

# A tutorial on Bayesian models of perception

Benjamin T. Vincent

*School of Psychology, University of Dundee, UK*

---

## Abstract

The notion that perception involves Bayesian inference is an increasingly popular position taken by many researchers. Bayesian models have provided insights into many perceptual phenomena, but their description and practical implementation does not always convey their theoretical appeal or conceptual elegance. This tutorial provides an introduction to core concepts in Bayesian modelling and should help a wide variety of readers to more deeply understand, or to generate their own Bayesian models of perception. Core theoretical and implementational issues are covered, using the 2 alternative-forced-choice task as a case study. Supplementary code is available to help bridge the gap between model description and practical implementation.

*Keywords:* Bayesian network, Bayesian inference, alternative forced choice, MCMC, ideal observer, psychometric function, probabilistic generative model.

---

## 1. Introduction

The question of how our perceptions derive from sensory observations of the world has been long debated. One popular position is that our perceptions arise from unconscious inferences (Helmholtz, 1856), with Gregory (1980) suggesting that perceptions are hypotheses about the world. Under this view, in contrast to the claims of Gibson (2002), sensory data are ambiguous and prior knowledge or Gestalt-like assumptions about the world are required to make accurate perceptual inferences (Pizlo, 2001). Because Bayesian inference provides a recipe for optimally combining prior knowledge

---

*Email address:* [b.t.vincent@dundee.ac.uk](mailto:b.t.vincent@dundee.ac.uk) (Benjamin T. Vincent)

*URL:* <http://www.inferencelab.com> (Benjamin T. Vincent)

with new sensory observations, it has the potential to provide significant insight into perception (e.g., [Knill and Richards, 1996](#); [Kersten et al., 2004](#)). Determining the extent and limits of these insights is, and will be, an important task. In doing so, it is clear from recent discourse (e.g., [Bowers and Davis, 2012a](#); [Griffiths et al., 2012](#); [Bowers and Davis, 2012b](#)) that clarity is required in conveying the details of, and theoretical claims made by, Bayesian explanations of behaviour. The main aim of this tutorial is to help facilitate this broad accessibility. This is achieved by presenting a series of models of an important task in perceptual research, the 2-alternative-forced-choice task. In order to convey these theoretical claims, it is important to understand how these models work and how they are evaluated, and so this tutorial also acts as a primer to methods of Bayesian model evaluation. Sometimes, this understanding of the details of a model and how they are practically implemented is important to be able to understand and critique models, and so Supplementary Code is provided to help in this way.

To best understand the claims made by Bayesian explanations of perception it is first worth outlining how Bayesian methods are used in alternative contexts. The first is where a scientist may use Bayesian inference to evaluate a *data analysis model* (Figure 1a). Examples would be general linear models such as regression or ANOVA. These can serve as descriptions of the data, no claims about the perceptual, cognitive, or neural processes that gave rise to the data are being made. [Kruschke \(2015\)](#) provides a comprehensive overview of using Bayesian methods for a wide variety of data analysis models.

Under a second approach, a candidate model is constructed to describe processes that give rise to behavioural data (Figure 1b). While Bayesian methods are used to evaluate the models, under this approach there is no claim that Bayesian processes are occurring within the observer themselves. This is perhaps the most common modelling approach where researchers propose explanations for behavioural data and evaluate these through formal modelling. Frequently, non-Bayesian methods are used in evaluating these models, see [Lewandowski and Farrell \(2011\)](#) for a comprehensive overview. However, it is becoming more common to describe and evaluate models using Bayesian methods. The text by [Lee and Wagenmakers \(2014\)](#) is an excellent resource for readers interested in constructing and evaluating a wide range of models of this class.

