

# 9 Model-Based Data Analysis

Just because we have the best hammer does not mean that every problem is a nail.  
—Barack Obama

## 9.1 Introduction

Ultimately, the models described in this book are only useful if they can answer scientific questions. In this chapter, we focus on the ways in which Active Inference can be applied in understanding empirical data. The central idea is that we, as scientists, can appeal to the same maths as we have assumed the brain uses in previous chapters. Our general goal is to recover the parameters of the generative model that a subject's brain uses to produce behavior—the *subjective* model. For this, we can use our own generative model (of how the subjective model produces behavior)—the *objective* model. We can invert our objective model on the basis of the behavior we observe to draw inferences about the parameters of the subjective generative model. This meta-Bayesian inference affords the opportunity to test hypotheses about the model we assume the brain uses and to phenotype individuals on the basis of the prior beliefs they would have to hold for their behavior to be Bayes optimal. Belief-based computational phenotyping of this sort holds promise in the emerging fields of computational psychiatry, neuropsychology, and neurology.

## 9.2 Meta-Bayesian Methods

This chapter deals with the utility of Active Inference formulations in analyzing data from behavioral experiments. This goes beyond the proof-of-principle

simulations we have seen in previous chapters and instead exploits Active Inference in answering scientific questions. We have seen already that a subject's generative model is the key determinant of behavior under Active Inference. This implies that hypotheses about the causes of empirical behavioral measurements must be framed in terms of the alternative generative models used to select those actions. Our challenge, then, is to fit an Active Inference scheme to observed data by manipulating the parameters (i.e., prior beliefs) of the generative model.

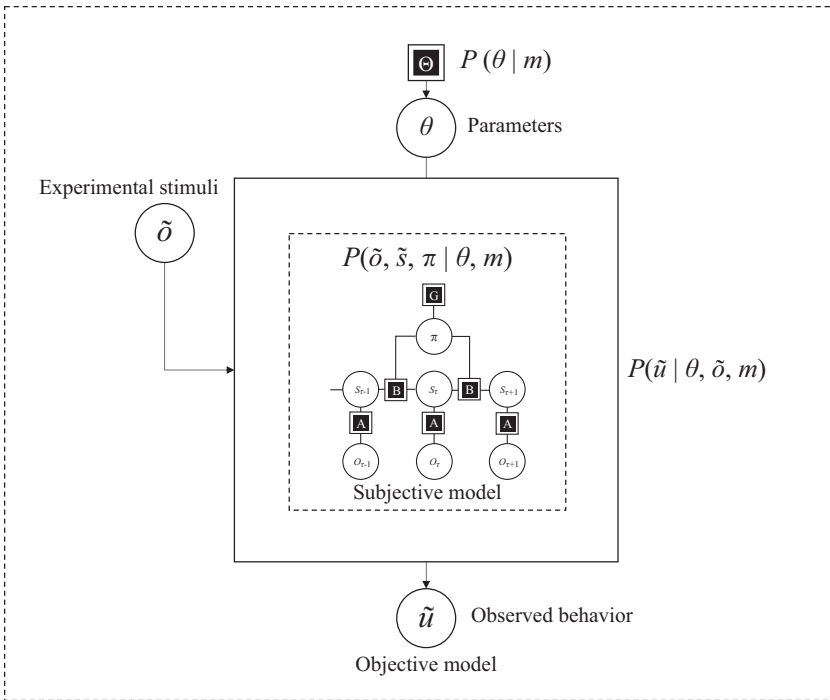
Broadly speaking, there are two (related) reasons for fitting a computational model to observed behavior. The first is to estimate parameters of interest from that model that best explain the behavior of a specific subject or group of subjects. This is useful in characterizing subjective behavior in terms of the computations that generate it, a process known as *computational phenotyping* (Montague et al. 2012, Schwartenbeck and Friston 2016, Friston 2017). Computational phenotypes may be used in combination with other measures (e.g., to establish links between neuroimaging findings and function) or may be used alone in forecasting behaviors in other settings (e.g., following a therapeutic intervention).

The second reason is to compare alternative hypotheses, expressed as models, that represent different explanations for a behavioral phenomenon (Mirza et al. 2018). These two agendas—parameter estimation and model comparison—map to one side of Bayes' theorem. Parameter estimation is the process of finding the posterior probability, under a model, of a parameter setting. Model comparison rests on finding the marginal likelihoods (i.e., evidence) for each model. To recap, Bayes' theorem is

$$\underbrace{P(u|\theta, m)}_{\text{Likelihood}} \underbrace{P(\theta|m)}_{\text{Prior}} = \underbrace{P(\theta|u, m)}_{\text{Posterior}} \underbrace{P(u|m)}_{\text{Evidence}}. \quad (9.1)$$

The right-hand side deals with the posterior probability of parameters ( $\theta$ ) given behavioral data ( $u$ ) under a model ( $m$ ) and the model evidence, and the left-hand side tells us what we need to specify for our model: we need prior beliefs about our parameters of interest and a likelihood function.

