatic value highlights the (epistemic) drive toward information gathering and the (pragmatic) drive toward realizing prior beliefs (**C** in figure 7.3). We will attempt to provide a deeper intuition for the *epistemic value* in the next section, but it can be thought of simply as the amount of information we stand to gain under a specific policy. The form of the *pragmatic value* effectively treats the probability of outcomes, averaged over all policies, as if it were a prior. In doing so, those policies with consequences consistent with this prior become more probable, as they are associated with lower expected free energy. To put this in more intuitive terms, if we consider a certain sort of observation to be very probable, we will act to fulfill our belief that we will encounter these. Therefore, the log probability of outcomes may be thought of as equivalent to a utility function in other formalisms, such as optimal control theory and reinforcement learning. The fact that utility and the value of information emerge as two components of the expected free energy means that we do not need to worry about balancing exploration and exploitation. Both are in service of optimizing the same function.

To see how this unfolds in the T-maze example, we need to formalize the generative model in the same way as with the HMM above. Figures 7.4–7.6 illustrate the likelihood and transition probabilities that comprise the generative model for the T-maze. We will go through this in some detail, as this minimal example provides the building blocks from which readers can construct their own generative models. The first thing to do is to decide on the number of outcome modalities that represent the (sensory) data our model is supposed to explain. This tells us the number of **A**-matrices we must specify. Here, we have two modalities that represent exteroceptive data pertaining to *where* the rat is in the maze (**A**¹) and a *what* modality that may be the interoceptive data the rat experiences when it has found the attractive (edible) stimulus (**A**²). The levels in these modalities (i.e., the alternative observations that could be made in each) determine the rows of each **A**-matrix. The next decision is the number of hidden state factors that may be used to explain these data; this is the number of **B**-matrices we require. We consider two factors here: the position of the rat in the maze, and the context (attractive stimulus on left or right). These have four and two levels, respectively. We now must specify, for each combination of hidden states, the probability of each outcome. Context 1 is shown in figure 7.4; context 2 is shown in figure 7.5. For the first modality, our **A**¹ associates each location with an outcome with

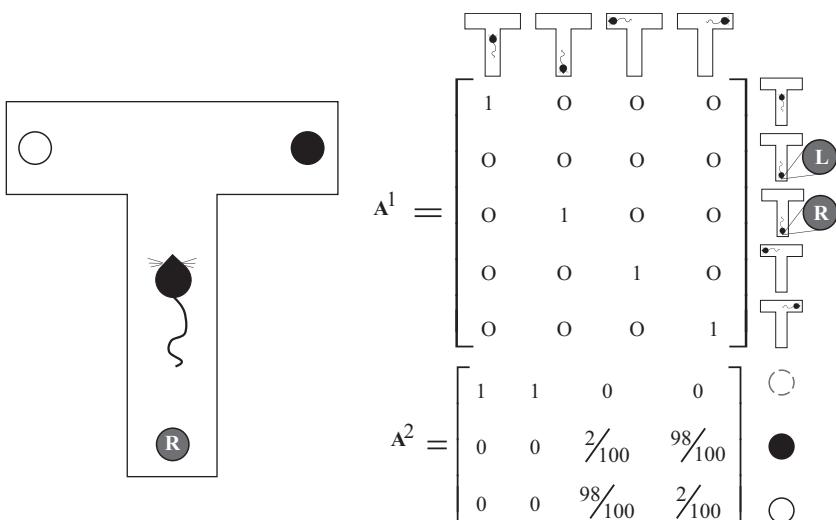


Figure 7.4

Likelihood in context 1. *Left:* T-maze configuration of cues and stimuli: the attractive stimulus is on the right and the aversive stimulus is on the left. *Right:* Likelihood or observation model specifies the probabilistic mapping from location to exteroceptive cues (A^1) and to interoceptive cues (A^2). Each element of these matrices is the probability of the outcome illustrated at the end of the row, conditioned on the context being one, and on being in the location indicated by the row. The exteroceptive outcomes are visual or proprioceptive input associated with each location, whereby the cue location can give rise to a rightward or a leftward cue. The interoceptive outcomes are absent (circle with dashed outline), attractive (filled circle), or aversive (unfilled circle).

probability one. The cue location may be associated with a left or a right cue, depending on the context. The interoceptive modality (A^2) associates a neutral outcome with the start and cue locations and a 98 percent chance of finding the attractive outcome when the context matches the arm of the maze the rat has entered. Technically, these A -matrices are tensor quantities, because their elements are specified by three numbers (outcome, location, and context), while a matrix is only specified by two (row and column).

We then need to specify transition probabilities. The B -matrices specify the probability of transitioning from a state (column) to another state (row), depending on the choice of policy (π). These specify the transitions pertaining to the position of the rat in the maze (B^1) and transitions in the context