The third approach, which is the focus of this tutorial, goes further to propose that Bayesian processes occur within the brains of observers (Figure 1c). This has been termed by some the ‘Bayesian Brain’ hypothesis

(Knill and Pouget, 2004; Doya et al., 2007; Colombo and Seriès, 2012), and has been applied across a wide range of domains in psychology. In the perceptual domain, the core assertions are that: a) an observer has an internal mental model which represents the processes that gave rise to their sensory observations, b) that observers conduct Bayesian inference using their mental model in order to infer probable states of the world from sensory data, and c) that observers have prior beliefs over states of the world.

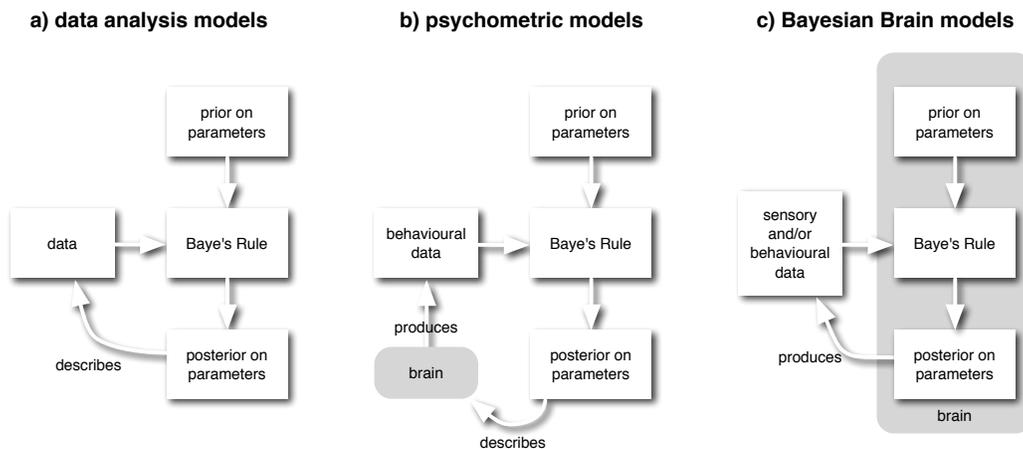


Figure 1: Three contexts in which Bayesian modelling are used. Adapted with permission from [Kruschke \(2011\)](#).

### 1.1. Bayesian observer models

A forward generative model, in the context of perception being Bayesian, is an internal mental model which describes and simulates the processes taking place in the world that give rise to sensory observations (see Figure 2, middle). Forward models allow ‘what if’ questions to be asked; if the world was like this, how likely is it that a range of possible observations would be observed. Forward models can be formalised by a likelihood term describing the probability of observing some sensory data for a given state of the world,  $P(\text{data}|\text{world})$ . This concise term is a simple summary of what will turn out to be a much more elaborate set of descriptions or equations for each particular modelling context.

However, observers do not have access to the true state of the world, they have access to sensory observations of the world, and must infer likely states

of the world based upon that sensory data (see Figure 2, right). This is done by evaluating the the Bayesian posterior,  $P(\text{world}|\text{data})$ . This ‘inverse problem’ is not trivial to solve because many different possible states of the world could be consistent with the observed sensory data, most obviously apparent in the case of visual illusions (Gregory, 1980). Bayes’ equation specifies how to solve the inverse problem and make inferences about the world given the data available,

$$P(\text{world}|\text{data}) \propto P(\text{data}|\text{world}) \times P(\text{world}). \quad (1)$$

In words, the observer’s task of inferring beliefs about different possible states of the world given some data (the posterior) is conducted by multiplying the probability of observing that data for a given state of the world (the likelihood term) with a belief about the plausibility of that state of the world (the prior).

What has been described so far is one major component in Bayesian Decision Theory which describes the way in which observers can arrive at beliefs about the world. The next component, which will not be discussed in this tutorial, translates this belief about the world into an action, taking into account the possible costs or benefits of each action dependent upon the true state of the world. Interested readers are referred to North (1968) and Kording (2007) for introductions and to Maloney (2002) and Maloney and Zhang (2010) for in depth information.