Importantly, while we appeal to the same Bayesian inference scheme as used in previous chapters, our purpose is different here. This rests on the fact that there are two inference processes going on (figure 9.1). The first is that creatures use their model of the processes generating their sensory data to draw inferences about their world (and about how to act). This has been the focus of the preceding chapters. The second is that we as objective



**Figure 9.1**

Relationship between the subjective and objective models of meta-Bayesian inference. *Inner dashed box:* Subjective model assumed to be used by an experimental subject. This could be a POMDP model as illustrated or some other form of model. The important features are that it depends on parameters ( $\theta$ ) whose value we do not know and that it generates sensory data ( $o$ ). *Outer dashed box:* Experimenter’s objective model ( $m$ ) includes prior beliefs about the parameters and predicts the behavior ( $u$ ) we would expect on presenting experimental stimuli (sensory data from the subjective model’s perspective). Crucially, the likelihood distribution of the objective model depends on the subjective model. This means we evaluate the likelihood of parameters taking a particular value as follows. First, we incorporate the parameters in the subjective model. We then use the Active Inference schemes described in previous chapters to solve this model, presenting our experimental stimuli as sensory data, and infer a distribution over the most probable course of action. Finally, we evaluate the probability of the observed actions or choices, given this distribution. This is the likelihood of observed behavior given parameters and stimuli—i.e., the likelihood distribution in the objective model.

scientists observe the creature's behavior and seek to draw inferences about the (*subjective*) generative model it is using by inverting our own (*objective*) generative model. The implication here is that we are drawing inferences about an inferential process—sometimes referred to as “meta-Bayesian” inference (Daunizeau et al. 2010).

More formally, this approach defines the likelihood distribution in terms of the solution to an Active Inference problem. By using a given parameter setting, we can simulate behavior under Active Inference and quantify the likelihood that a series of actions would have been taken. Equipped with prior beliefs about the value of these parameters, we have a generative model of how a creature uses its generative model to produce actions. While our focus is on Active Inference (and discrete-time models specifically), the generic methods used here may be used with any arbitrary likelihood function. Other normative models of behavior (such as those used in reinforcement learning) may be substituted in place of the Active Inference models.

The following sections unpack an example of a generic inference scheme that may be used for meta-Bayesian inference (namely, variational Laplace) and the use of hierarchical models for model comparison. We then provide a simple recipe for model-based data analysis and finally review a key example of this procedure. It is important to emphasize that understanding the technical details is not required to use these methods effectively; thus, readers uninterested in these details are invited to skip sections 9.3 and 9.4.

In brief, the basic idea is to evaluate the likelihood of any observed set of choices, given the unknown parameters of interest—namely, the parameters of a subject's prior beliefs. We then combine this likelihood with our objective prior over those parameters to evaluate the posterior over the subject's priors, in the usual way. If we have several subjects, these posteriors can be combined to make inferences about group or between-subject effects, using parametric empirical Bayes (PEB). The requisite likelihood is simply the probability of sampling the observed sequence of choices, under the subject's posterior beliefs about action. These posterior beliefs depend on what the subject sees (i.e., cues or stimuli) and her prior beliefs—and are evaluated in a straightforward way by solving the appropriate Active Inference scheme. Note that we are using Bayesian procedures twice: first to evaluate the subject's posterior beliefs about action, and second to evaluate our posterior beliefs about the unknown priors that characterize the subject. We now rehearse the various parts of this meta-Bayesian procedure.

### 9.3 Variational Laplace

Variational Laplace is an inference scheme based on the same principles as predictive coding (Friston, Mattout et al. 2007). However, it may be used for more generic likelihood functions than those encountered earlier—which were defined as Gaussian. We will start this section with an overview of the likelihood function  $\mathcal{L}(\theta)$  of interest here. This should give the probability of actions for an Active Inference scheme with a generative model with parameters set at value  $\theta$ . The actions selected depend on the observations made:

$$\begin{aligned}
 \mathcal{L}(\theta) &= \ln P(\tilde{u} | \theta, m, \tilde{\delta}) \\
 P(\tilde{u} | \theta, m, \tilde{\delta}) &= \tilde{u} \cdot \sigma(\theta_\alpha \ln \tilde{\mathbf{u}}) \\
 \tilde{\mathbf{u}} &= \boldsymbol{\pi} \cdot U \\
 \boldsymbol{\pi} &= \arg \min_{\boldsymbol{\pi}} F
 \end{aligned}
 \tag{9.2}$$