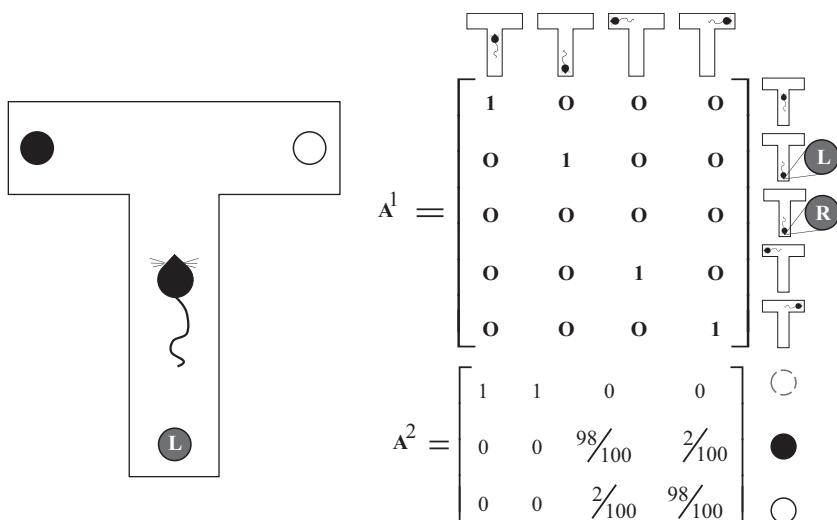


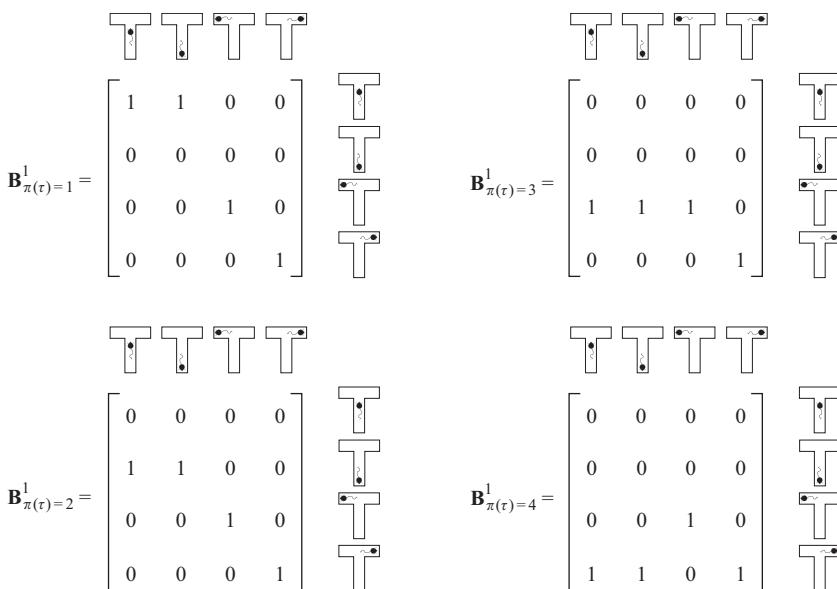
Figure 7.5

Likelihood in context 2. Nearly identical to figure 7.4—in this context, the aversive and attractive stimuli have been swapped. This is reflected in the probability of the exteroceptive outcomes in the cue location and the probabilities of the interoceptive outcomes in the right and left arms of the maze.

(B^2). Figure 7.6 shows the controllable B^1 -transitions. Each matrix shows the probabilities under a different action choice (subscripted). These allow a move from any location to any other location, except for the two arms of the maze, which are absorbing states. This means that once there, the rat must stay there, regardless of the actions it chooses. In contrast, the rat has no control over the context (i.e., whether it is in context 1, shown in figure 7.4, or context 2, shown in figure 7.5). Context stays constant over time and can be represented as an identity matrix:

$$B_\pi^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (7.5)$$

Here each column (and row) refers to a state indexing either figure 7.4 or figure 7.5. This means that whichever context we start in stays constant (transitions to itself) over time. This is true regardless of the policy selected. The C^1 -vector shows prior preferences for each of the outcomes in this modality, with uniform preferences except for a slight aversion (-1) to

**Figure 7.6**

Controllable transition probabilities for moving between the different locations. Each of the four matrices corresponds to an alternative action the rat can choose. These allow for a move from any state (except the right and left arm) to any other state. The right and left arms are absorbing states, in which the rat must stay once entered.

the start location. The \mathbf{C}^2 -vector specifies preferences (+6) for the attractive stimulus and aversion (-6) to the aversive stimulus. The absence of either is considered neutral (0).

$$\begin{aligned}\mathbf{C}^1 &= \sigma([-1, 0, 0, 0, 0]^T) \\ \mathbf{C}^2 &= \sigma([0, 6, -6]^T)\end{aligned}\tag{7.6}$$

The order of elements in these vectors corresponds to the order of rows in the corresponding \mathbf{A} -matrices. The softmax function (σ) allows us to specify preferences in terms of positive and negative values (corresponding to unnormalized log probabilities), which are then converted to probabilities. This preserves the difference in log probabilities (or the relative probability) while ensuring normalization. Practically, this formulation means the attractive stimulus is considered e^6 (≈ 400) times more probable than the neutral stimulus under the rat's generative model. This is a very strong preference that means the rat believes its actions are much more likely to lead to the attractive outcome. This constraint on inference about action is

crucial for the behavior that follows. Finally, the **D**-vectors specify the prior probabilities for the initial states:

$$\begin{aligned}\mathbf{D}^1 &= [1, 0, 0, 0]^T \\ \mathbf{D}^2 &= \frac{1}{2}[1, 1]^T\end{aligned}\tag{7.7}$$

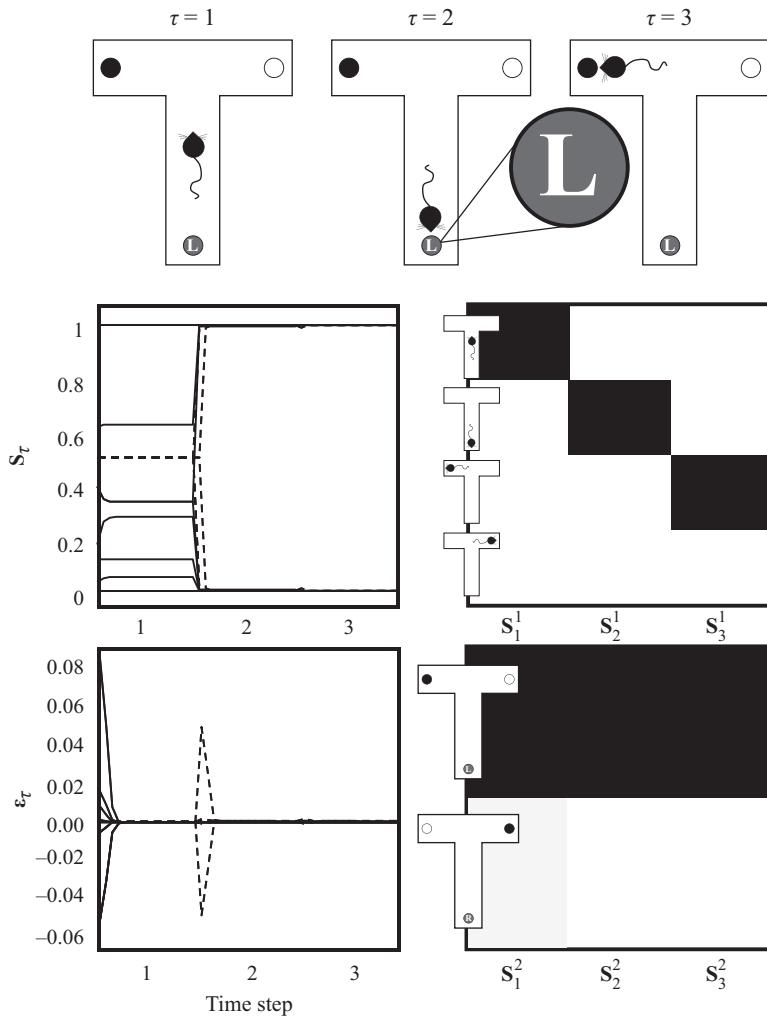
The order of elements in these vectors matches those of the **B**-matrices. The \mathbf{D}^1 -vector indicates a confident belief in starting at the center of the maze. The \mathbf{D}^2 -vector indicates that the two contexts (figure 7.4 or 7.5) are considered equally probable at the start.