While it is the multiplication of prior and likelihood that is the essence of what it means to conduct Bayesian inference, this simple mathematical operation is not the core theoretical appeal. Rather, if we accept that we do not have direct access to the true state of the world, and the data are insufficient to provide this knowledge unambiguously, then Bayesian inference becomes appealing as *the only* principled solution to the inverse problem. Readers are referred to recent texts for wide ranging examinations of the promise and problems of Bayesian approaches applied to perception, cognition and action (Glymour, 2001; Frith, 2013; Hohwy, 2013).

### 1.2. *Optimal vs. suboptimal Bayesian observers*

We can differentiate two subclasses of Bayesian models based upon whether they are optimal or suboptimal (Ma, 2012). The Bayesian optimal observer approach would claim that observer’s mental models of the world are veridical, and that their prior beliefs are matched to the statistics of the environment. In hypothetical ideal observer models, observers are given precise

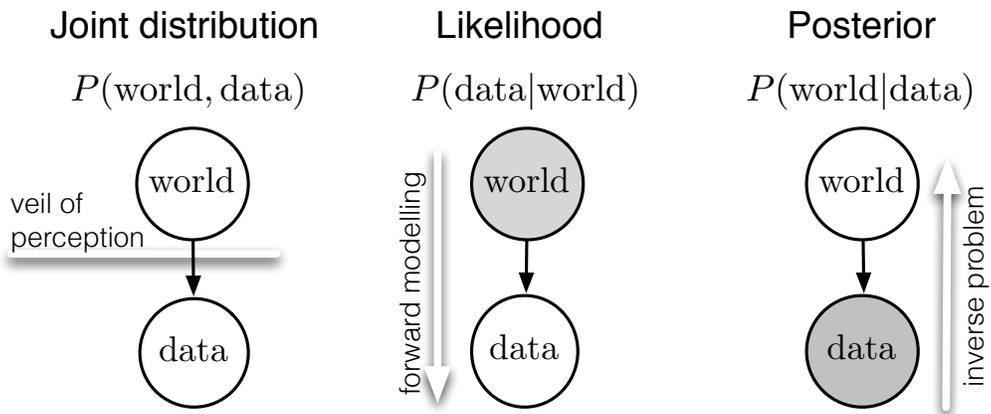


Figure 2: Schematic representation of Bayesian inference. Everything we know about the probabilistic generative model is captured by a joint probability distribution (left). The generative nature of the model can be used to simulate (deduce) data observations given an assumed state of the world, and to calculate the likelihood of observing some data given this state of the world (middle). The model also allows the perceptual process to infer a distribution of states of the world, given some observed sensory data (right). One example would be a forward model (middle) that describes the process giving rise to sensory data for a given world state  $P(\text{data}|\text{world})$ . Shaded nodes represent observed, known quantities.

knowledge of certain aspects of the world in order to calculate the theoretical maximum performance. While these ideal observer models provide compelling accounts of behaviour in perceptual similarity judgements (Ma et al., 2012), covert attention tasks (reviewed by Vincent, 2015), visual working memory tasks (Sims et al., 2012), visual discrimination (Geisler, 1989), and perceptual research in general (Geisler, 2011), they are not always serious explanations of human behaviour in all contexts. Instead, they act as important baselines, allowing the researcher to propose suboptimal observer models (Geisler, 2003).

Another approach is that observers are Bayesian yet suboptimal (Beck et al., 2012; Acerbi et al., 2014). The first reason for this suboptimality could be that an observer's internal mental model of the world is mismatched with the actual generative process (the real world) which gave rise to their sensory observations. Beck et al. (2012) argue that this is likely to be the case in many real world tasks where reality is too complex to formulate a veridical generative model. Low-level perceptual tasks, however, are generally simple and it is perfectly reasonable to assume observers have accurate generative models. Does this mean Bayesian Brain modelling is only usefully applied to low-level perceptual tasks or phenomena? No, it is the task of the scientist to infer how people may approximate the true generative process by a simpler internal generative model.