Unpacking this, the first term gives the log likelihood of an observed sequence of actions ( $\tilde{u}$ ) as a function of parameters ( $\theta$ ), the model ( $m$ ), and a sequence of stimuli ( $\tilde{\delta}$ ) presented during a real experiment. The probability of these actions is found by using the parameters to set the prior beliefs in a POMDP model of the sort described in chapter 7. We can then solve the POMDP as described in chapters 4 and 7, forcing the simulation to take the observed action sequence and presenting it with the same experimental stimuli. As we described in preceding chapters, this involves computing the beliefs ( $\boldsymbol{\pi}$ ) a synthetic subject holds about the policy or course of action she chooses to pursue. This minimizes the free energy ( $F$ ) associated with her generative model of the world. We can then take these beliefs and calculate the average probability of pursuing an action sequence. This requires us to distribute the probability for each policy over the actions implied by that policy (indexed by an array  $U$ ). Finally, a softmax temperature parameter ( $\theta_\alpha$ ) is applied to account for randomness (*shaky-handedness*) in behavior not accounted for by the model. If this softmax parameter is one, we are effectively assuming that the subject samples her actions from posterior beliefs about her actions; sometimes, this is called matching behavior. Alternatively, if the softmax parameter is very large, the action emitted is the action with the greatest subjective posterior—that is, the subject always chooses the most likely option. This softmax parameter can itself be estimated.

The result is the probability of the actions under the model, given a sequence of stimuli and parameters—that is, a likelihood of behavioral

data given a model. Equipping the objective parameters with Gaussian priors<sup>1</sup> ( $\theta \sim \mathcal{N}(\eta, \Pi^{(1)})$ ), we can use the Laplace assumption to express a (free energy) approximation to model evidence:

$$\begin{aligned} \ln P(\tilde{u}|m, \tilde{\delta}) &\approx \mathcal{L}(\mu) + \frac{1}{2} \left( \varepsilon \cdot \Pi^{(1)} \varepsilon + \ln \left| \nabla_{\mu\mu} \mathcal{L}(\mu) - \Pi^{(1)} \right| \right) \\ \varepsilon &= \eta - \mu \\ \mu &= \arg \max_{\mu} \left\{ \mathcal{L}(\mu) - \frac{1}{2} \left( \varepsilon \cdot \Pi^{(1)} \varepsilon + \ln \left| \nabla_{\mu\mu} \mathcal{L}(\mu) - \Pi^{(1)} \right| \right) \right\} \end{aligned} \quad (9.3)$$

Equation 9.3 is the same as that unpacked in box 4.3 (generalized to a multidimensional parameter space), but here we have substituted an explicit form for the posterior covariance and assumed a normally distributed prior. In chapter 4 and in the applications in chapter 8, we ignored the terms in equation 9.3 that did not depend on the mode. However, it is important to include these here when we consider model comparison problems.

To find the value of  $\mu$  that maximizes the last line of equation 9.3, we perform a gradient ascent. Under quadratic assumptions, this reduces to the following:

$$\dot{\mu} = \nabla_{\mu} \mathcal{L}(\mu) + \Pi^{(1)} \varepsilon \quad (9.4)$$

While an explicit form for the gradient of the log likelihood used here may not be available, finite difference methods<sup>2</sup> may be used to calculate a reasonable numerical approximation. These may also be used to find the posterior precision, which is the second derivative (or Hessian) of the negative log likelihood plus the prior precision. Equation 9.4 is the simplest form of update, but often more sophisticated methods based on the local curvature are used.

## 9.4 Parametric Empirical Bayes (PEB)

The variational Laplace procedure in the previous section lets us draw inferences about, and quantify the evidence for a model of, choice behavior. This enables us to computationally phenotype an individual and to compare alternative hypotheses about that individual. However, the interesting questions often lie at a group level. For example, we might be interested in how a parameter—such as the precision of prior preferences—varies with age. To answer this question, we can use the approach of section 9.3 to fit

models to the behavior of individual participants with a range of ages. We then formulate a general linear model that generates the parameter of interest, taking age into account:

$$P(\theta|\beta, X) = \mathcal{N}(X\beta, \Pi^{(2)}) \quad (9.5)$$

Here,  $X$  is a matrix whose columns are alternative explanatory variables and whose rows indicate each participant. The first column of  $X$  typically comprises a matrix of ones (to indicate the effect of the mean parameter over subjects). The second column, in our example, might be the age of each participant. The  $\beta$  vector indicates the size of effect of each of the explanatory variables in  $X$ . The first element of  $\beta$  is then the average value of the precision (or any other parameter), while the second is the effect of age on precision. This value is the slope of the line in a plot of age ( $x$ -axis) against predicted precision ( $y$ -axis). There may be an arbitrary number of columns of  $X$ , with an arbitrary number of elements in  $\beta$ .