Figure 7.7 shows what happens when we invert the generative model of figures 7.4–7.6. The upper row illustrates what we would see if observing the rat's behavior. It starts in the center and then goes to the informative cue. This is due to the high *epistemic value* associated with this location (i.e., the observations made at this location have the potential to resolve uncertainty about the context). On seeing the cue that indicates a left context (context 1), the rat chooses the left arm of the maze and finds the rewarding stimulus. This move is driven by the high *pragmatic value* attributed to this location. The lower plots illustrate the belief updating that occurs during this simple trial. As in figure 7.2, this is shown in the form we might expect to observe in an idealized rat if we were measuring neuronal activity (i.e., firing rates and local field potentials [LFPs]). Note the rapid change in beliefs at the second time step, when the rat reaches the informative cue location, and associated LFP (dashed line).

7.4 Information Seeking

The simulation in section 7.2 illustrates a simple example of an exploration-exploitation trade-off, which is solved by foraging for information until uncertainty is resolved, then exploiting what has been inferred to fulfill prior preferences. In this section, we unpack the concept of epistemic value in greater detail. As we saw in equation 7.4, this comprises two terms:

$$\begin{aligned}\underbrace{\mathcal{I}(\pi)}_{\text{Epistemic value}} &= \underbrace{H[Q(\tilde{o}|\pi)]}_{\text{Post. pred. entropy}} - \underbrace{E_{Q(\tilde{s}|\pi)}[H[P(\tilde{o}|\tilde{s})]]}_{\text{Expected ambiguity}} \\ &= \underbrace{D_{KL}[P(\tilde{o}|\tilde{s})Q(\tilde{s}|\pi) \| Q(\tilde{o}|\pi)Q(\tilde{s}|\pi)]}_{\text{Mutual information}} \\ &= \underbrace{E_{Q(\tilde{o}|\pi)}[D_{KL}[Q(\tilde{s}|\pi, \tilde{o}) \| Q(\tilde{s}|\pi)]]}_{\text{Information gain, salience, Bayesian surprise}}; Q(\tilde{s}|\pi, \tilde{o}) \triangleq \frac{P(\tilde{o}|\tilde{s})Q(\tilde{s}|\pi)}{Q(\tilde{o}|\pi)}\end{aligned}\tag{7.8}$$

**Figure 7.7**

Simulated epistemic and pragmatic behavior of a rat foraging in a T-maze. The rat starts in the central location but then chooses to sample the informative cue in the lower arm of the maze. This location is associated with the greatest *epistemic value*, as observing the cue in this location reveals the context (reward right or left) that the rat finds itself in. On observing the cue, the rat undergoes rapid belief updating (s), inducing an LFP (ϵ). With no more uncertainty to resolve, the rat selects the *pragmatically valuable* option and goes to the left arm of the maze. The two plots on the right show the beliefs held by the rat at the end of the trial about all previous times (i.e., these are retrospective beliefs and not the beliefs of the rat at the moment of the decision). It believes (correctly) that it started in the central location, went to the cue arm, and then went to the left arm. For the context hidden state factor, the rat believes that the context was the left context throughout.

These are the *posterior predictive entropy* and the *expected ambiguity*, respectively. Below these, we highlight the correspondence between these and other rearrangements. To unpack these in an intuitive way, we will frame this in terms of a visual paradigm, where alternative saccades (π) lead to different transitions between fixation locations (s). In addition to fixation locations, the hidden states include the identity of a stimulus at each location. A combination of stimulus and fixation generate visual and proprioceptive consequences (o). With this in mind, we can interpret the posterior predictive entropy as the dispersion (or uncertainty) associated with “what I would see if I performed this eye movement.” From the perspective of a scientist, this quantifies how uncertain we might be about the data we would obtain on performing a given experiment. Under this perspective, it makes sense that we should select those saccades (or experiments) that are associated with the greatest posterior predictive entropy, as these offer the greatest potential for uncertainty resolution. We would gain nothing by performing an experiment if we already knew what the results would be with a high degree of confidence.

However, the predictive entropy only tells us the total amount of uncertainty. It does not tell us how much uncertainty is actually resolvable. We will always be uncertain about the next number in a sequence of randomly generated numbers, but we will never resolve our uncertainty about the process generating them by fixating on these. This is where the expected ambiguity comes in. This quantifies the degree to which observations and states are independent of one another. If states always generate the same observation, this quantity will be zero. It will be maximal if, as in the random number generator, there is no association between states and outcomes. In the visual domain, this implies that the best saccade will be that toward a well-lit stimulus, where there is little ambiguity about “what I would see if I looked at this stimulus.” Taken together, this says that the best saccades (i.e., perceptual experiments) are those for which there is the greatest uncertainty to resolve (posterior predictive entropy) but only if that uncertainty can be resolved (negative ambiguity). Interestingly, this has exactly the same form as expressions developed in statistics to score experimental design in terms of information gain (Lindley 1956).