A second reason for observers being suboptimal could be that their knowledge is mismatched with the true state of the world. Real people have uncertainty about the state of the world, and this is captured by their prior beliefs about the world. How well these priors reflect the true uncertainty about the world is another question (Vincent, 2011; Fennell and Baddeley, 2012).

### *1.3. Inferences made by modellers*

While modelling can be a complicated endeavour, it is possible to define some broad categories. This tutorial provides an overview of these broad categories and acts as an introductory exposure to each. There are a number of accessible texts which provide in depth coverage of these topics in the context of data analysis (Kruschke, 2015; Gelman et al., 2013) and cognitive modelling (Lewandowski and Farrell, 2011; Lee and Wagenmakers, 2014), but this tutorial is focussed on applying these approaches in the context of Bayesian Brain models.

**Describing the generative probabilistic model.** The first task of modelling from a Bayesian perspective is to define a probabilistic generative model. This, as the name implies, is a model of the processes which generated observed data, and provides a joint probability distribution which can then be used in a variety of inference steps, such as the three listed below.

**Generating simulated observer behaviour.** One of the main advantages of quantitative modelling in general is to make reliable and accurate predictions of the logical consequences of a hypothesis. Trying to mentally simulate a model’s predictions can be difficult, error prone, and only qualitative (Farrell and Lewandowski, 2010). We can use the generative aspect of the probabilistic models to create simulated datasets and make predictions of the model.

**Parameter estimation/recovery.** Having acquired some behavioural data from a human observer one aim of an experimenter could be to infer the values of the model’s parameters that are most consistent with the data. This is commonly termed ‘fitting the parameters to the data’ but a more accurate term would be parameter estimation (Tarantola, 2004, 2006). That is, the Bayesian approach estimates a distribution of belief of how plausible a whole range of parameter values are, given the observed data. But how do we know that our parameter estimation process is reliable? As modellers, we can calculate an observer’s responses for a set of parameter values that we specify. We can then conduct parameter recovery, testing how well we can infer the known parameter values just based upon the observer’s responses. If parameter recovery does a bad job, then we will have to collect more data, or refine the model’s formulation so that the data can more accurately inform us of the true parameter values.

**Goodness of fit to data.** While the previous step of parameter estimation shows that we can infer a distribution of belief over parameter values, this does not yet inform us of whether the model is doing a good or a bad job of accounting for experimental data. We can do this by using the generative aspect of the model again by making model predictions where the parameters are constrained by the data. This step of *posterior prediction* allows us to plot the model’s predictions for a visual check.

#### 1.4. Practical evaluation of Bayesian models

There are a variety of approaches in how we might practically evaluate a probabilistic generative model (Jordan, 2004). The two methods we focus upon here will be grid approximation and MCMC sampling (see Figure 3)

and they have advantages and disadvantages depending upon the modelling context. Both methods involve calculating the joint probability of the model parameters for a given set of latent parameter values and observed experimental data. This is analogous to the non-Bayesian approach of evaluating the goodness of fit of a model to data (e.g., sum squared error). However, the Bayesian approach introduces priors over parameter values and replaces the sum squared goodness of fit term with a data likelihood (see [Myung, 2003](#), for a tutorial on likelihood estimation).

Grid approximation evaluates the model for a range of possible values of the parameters (Figure 3, left). The more parameter values evaluated, the closer the posterior will be approximated. MCMC sampling takes a different approach and instead attempts to draw many samples of the parameter values in proportion to the model's posterior probability (Figure 3, right). The more MCMC samples generated, the closer these samples will approximate the true posterior.