Once we have fit the model expressed in equation 9.5, supplemented with priors for the  $\beta$  values, we can ask questions about the role of the explanatory variables. For example, we can ask whether age has an effect on the precision of prior preferences by comparing the evidence for a model in which the second element of  $\beta$  is allowed to deviate from zero with the evidence for a model with a precise belief that it is zero. Practically speaking, this can be done without multiple model inversions through use of Bayesian model reduction (Friston, Parr, and Zeidman 2018).

## 9.5 Instructions for Model-Based Analysis

In practice, we follow the steps outlined below to analyze empirical choice behavior using active inference (Schwartenbeck and Friston 2016). These refer to the relevant routines available in the SPM12 Matlab package.

1. **Collect behavioral data**, including the choices made and the sensory input available to the person making that choice. In addition, collect data of interest for second-level, between-subject analysis (e.g., whether the subject is a patient or a control subject, relevant demographic information, and so on).
2. **Formulate a POMDP model** as in chapter 7. This should be a function that takes parameters as inputs and outputs a fully specified (but not yet solved) POMDP.

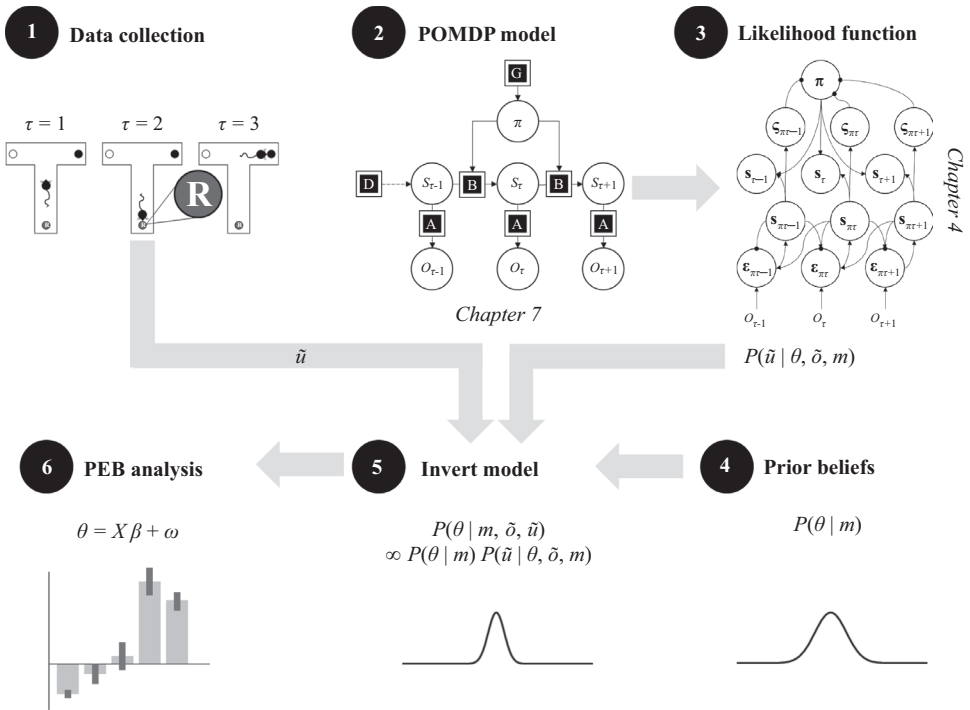
3. **Specify a likelihood function** (i.e., equation 9.2). This tells us how the model should be used to calculate a likelihood. This typically calls a POMDP solver (like the `spm_MDP_VB_X.m` routine) to simulate behavior and quantifies the likelihood of observed actions.
4. **Specify prior beliefs** about the parameters in terms of expectations and precisions. Often these will be centered on zero, with precisions reflecting plausible ranges.
5. **Solve for posterior probability and model evidence.** This uses a standard inference scheme such as the variational Laplace procedure outlined above (equation 9.4). The `spm_nlsi_Newton.m` routine will do this automatically.
6. **Perform group-level analysis.** This typically makes use of PEB, which treats the estimated parameters for each individual as if they were generated by a second-level model. This allows us to test hypotheses about the causes of those parameters. Practically, this may be performed using the `spm_dcm_peb.m` routine. Alternative analyses include standard statistical tests of association between the inferred parameters for each subject and other subject-specific measures. For example, a canonical variates analysis may be used to assess the relationship between questionnaire scores and inferred parameters.