Figure 7.9 illustrates what happens in a saccadic paradigm (Parr and Friston 2017b) when we simulate manipulations to the ambiguity and posterior predictive entropy. This shows four stimuli (squares), each of which may change color from moment to moment. Superimposed on these is a

simulated eye-tracking trace, as if we were measuring where an experimental participant was looking. Crucially, we specify prior beliefs about outcomes to be uniform (i.e., pragmatic value to be absent), precluding any preference-based choices. This means each saccade is selected to maximize epistemic value. When the generative model treats all four stimuli as equivalent (left image), all are sampled with approximately the same frequency. However, we can modulate the uncertainty associated with each stimulus (see box 7.1). If we set one stimulus to have a greater ambiguity (by increasing the value of off-diagonal elements of the corresponding \mathbf{A} -matrix), this square is ignored (middle image). This is an example of the famous “streetlight” effect (Demirdjian et al. 2005), which takes its name from the metaphor of

Box 7.1

Uncertainty and precision

The example in figure 7.7 appeals to the concept of precision—an important idea in this book. Precision is the inverse of variance and scores our confidence in a given probability distribution. This is closely related to the negative entropy (*negentropy*) of a distribution:

$$-H[P(s)] = \mathbb{E}_{P(s)}[\ln P(s)]$$

A simple way to parameterize a distribution such that it can be made more or less precise is to use a Gibbs form with an *inverse temperature* parameter (ω). This has the following form:

$$P(s|\omega) = \text{Cat}(\sigma(\omega \ln \mathbf{D}))$$

Note that the precision multiplies the log prior, so it can be interpreted as a gain-control device (amplifying as opposed to adding to neural signals). The plots in figure 7.8 show how the probability distribution (each column representing the probability of an alternative state) changes for a given \mathbf{D} when we vary ω . Note the increasing confidence with increasing precision.

This sort of parameterization may be applied to any of the distributions used in a POMDP. In addition, we can define priors over the precision and infer



Figure 7.8

Box 7.1 (continued)

this just as we infer other latent variables (i.e., through free energy minimization). Assuming the prior has a Gamma distribution (precluding negative values of the precision), we get the following updates (see appendix B for details):

$$\begin{aligned} P(\omega) &= \Gamma(1, \beta_\omega) \\ Q(\omega) &= \Gamma(1, \beta_\omega) \\ \Rightarrow \hat{\beta}_\omega &= (\mathbf{D}^{\beta_\omega^{-1}} - \mathbf{s}) \cdot \ln \mathbf{D} + \beta_\omega - \beta_\omega \end{aligned}$$

There is an increasing recognition that the biological substrate of these precision parameters may be the neuromodulatory systems that set the gain of neural responses. Chapter 5 discusses the evidence relating these parameters to specific neurochemicals.

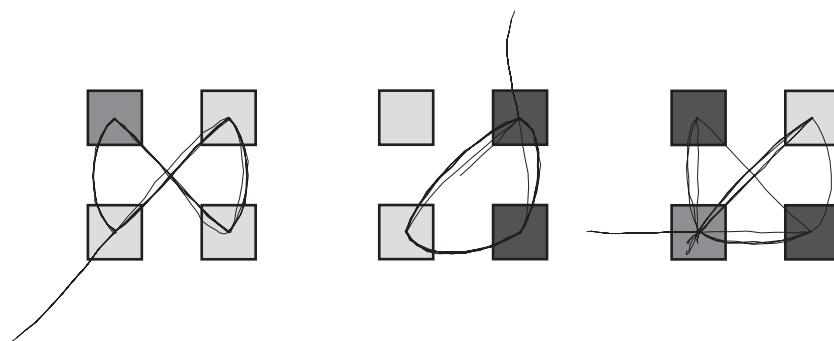


Figure 7.9

Simulated epistemic visual search paradigm (Parr and Friston 2017b) with the synthetic eye-tracking trace superimposed on the four stimulus locations. Each stimulus (shaded square) is associated with a transition matrix that may be more or less predictable and a likelihood matrix that may be more or less ambiguous. *Left:* When transitions and likelihoods are equally predictable for all four locations, all locations are sampled with about the same frequency. *Middle:* The viewer shows aversion to the upper-left square when it is specified with a less precise (more ambiguous) likelihood mapping. *Right:* The lower-left square is epistemically attractive when the transition probabilities are specified as more uncertain.

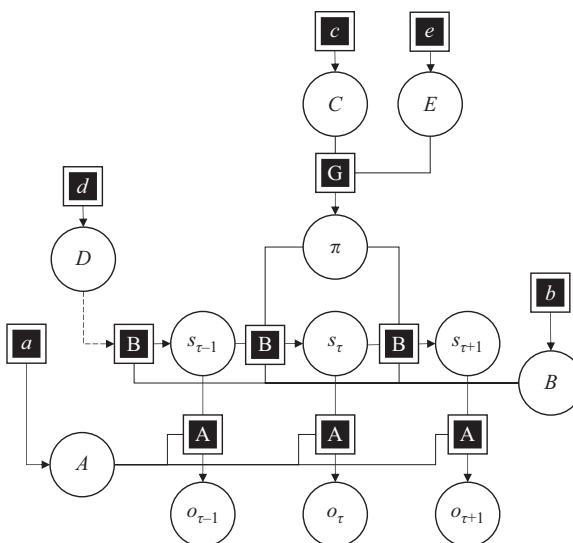
people who have lost their keys late at night. The first place they might look is under the streetlight—not because the keys are most likely to be there, but because it is the best place to find high-quality, unambiguous, uncertainty-resolving information. The simulation shows how the ambiguous (e.g., poorly lit) square is ignored, reproducing an *in silico* streetlight effect.

In contrast, the right image in figure 7.8 shows what happens when we make the transitions less predictable for the lower-left square. We accumulate uncertainty about this location very quickly, ensuring a high posterior predictive entropy with no change to the ambiguity. As we can see, this leads to more frequent fixation on this location, as there is always new uncertainty to resolve here. Intuitively, if I know something has very predictable dynamics, I do not have to look at it very often to be confident about its state. In contrast, if something may have changed in the time that I have been looking at something else, it is worth looking back at to check. These simulations are designed to offer an intuition for the two parts of the epistemic value, to see how minimization of expected free energy ensures we actively select our sensory data to find out about the world.