The Metropolis-Hastings algorithm is perhaps the simplest of many MCMC sampling strategies ([Cum and Greenberg, 1995](#)). While a deep understanding is not required, a basic familiarity is certainly advantageous. The posterior probability of the parameters given the data are estimated based upon an initial point in parameter space. The parameter space is explored by proposing a new set of parameters by sampling from a distribution centred on the current point in parameter space. Before each new proposed parameter vector is accepted into the list, acceptance criteria are defined based on whether the posterior probability of the proposed sample is higher or lower than the current sample. In the former situation, new samples are always accepted, in the latter case, new samples are probabilistically accepted. Doing this means that the number of samples returned from points in parameter space will be proportional to the posterior density (eg. Figure 3, right). Readers interested in learning more of the technical details are referred to the algorithmic implementation in the Supplementary Code, chapter 29 of [MacKay \(2003\)](#), and to [Kruschke \(2015\)](#) who describes more sampling algorithms.

### *1.5. Overview of the tutorial*

In the remainder of the tutorial, some of the core aspects of Bayesian modelling will be demonstrated using the case study of the 2 alternative forced choice (2AFC) task. Three different models of this task are described, and used in order to demonstrate important points about the formulation

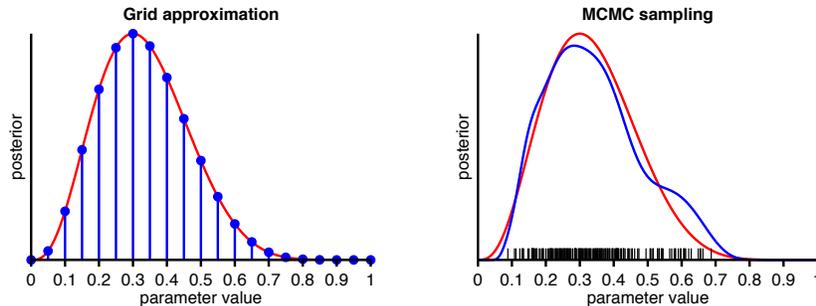


Figure 3: A true but unknown posterior distribution (red) can be approximated (blue) by evaluation over a grid of parameter values (left) or by drawing MCMC samples (right). Drawing more samples than the 200 shown will result in a more accurate approximation of the posterior.

and practical evaluation of Bayesian models of perception. Supplementary Code is available for readers to learn more about the practical evaluation of Bayesian models, and is available at <https://github.com/drbenvincent/bayesian2afc>.

**Model 1** shows how to formulate the SDT account of 2AFC in terms of a probabilistic generative model. Grid approximation is used to demonstrate how probabilistic generative models can be constructed and practically evaluated. MCMC methods build upon grid approximation and will become particularly useful for more complicated models with more parameters. However, we see that MCMC methods are approximate and necessitate some extra diligence.

**Model 2** formulates the SDT account (SDT) of 2AFC on a trial to trial basis. It models the stimulus generation process as well as the observer’s response selection. While the model is more complex, it allows modelling of suboptimal aspects of behaviour such as response errors.

**Model 3** is a Bayesian ideal observer of the 2AFC task, also formulated on a trial to trial basis. The nature of the Bayesian optimal aspect of the model becomes more apparent as it is a model of the causal generative structure of the 2AFC task environment. This forward model is then putatively used by observers to make inferences about the state of the world. Suboptimal Bayesian inference is demonstrated through an observer with incorrect priors of the state of the world.

## 2. The 2AFC case study

The 2AFC task has played a crucial role in the psychophysical study of perception. An experimental trial is simple: an observer indicates which of two locations contains a signal item as opposed to a noise item (see Figure 4, top). These items consist of simple visual features, such as the orientation of a line, the size of a dot, or the luminance of a patch. Each response will either be correct or incorrect, and the probability of responding correctly decreases as the signal and noise items become more similar. By repeating many trials while varying the difference between signal and noise items, a psychometric function can be measured (see Figure 4, bottom). One reason for this is that noise will corrupt the sensory measurements of the actual stimuli being displayed, such that sometimes the identity of a signal and noise items become confused and an incorrect localisation results (see Figure 5).