Figure 9.2's summary of these instructions are based on the behavior of the rat in the T-maze task described in chapter 7. First, we place a rat in a T-maze with a rewarding stimulus in either the left or right arm and an informative cue in the central arm. We then record the sequence of actions taken by the rat. This procedure may be repeated over multiple trials to record learned behavior, and it may be repeated for multiple different rats under different interventions (e.g., pharmacological or optogenetic).

Once these behavioral data have been obtained, we need a likelihood function that lets us quantify the probability of behavior (for a given rat in a given condition) under specific parameter settings. We can do this by formulating the POMDP model we considered in chapter 7. This must be parameterized in terms of the parameters whose likelihood we seek to find. For example, if we wanted to assess the precision associated with preferences, we might include a log scale parameter that makes the preference distribution more or less peaky.

Having set up the generative model (from the perspective of the rat), we can automatically solve the POMDP using the belief-update equations





**Figure 9.2**

Roadmap of the six-step inversion procedure for model-based data analysis, as outlined in the main text (with reference to the chapters where more detail may be found). The arrows indicate the dependencies between each part of the process. The POMDP model must be defined for the likelihood to be evaluated. The model inversion requires collected data and the likelihood and priors; the PEB analysis cannot take place until after model inversion for each subject. Steps 4 and 5 schematize the update from a prior distribution over parameters under a model to a posterior distribution. The evidence and posterior from each model can then be combined using PEB to find posterior densities (shown as expectations with accompanying credible intervals) for the  $\beta$  coefficients of a linear model predicting these parameters.

in chapter 4. This lets us calculate the probability of the data (i.e., the sequence of arms visited) conditioned on the model with the (preferences) scale parameter at a particular value. Combining this likelihood with our prior completes the specification of a generative model for behavior (from the perspective of the scientist). This may be solved using variational Laplace to find a posterior probability distribution over the scale parameter for each rat.

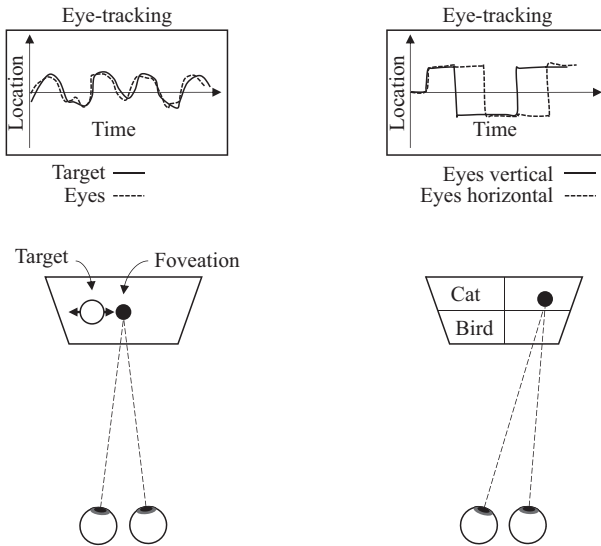
In practice, before analyzing actual data, we may want to check the face validity of the POMDP model by using it to generate fictive data—and considering whether they are qualitatively plausible given the problem at hand. A second sensible test for the model is *parameter recovery*. This entails generating fictive data under some (known) parameterization to see whether these parameters can be recovered on inverting the model. This is useful to verify whether some parameters (or their combinations) are possible to recover (i.e., identifiable).

Finally, we can construct a design matrix for a linear model, each row of which represents a computational phenotype (e.g., a posterior density over each subject's preferences), with columns representing different attributes of those subjects. These attributes are variables that could explain differences in a rat's preferences. In addition to a column of ones indicating the average preferences over all rats, these will include things like their age, whether a drug has been administered, and so on. With this model of between-subject effects, we now perform a PEB analysis to assess the contribution of these explanatory variables to prior preferences.

## 9.6 Examples of Generative Models

In this section, we leverage two examples in the literature illustrating the use of continuous and discrete generative models. First, we briefly overview the methods used by Adams, Aponte et al. (2015) and Adams, Bauer et al. (2016) (hereafter in this section, Adams et al.) to model smooth pursuit eye movements as a way of quantifying the precision parameters of each subject's generative models. An important aspect of this design was the simultaneous collection of electrophysiological data (via magnetoencephalography) that enabled the authors to ask questions about the neurobiological substrates of precision or confidence encoding. We then turn to an analysis of saccadic eye movements by Mirza et al. (2018; hereafter in this section, Mirza et al.) formulated as a POMDP model. Each of the associated experiments is cartooned in figure 9.3. Our hope is that these examples will help readers understand how the generic methods outlined above can be used empirically to answer scientific questions.