7.5 Learning and Novelty

Sections 7.2–7.4 set out everything that is required for the majority of practical applications of Active Inference. However, we have assumed that the generative model is already known and does not change as an effect of learning. In some practical applications, we may want to consider how one or more parts of the generative model (e.g., the **A**- or **B**-matrix) are learned during an experiment or, more broadly, how we optimize the structure of the generative model itself, given some data (Friston, FitzGerald et al. 2016). In doing so, Active Inference extends to active learning, and the *salience* (equation 7.5) describing information gain about states is complemented by *novelty*, which deals with resolution of uncertainty about (for example) the elements of the **A** matrix shown in equation 7.1, the **B** matrix shown in equation 7.2, or any other parameters of the generative model. These beliefs can now vary with time rather than being fixed, as assumed so far (Schwartzenbeck et al. 2019). To get to this, we first have to extend the generative model as in Figure 7.10 to include beliefs about these model parameters.

Conceptually, including beliefs about parameters in the generative model permits treating learning as another form of Bayesian inference—namely,

**Figure 7.10**

This generative model for learning uses the same POMDP structure as in figure 7.3, but the priors for each of the hidden states now depend on variables (in circles), which themselves now come equipped with prior beliefs. These have the form of Dirichlet distributions, which are conjugate (see box 7.2) to the categorical distributions considered thus far. The model shows how the likelihood of outcomes given states now also depends on a variable A (which is the same for all time-points), the transition probabilities are now conditioned on a variable B , the preferences depend on C , the initial states depend on D , and the fixed form policy prior depends on E . By making prior beliefs about the parameters of the generative model explicit, this figure emphasizes that both inference and learning are free energy minimizing processes, but they are distinct. In short, inference describes the optimization of beliefs about the state of the world as it is (s), including beliefs about the way in which we are acting (π). In contrast, learning describes optimization of beliefs about the relationships between these variables (A, B, C, D , or E). The latter vary much more slowly than the former and may only be learned when the states have been inferred. We will return to this separation of timescales below when we consider hierarchical generative models.

as the passage from prior to posterior beliefs about model parameters. This highlights the fundamental similarity of perception and learning: in the same way that perception can be described as the inversion of a generative model to infer hidden states from observations, learning can be described as the inversion of a generative model to include beliefs about parameters (although normally this inversion may operate on a slower timescale).

Box 7.2

Conjugate priors

When setting up a generative model of the form in figure 7.10, it is important to carefully select the appropriate distribution for prior beliefs. Typically, these will be the *conjugate* prior distribution associated with the likelihood. A conjugate prior belief means that, when used to perform Bayesian inference, the posterior belief will be the same type of distribution. For example, using Bayes' rule:

$$P(D|s) \propto P(D) P(s|D)$$

If $P(s|D)$ is a *categorical* distribution, when we choose a *Dirichlet* distribution (conjugate to *categorical*) for $P(D)$, we can guarantee that $P(D|s)$ is also a *Dirichlet* distribution. Put formally:

$$\left. \begin{array}{l} P(D) = Dir(d) \\ P(s|D) = Cat(D) \end{array} \right\} \Rightarrow P(D|s) = Dir(d)$$

The simplest way to choose the right kind of prior is to look up the conjugate prior for whatever form the likelihood distribution takes. For the categorical distributions used here, a Dirichlet distribution is the appropriate choice for beliefs about parameters (see box 7.2). Having included these additional prior beliefs, we can now optimize posterior beliefs about the structure of the generative model. This means incorporating these into the free energy (as we did for states in chapter 4) and finding the free energy minima.

$$\theta = (A, B, C, D, E) \quad (7.9)$$

$$F = \mathbb{E}_{Q(\pi, \theta)}[F(\pi, \theta)] + D_{KL}[Q(\theta) \parallel P(\theta)] + D_{KL}[Q(\pi) \parallel P(\pi)]$$

Dirichlet distributions are parameterized by counts (or pseudo-counts) that index the number of times a given categorical variable has been seen (or, in the case of the priors, as if it had been seen that number of times). For the derivation of the update rules for these parameters, see appendix B. For now, we summarize the update rule and key properties of a Dirichlet distribution, focusing on the a and \mathbf{a} concentration parameters associated with the prior and posterior over A .

$$\mathbf{a} = a + \sum_{\tau} \mathbf{s}_{\tau} \otimes \mathbf{o}_{\tau}$$

$$\mathbb{E}_Q[A_{ij}] = \mathbf{A}_{ij} \approx \frac{\mathbf{a}_{ij}}{\mathbf{a}_{0j}} \quad (7.10)$$

$$\mathbb{E}_Q[\ln A_{ij}] = \ln \mathbf{A}_{ij} = \psi(\mathbf{a}_{ij}) - \psi(\mathbf{a}_{0j})$$

$$\mathbf{a}_{0j} \triangleq \sum_k \mathbf{a}_{kj}$$

The first line here expresses the update from prior to posterior concentration parameters following a series of observations, with beliefs about the states that caused them. The cross in the circle indicates a Kronecker tensor product (or outer product in the case of two vectors), here giving rise to a matrix in which each element is the product of a pair of elements in s_τ and o_τ . This update rule may be interpreted simply as a form of activity-dependent plasticity. When an outcome is observed in combination with a posterior belief that a particular state caused it, the element of the matrix representing the relationship between the two is incremented. The second line of the equation highlights the interpretation of the Dirichlet concentration parameters in terms of counts. For a given state (column), each element of \mathbf{a} is the number of times the corresponding outcome has been seen. Dividing by the sum of the elements in the column (total number of observations or pseudo-observations) gives the probability of each outcome given that state. To understand why this (pseudo) counting method makes intuitive sense, consider the amateur musician example from the beginning of this chapter. If one counts how many times the musician hits the first note when she intends to do so (first row and column), how many times she hits the second note when she intends to do so (second row and column), and so on, and divides these by the total number of times she intends each note, one will eventually converge to the correct numerical values of the \mathbf{A} -matrix shown in equation 7.1—namely, that the musician hits all her intended notes 70 percent of the time. The counting method has another important consequence that we will return to: The number of counts or pseudo-counts preceding an observation tells us how likely we are to update our beliefs on making the observation. Imagine flipping a coin five times and getting five heads in a row. This might lead us to update our beliefs to favor the hypothesis that this is an unfair coin. However, if this had been preceded by 100 flips with 50 heads and 50 tails, the final five heads would do little to influence our beliefs about whether this is a fair coin. The third line of equation 7.10 shows a useful identity associated with Dirichlet distributions: the expected log of the random variable is given by the difference in two digamma functions (derivative of a gamma function).