In terms of the sequence of steps outlined in figure 9.2, Adams et al. collected data (step 1) from a task in which subjects had to maintain fixation on a moving visual target. The details are not important, but this task



**Figure 9.3**

Two experiments outlined in section 9.5. The details are not important but highlight where meta-Bayesian inference has been successfully exploited and the kinds of behavioral data it can be applied to. *Left*: Experiment of Adams et al., who measured smooth pursuit eye movements as subjects tracked a moving target. *Right*: Experiment from Mirza et al., who measured saccadic eye movements during an exploration task. The visual display was divided into four quadrants, two of which included stimuli (cat and bird). Different scene categories involved different configurations of stimuli, meaning participants had to select which quadrants to foveate to gain sufficient information to categorize the scene. The Adams task (*left*) generates continuous eye-tracking data, while the Mirza task (*right*) leads to a sequence of fixations and may be discretized. These are the behavioral data ( $u$ ) from step 1 in figure 9.2.

comprised two conditions. In the first, the target moved according to a predictable sinusoid. In the second, it followed the same trajectory with additive Gaussian noise. The data collected included the eye-movement trajectories. The authors formulated a *subjective* model (step 2). Unlike the POMDP model shown in figure 9.2, they opted for a continuous model of the sort described in chapter 8. In brief, the model predicted proprioceptive and visual input from the eyes, where the fixation point was assumed to be attracted to the target location. The likelihood (step 3) is constructed using the (active) predictive coding schemes outlined in chapter 4. This quantifies the probability of the actions (eye movements) under a set of

log precision parameters and the model set out in step 2. Moving on to step 4, the authors specified prior beliefs as normal distributions over the log precisions. They inverted the model (step 5) to find posterior distributions over these precision parameters. Step 6 did not use a PEB analysis but used the neuroimaging data collected concurrently with the behavioral task. The authors used dynamic causal modeling to estimate the gain of superficial pyramidal cells in the primary visual cortex. This means they had estimates of precision and synaptic gain for each subject. This allowed the authors to perform a group-level analysis by assessing the correlation between parameters of the subjective model and their biological substrates. Their demonstration of this correlation provides an important example of how Active Inference formulations of behavior let us ask (and answer) questions about the relationship between belief updating and neurobiology.

In our second example, Mirza et al. used the POMDP formulation of Active Inference to address the role of information gain in driving human behavior. Again, we unpack this in terms of figure 9.2's steps. Mirza et al. collected behavioral data (step 1) while subjects performed a visual foraging task. Here, the aim was to classify a visual scene into one of several groups. Each element of the scene was only revealed once subjects fixated those locations; this meant multiple fixations were required to acquire enough evidence for a given scene category. The data collected by the authors included the sequence of saccades (fast eye movements) performed. The model (step 2) used was a POMDP model described in Mirza et al. (2016) that predicted discretized proprioceptive, visual, and feedback outcomes, conditioned on the current fixation location and the scene category. Preferences (see chapter 7) were placed over the feedback outcome such that the model anticipates (and thus prefers) being correct in the categorization. The likelihood function (step 3) was obtained by solving the model using the scheme outlined in chapter 4 under different parameter settings. The parameters in question included (among others) a log scaling parameter for the precision of the preference distribution. The authors specified prior distributions (step 4) over the log scaling (and other parameters) and inverted the model for each subject (step 5). They used the log evidence estimated for each subject to assess the evidence for models that did or did not motivate behavior using the epistemic component of the expected free energy, finding greater evidence for those models that included epistemic affordance in all subjects. They then employed a PEB analysis (step 6) to

assess changes in prior beliefs for subjects over the course of multiple trials, finding evidence in favor of changes in belief parameters (i.e., active learning). Finally, they used a canonical covariates analysis to assess the relationship between linear combinations of the phenotypic variables estimated for each subject (e.g., precision of preferences) and linear combinations of performance measures (e.g., percentage correct and reaction time).

## 9.7 Models of False Inference

Men in general are quick to believe that which they wish to be true.  
—Julius Caesar

Given the relevance of these methods for fields like computational psychiatry (Friston, Stephan et al. 2014), we end with an overview of false inference, which is central to the notion of psychopathology as a failure of belief updating. One benefit of using an inferential framework like Active Inference is that it simultaneously addresses multiple dimensions of psychiatric disorders, linking together maladaptive behavior (e.g., compulsions or addictions) and psychological-level (e.g., false beliefs) and biological-level phenomena (e.g., abnormalities of neuromodulators).