The inferential approach to learning highlights an important difference between Active Inference and most other approaches to computational neuroscience and machine learning, which incorporate various learning rules (e.g., Hebbian rules or error backpropagation) that are considered biologically realistic or computationally efficient. In Active Inference, the

update rules that govern learning are derived from statistical considerations, yet they turn out to be remarkably similar to biologically motivated rules for activity-dependent plasticity (see the above considerations on the first line of equation 7.10). This exemplifies one of the appeals of normative approaches, which start from first principles to explain what we know about brains and behavior—and things that we did not know.

A further difference between Active Inference and most machine learning approaches is that learning is naturally described as an *active* process, in which creatures autonomously select the most appropriate data to improve their generative models. This becomes evident if one considers that when including beliefs about parameters in the model, the expected free energy acquires an additional term:

$$\begin{aligned}
 G(\pi) &= \underbrace{D_{KL}[Q(\tilde{o}|\pi) \parallel P(\tilde{o}|C)]}_{\text{Risk}} + \underbrace{\mathbb{E}_{Q(\tilde{s}|\pi)}[H[P(\tilde{o}|\tilde{s})]]}_{\text{Ambiguity}} \\
 &\quad + \underbrace{\mathbb{E}_{\tilde{Q}(\tilde{o}, \tilde{s}, \theta|\pi)}[\ln Q(\theta) - \ln P(\theta|\tilde{o}, \tilde{s})]}_{\text{Parameter information gain}} \\
 &= -\underbrace{\mathbb{E}_{Q(\tilde{o}|\pi)}[D_{KL}[Q(\tilde{s}|\pi, \tilde{o}) \parallel Q(\tilde{s}|\pi)]]}_{\text{Salience}} \\
 &\quad - \underbrace{\mathbb{E}_{\tilde{Q}(\tilde{o}, \tilde{s}|\pi)}[D_{KL}[Q(\theta|\tilde{o}, \tilde{s}) \parallel Q(\theta)]]}_{\text{Novelty}} - \underbrace{\mathbb{E}_{Q(\tilde{o}|\pi)}[\ln P(\tilde{o}|C)]}_{\text{Pragmatic value}}
 \end{aligned} \tag{7.11}$$

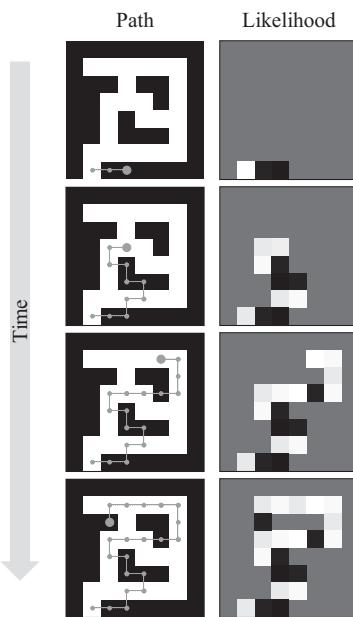
The *salience* and *pragmatic value* terms were already in place in equation 7.4, but the *novelty* term is new. The final equality here shows an arrangement that highlights the relationship between salience and novelty. In short, salience is to inference what novelty is to learning. Both are expressions of the change in beliefs anticipated once a perceptual experiment (i.e., an action in a policy) is performed. As with scientific experiments, the greater the change in beliefs following data collection, the better the experiment. Returning to the analogy of flipping a coin and accumulating counts, this tells us something useful. If we have two coins and can choose to flip either one, we can elicit the greatest change in beliefs by flipping the coin we had flipped only five times previously rather than the coin with 100 previous flips. There is greater novelty associated with flipping the former (less familiar) coin. Similarly, if we have confident prior beliefs *as if* we had observed something many times, policies that interrogate these variables are associated with less novelty than those about which we have less confident beliefs.

To illustrate how this works in practice, imagine we have a very myopic creature standing on a tiled floor. This creature can only see the color of

the tile it is standing on and can only move one tile at a time. For any suitably large landscape with many tiles, it is very computationally expensive to represent the color of each tile as a different hidden state. However, a simpler form of model is available. If we associate hidden states only with location, and colors only with outcomes, we can efficiently represent beliefs about “what I would see if I went over there” in the \mathbf{A} matrix that generates colored tiles from locations. By accumulating Dirichlet parameters (equation 7.10), our creature can optimize these beliefs on the basis of observations. We might interpret this as a form of synaptic memory as opposed to the maintenance of persistent activity in neurons representing beliefs about the color of a given tile. Given this sort of generative model, wherein all of the uncertainty is in the parameters of the likelihood distribution, it is interesting to see what happens in the absence of any preferences (i.e., when the novelty term of equation 7.11 dominates policy selection). Figure 7.11 shows a simulation of a simple environment comprising 64 black or white tiles. As each tile is visited, beliefs about the likelihood of observing black or white in that location are updated through accumulation of Dirichlet parameters. As large Dirichlet parameters preclude large belief updates, the drive to novelty resolution given by expected free energy minimization leads our simulated creature to avoid any previously visited locations.

The same principles could be applied to a range of other paradigms (e.g., if we reinterpret the path taken by our creature as a saccadic scan path, this could be applied to active visual sampling). In the domain of active vision, this has been used to simulate the kinds of visual search behavior induced by target cancellation tasks (Parr and Friston 2017a). Subsequently, evidence for the short-term plasticity required in accumulating Dirichlet parameters in this setting has been demonstrated (Parr, Mirza et al. 2019).

Just as we can extend ideas about inference to learning, it is possible to go (at least) one step further and think about structure learning: the process of not just optimizing the parameters in the model but selecting between different models with more or fewer parameters in play. Box 7.3 sets out a way of doing this that involves efficient post hoc comparisons of alternative hypothetical models. This has been used as a metaphor for sleep (Friston, Lin et al. 2017) and resting spontaneous activity (Pezzulo, Zorzi, and Corbetta 2020), where no new data are collected but the structure of the model may still be refined and simplified.