As we cannot do justice here to the extensive literature that uses Active Inference in the modeling of disease processes, this section provides the briefest of overviews to suggest a framework in which to think about computational pathologies. See table 9.1 for a nonexhaustive sampling of illustrative examples, which include models specified in discrete and continuous time (based on those in chapters 7 and 8, respectively). In our discussion, we will appeal to the structure of POMDP-like models; the principles that underwrite false inference in these settings are largely the same.

The hypothesis underlying inferential approaches (like Active Inference) is that psychopathological conditions may be conceptualized as disorders of inference. The term *disorder* does not necessarily imply that the inferential mechanism is flawed (e.g., generates incorrect posterior probabilities). In most of the studies reviewed in table 9.1, the inferential mechanism operates normally, but based on a flawed generative model (i.e., a generative model endowed with aberrant prior beliefs). This means that, ultimately, pathology is a consequence of aberrant prior beliefs—and one can recover these priors using the model-based data analysis outlined in this chapter.

Table 9.1

## Computational pathology

| Pathology                                | Sources  | Notes   |
|--|--|---|
| Addiction, impulsivity, and compulsivity | FitzGerald, Schwartenbeck et al. 2015<br>Schwartenbeck, FitzGerald, Mathys, Dolan, Wurst et al. 2015<br>Mirza et al. 2019<br>Fradkin et al. 2020 | Addiction is an important example of behavior that appears aberrant but can be framed as optimal inference under the right sort of generative model. Work by Schwartenbeck et al. illustrated this using a limited offer task wherein participants are more or less confident about whether they will receive a reward on waiting. Low confidence leads to compulsive behavior of the sort associated with addiction. Subsequent work on this theme looks at the prior beliefs associated with more or less impulsive behavior, using the patch-leaving paradigm, and examines the role of attenuated prior precision in obsessive compulsive disorder. |
| Delusions                                | Brown et al. 2013<br>Friston, Parr et al. 2020   | Delusions, characterized by fixed, false beliefs, are simply articulated in Active Inference as precise posterior probability distributions in the absence of supportive evidence. If sufficiently precise, they will be fixed even in the face of (subsequent) contradictory evidence. The mechanisms underlying each delusion may be different. For example, failures of sensory attenuation may be central to delusions of agency. Recent work provides an example of a shared delusion ( <i>folie à deux</i> ), which depends on two agents—with no information—reaching a confident consensus about the state of the world.                        |
| Hallucinations                           | Adams, Stephan et al. 2013<br>Benrimoh et al. 2018<br>Parr, Benrimoh et al. 2018<br>Corlett et al. 2019  | These simulations rest on imbalances between prior and likelihood precisions. Overinterpretation of spurious sensory data due to a failure of attenuation of likelihood precision, or a failure to correct prior beliefs due to excessive attenuation, each offer mechanisms for false perceptual inference.  |

**Table 9.1**

(continued)

| Pathology                               | Sources   | Notes  |
|---|---|--|
| Interpersonal and personality disorders | Moutoussis et al. 2014<br>Prosser et al. 2018               | Interpersonal inference depends on having models about other people and how they may react to our decisions. This has prompted the development of models of trust games, which rely on interactions between two (or more) parties, and charity games. The latter have been used to reproduce the self-aggrandizing and remorselessness associated with psychopathy. These traits are simulated by modulating the degree to which beliefs about self-worth depend on decisions to be charitable versus selfish and sensitivity to the approval of others. |
| Oculomotor syndromes                    | Adams, Perrinet, and Friston 2012<br>Parr and Friston 2018a | In these papers, continuous-time generative models are employed to predict dynamic evolution of Newtonian systems. By rendering various aspects of the generative model conditionally independent of others, oculomotor syndromes such as internuclear ophthalmoplegias may be induced.  |
| Pharmacotherapy                         | Parr and Friston 2019b                                      | Given the associations we proposed between precision parameters and neurochemicals in chapter 5, it should be possible to simulate the consequences of pharmacological manipulation of these systems. This work illustrates the consequences of several synthetic pharmacological interventions on performance of an oculomotor delay-period task, providing a proof of principle that these methods can be used to simulate not only pathology but also the influence of therapeutics.  |
| Prefrontal syndromes                    | Parr, Rikhye et al. 2019                                    | These simulations set out a difference between medial and lateral prefrontal syndromes by attenuating the precision of transitions to impair the performance of a memory guided task (lateral) versus the precision of an interoceptive likelihood that determines motivation to engage in the task (medial).  |

(continued)

**Table 9.1**  
(continued)

| Pathology                            | Sources  | Notes   |
|--------------------------------------|--|---|
| Visual neglect                       | Parr and Friston 2017a   | Inattention to the left side of space may be induced by several alternative lesioned priors. Among these, an increase in Dirichlet parameters for this side of space reduces the novelty associated with saccades to the left, increasing visual sampling of the right instead. Alternatively, setting preferences consistent with right-sided proprioceptive or visual outcomes or increasing habitual engagement in right-sided saccades reproduces qualitatively similar behavior.   |
| Disorders of interoceptive inference | Barrett et al. 2016<br>Allen et al. 2019<br>Maisto, Barca et al. 2019<br>Pezzulo, Maisto et al. 2019<br>Barca and Pezzulo 2020<br>Tschantz et al. 2021 | Simulations of interoceptive inference (or active inference in the interoceptive domain) suggest that imbalances between prior and likelihood precisions about (for example) cardiac or gastric signals can cause false beliefs about the internal state of the body, misperceptions of bodily symptoms, and psychosomatic hallucinations. Furthermore, they can have cascading effects on autonomic regulation and action selection, causing various types of maladaptive behavior, such as hypervigilance, excessive medicine use, and excessive food restrictions. |

Aberrant priors may be about states or precisions, or they may be structural priors about the form of the generative model. A useful way of thinking about the causes of pathological behavior is to think about the prior belief used for policies and about how each part of this may be disrupted to give rise to abnormal policy selection. Policy priors depend on the expected free energy, which itself depends on posterior beliefs, the potential for information gain, and prior preferences (C). Priors over policies may additionally be equipped with a fixed form term (E), representing habitual biases.

Taking each of these in turn: Posterior beliefs depend on priors and likelihoods. To form an aberrant posterior belief, one or both must be disrupted. Typically, this disruption takes the form of under- or overestimation of the



balance of precisions. Excessively high likelihood, compared to prior, precision leads to overinterpretation of (potentially noisy) sensory input. This leads to *overfitting* in the sense that unwarranted conclusions may be drawn from spurious data. If the balance is disrupted in the opposite direction, favoring confidence in the prior, internally generated percepts become resistant to conflicting sensory input. Both mechanisms have been associated with the development of hallucinations, and the two can coexist when hierarchical models are employed. Given the association of various precisions with neuromodulatory chemicals (see chapter 5), it seems sensible that conditions such as Lewy body dementia, where cholinergic signaling is impaired, and schizophrenia, with abnormalities of the dopaminergic system, present with hallucinatory phenomena—that is, false perceptual inference.

Next, we consider the role of information gain. Here, the precision of the likelihood and the precision of prior beliefs tell us the degree to which uncertainty is resolvable and the amount of uncertainty there is to resolve, respectively. The precision of the prior beliefs applies to either parameters of the generative model (i.e., influences novelty) or to the states (i.e., influences salience). Interpreting the parameters of conditional probabilities as synaptic efficacies or the precisions as synaptic gains suggests that synaptic disconnection syndromes may be thought of as disruption of one or both of these. Absent synapses cannot be modulated, so this is very much like having extremely confident prior beliefs about a conditional probability, as new data cannot update the associated efficacy. This has important implications for the potential information gain of different policies, as has been exploited in modeling sensory neglect syndromes.

Finally, the preferences and policy priors provide a clear influence over behavior. These could underwrite the development of addictive habits or the apathy associated with various psychiatric and neurological syndromes (Hezemans, Wolpe, and Rowe 2020). In summary, defective prior beliefs at various places in the generative models described above provide a functional or teleological explanation for pathological behavior.

## 9.8 Summary

In this chapter, we outlined an approach that uses the theoretical models described in previous chapters to pose questions to empirical data. This lets us use Active Inference as a noninvasive tool to probe the computational

processes that individuals use to make decisions. We have focused on a few simple examples. However, Active Inference–based models have been developed for more realistic and complicated tasks (Cullen et al. 2018) designed to evoke richer behavior for computational phenotyping. In addition to setting out a six-step process for model-based analysis, we highlighted two examples of the use of these methods. These bring out key variations in how this may proceed, including the kinds of behavior measured (smooth trajectories or discrete choices), the choice of model (continuous or discrete), and the different scientific questions being asked. The last of these is the most important, as it determines the preceding choices. We saw the use of computational phenotyping in combination with neuroimaging (Adams, Bauer et al. 2016) to ask questions about the relationship between synaptic gain and precision. In addition, we saw how model inversion may be used to assess the contributions of alternative behavioral drives and predictors of performance (Mirza et al. 2018). Ultimately, the six steps in figure 9.1 provide a generic method for designing experiments to noninvasively interrogate the implicit generative models people (or other animals) use to drive behavior. This offers an opportunity to answer questions about the function of the nervous system in health and disease.

This is a section of [doi:10.7551/mitpress/12441.001.0001](https://doi.org/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](https://doi.org/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://lcn.loc.gov/2021023032>