**Figure 7.11**

Active learning is demonstrated by a synthetic creature exploring a simple world of black and white tiles (Bruineberg et al. 2018, Kaplan and Friston 2018). *Left:* Path taken by the creature, showing which tiles are white and which are black (dots correspond to visited locations). *Right:* A matrix of the creature and the beliefs (in terms of normalized Dirichlet counts) the creature has about what it would see on going to different locations. Cells in the \mathbf{A} matrix are white (or black) if the creature has a strong belief that the corresponding tile is white (or black); they are grey if the creature is uncertain about color. Crucially, these beliefs influence which path it takes via the novelty term of the expected free energy. Those locations about which it has confident beliefs afford relatively little opportunity for uncertainty resolution, so it does not revisit them. In other words, the phenomenon of “inhibition of return” (Posner et al. 1985) emerges naturally from the minimization of expected free energy.

7.6 Hierarchical or Deep Inference

In the previous section, we saw one method for hierarchical extension of the original generative model based on defining priors over the parameters of the generative model. Figure 7.12 shows a second form of hierarchy that speaks to the nesting of temporal scales. This generative model for hierarchical or deep inference can be conceived of as a hierarchical extension

Box 7.3

Structure learning and model reduction

The discussion in section 7.4 deals with an important, but limited, form of (parametric) learning. The next level of sophistication—in learning about the structure of the world—goes beyond the optimization of model parameters and asks whether we should expand or prune the model structure. This can be cast as a question of model comparison (Friston, Lin et al. 2017). In other words, would my free energy increase or decrease if I were to (for example) eliminate elements of a likelihood matrix? By comparing models with and without these elements, we can answer this question. However, it may be very costly to have to explicitly invert multiple models. Fortunately, an efficient method for doing this—known as Bayesian model reduction (Friston, Litvak et al. 2016; Friston, Parr, and Zeidman 2018)—is available and only requires inversion of a single full model. In a general setting, comparison between a full model and one with alternative priors (indicated by \sim) can be achieved through the following formulae:

$$\Delta F = F[\tilde{P}(\theta)] - F[P(\theta)] = \ln \mathbb{E}_{Q(\theta)} \left[\frac{P(\theta)}{\tilde{P}(\theta)} \right]$$

$$\tilde{Q}(\theta) \propto \exp(\ln Q(\theta) + \ln \tilde{P}(\theta) - \ln P(\theta) + \Delta F)$$

For the Dirichlet priors used in section 7.5, this takes the form (where B is the multivariate beta function):

$$\Delta F = \ln B(\tilde{\mathbf{d}}) - \ln B(\mathbf{d}) + \ln B(\mathbf{d}) - \ln B(\tilde{\mathbf{d}})$$

$$\tilde{\mathbf{d}} = \mathbf{d} + \tilde{\mathbf{d}} - \mathbf{d}$$

This form of model reduction may be important in understanding offline model optimization, of the sort that may occur during sleep. We will briefly revisit Bayesian model reduction in chapter 8, when considering the optimization of hierarchical models with both discrete and continuous components.

of the shallow model shown in figure 7.3: it includes a series of POMDP models at the lower level that are the same as in Figure 7.3 (one example is outlined with the dashed box), contextualized by a higher-level POMDP.

Importantly, this generative model includes variables that evolve at different timescales: slower for higher levels and faster for lower levels (Friston 2008; Friston, Rosch et al. 2017; Pezzulo, Rigoli, and Friston 2018). This becomes evident if one considers that the POMDP models at level 1 evolve over three time steps, but each of these short trajectories of states and outcomes depends on a single state at the higher level (level 2) that persists

throughout the entire trajectory at the lower level. In other words, for every time step from the perspective of the higher level, there are multiple (here, three) time steps for the lower level.

To gain some intuition for this separation of timescales—which underwrites deep temporal inference—it is worth thinking about a simple example of hierarchy in everyday life: reading. We draw inferences about words that combine to form sentences. Sentences combine to form paragraphs, pages, books, libraries, and so on. If we imagine that each state at the lower level of figure 7.12 is a word, each state at the higher level can be thought of as a sentence. Crucially, the duration of the sentence transcends that of any one word in the sequence.

The reading example is illustrated in more detail in figure 7.13, which is based on the example from Friston, Rosch et al. (2017), to which we refer

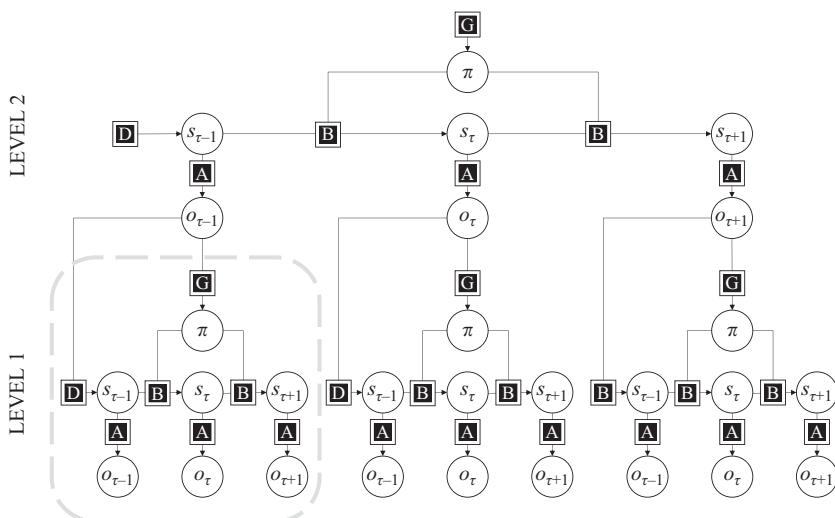


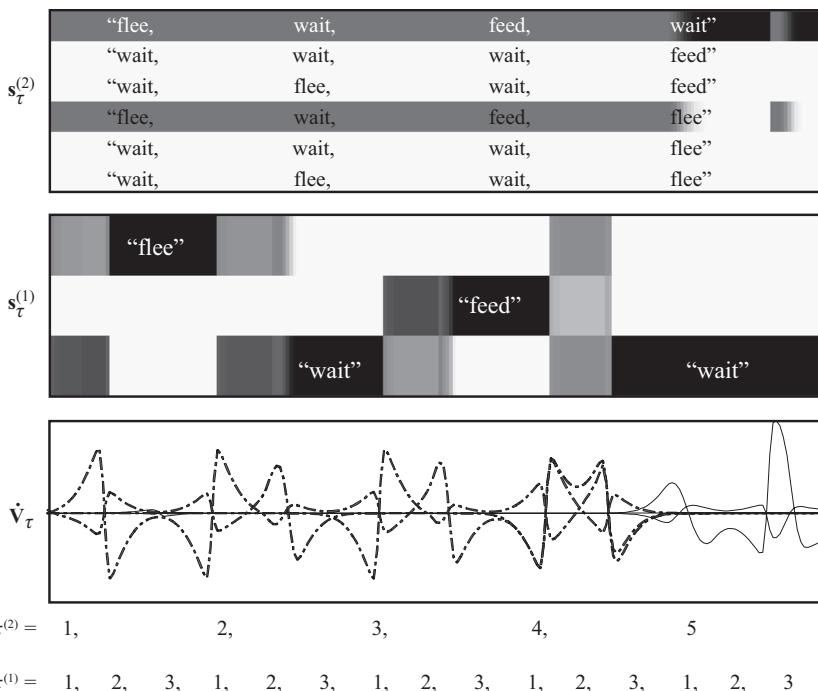
Figure 7.12

We can extend the (shallow) generative model set out in figure 7.3 so that it affords hierarchical or deep inference, which evolves over multiple timescales. The full generative model includes a slowly changing context (at level 2) that generates a series of short trajectories at the lower, faster level 1. The form of the POMDP is the same at the higher level as at the lower level (one of the POMDPs is outlined with the dashed box). The only difference is that it is stretched out in time (horizontally) and that the outcomes it generates are not directly observed. Instead, they form empirical priors for the lower level, which generates observable outcomes.

for more details. The model is structured as in figure 7.12 and represents sentences (at the higher level) and words (at the lower level) drawn from a very simple language. This language comprises three possible words (*flee*, *feed*, *wait*) that may be arranged into six possible four-word sentences. If the sentence is “*flee*, *wait*, *feed*, *wait*,” the higher level predicts the word *flee* for the first of the lower-level POMDPs, *wait* for the second, and so on. At the lower level, we start with an empirical prior (**D**) based on the higher level, which tells us which words are most plausible. For example, if we started with a uniform distribution over the sentences shown in the upper panel of figure 7.13, we see that the first word is *wait* in two-thirds of the sentences and *flee* in the other third. This means that at the first time step of the first low-level POMDP, our **D**-vector should ascribe these probabilities to these words.

The words at the lower level then generate observations, visual inputs based on which part of the word is currently foveated. Much as in the example of figure 7.9, the POMDP allows for selection of different foveal targets to accumulate evidence for or against each hypothetical word. This appeals to the same expected free energy minimizing processes outlined above; therefore, we will not detail the specific foveations made here, but we note that with each time step at the lower level, there is an increase in confidence about the word in play. In the sequence shown in figure 7.13, we see that evidence is accumulated for the word *flee* at the lower level over the first few time steps (over the fast scale, $\tau^{(1)}$). This inference is propagated back up to the higher level, where it provides evidence for the first and fourth sentences (each of which start with this word). Over subsequent time steps the evidence accumulated at the lower level is consistent with both sentences. At the fourth step (at the slow scale, $\tau^{(2)}$), we would predict *wait* under the first sentence and *flee* under the second. On inferring *wait* at the fast timescale, the first sentence is inferred at the slow scale. At the final step, the simulation selects the correct sentence and is rewarded with correct feedback. The resulting belief updating is seen in the LFP plot in the lower part of figure 7.12.

Deep temporal models of this sort have been used to simulate reading (Friston, Rosch et al. 2017), delay-period working memory tasks (Parr and Friston 2017c), and computation of empirical priors for visual inference (Parr, Benrimoh et al. 2018). In addition, they have been leveraged in theoretical accounts of motivation and control (Pezzulo, Rigoli, and Friston 2018). In

**Figure 7.13**

Belief updating occurs over multiple timescales by inverting a simulated hierarchical inference model. This relies on a generative model with a separation of timescales (shown as a slow timescale— $\tau^{(2)}$ —and a fast timescale— $\tau^{(1)}$). Belief updating at the higher level ($s^{(2)}$), representing sentences, is slower than at the lower level ($s^{(1)}$), representing words. *Lower panel:* LFPs, i.e., the rate of change of the log expectations—which is proportional to the prediction errors (ϵ) shown in previous figures.

principle, these models can be extended to an arbitrary number of levels, accounting for a deeply structured world with dynamics that play out over many different temporal scales.

We can draw an interesting parallel between hierarchical models of the sort in figures 7.12 and 7.13 and learning models of the sort in figure 7.10. Learning models can be considered hierarchical generative models, which highlight a separation of timescales between faster inferential dynamics (updates of beliefs about states) and slower learning dynamics (updates of beliefs about parameters). The models shown in figures 7.10 and 7.12 may be also combined to arbitrary levels of complexity, wherein

the relationships between variables on different levels may themselves be learned. This permits designing increasingly sophisticated generative models that address systems-level cognitive and neurobiological questions.

7.7 Summary

In this chapter, we saw some of the ways that discrete-time generative models may be constructed to address a range of cognitive and neurobiological problems, such as perceptual inference, decision-making and planning, balancing exploration and exploitation, parametric and structure learning, and novelty seeking. This is far from an exhaustive summary of applications of discrete models in Active Inference, but it serves to illustrate the key principles of this sort of modeling. The models outlined above may be combined hierarchically, with added priors over parameters, and with context-sensitive priors for policies or preferences. Importantly, inference using both simple and more complex generative models can always proceed through free energy minimization, which illustrates the generality of the approach. The fact that different aspects of Active Inference become apparent under distinct generative models (e.g., novelty seeking with priors over model parameters) opens up the possibility of exploring an open-ended set of cognitive and biological problems by designing the appropriate generative models.

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

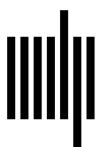
DOI: 10.7551/mitpress/12441.001.0001

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

© 2022 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC BY-NC-ND license.
Subject to such license, all rights are reserved.



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>