The psychometric function establishes the relationship between external stimuli and behavioural responses. However, the problem that has to be solved by an observer is how to infer the state of the world from the proximal sensory measurement. How is this done, and what processes are involved? Proposing models that relate the external stimuli to behavioural responses allows us to make inferences about the internal information processing mechanisms and the observer’s internal percept. The aim here is to utilise a simple yet fundamentally important task to highlight some core theoretical and practical issues associated with Bayesian explanations of perception.

One popular approach of modelling an observer’s performance in the 2AFC task is SDT (Green and Swets, 1966). This is a manifestation of Statistical Decision Theory (Maloney and Zhang, 2010) which is closely related to Bayesian Decision Theory, where in the latter, observers have priors over states of the world.

This simple 2AFC task is distinct from a number of other similar experimental paradigms. An alternative paradigm allows speed accuracy tradeoffs to be investigated by keeping the stimuli available until the observer responds. But in the simple version considered here, speed accuracy tradeoffs are eliminated by only displaying the stimuli for a fixed and brief duration (typically around 100ms) and by instructing observers to maximise the accuracy (not the speed) of their response. The 2AFC task is also distinct from the yes/no task which is commonly associated with SDT. The yes/no task requires an observer to indicate if a signal item was present or absent, whereas the 2AFC task requires the observer to indicate the location of a

## 2-AFC spatial localisation: where was the signal?

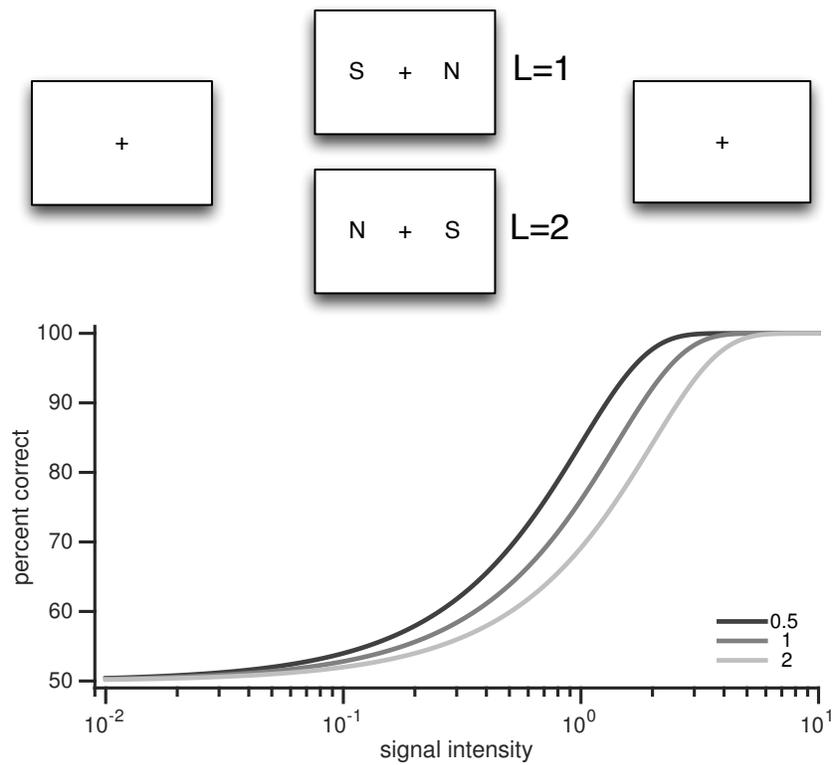


Figure 4: Overview of the spatial 2-AFC task. After central fixation, 1 of 2 stimulus displays is presented for a brief duration (top) before the observer gives an unspeeded response of whether they inferred the signal to be in location 1 or 2. By repeating many of these trials, and by varying the signal intensity, a psychometric curve can be established (bottom). Three psychometric curves are shown for different levels of observation noise variance (see legend).

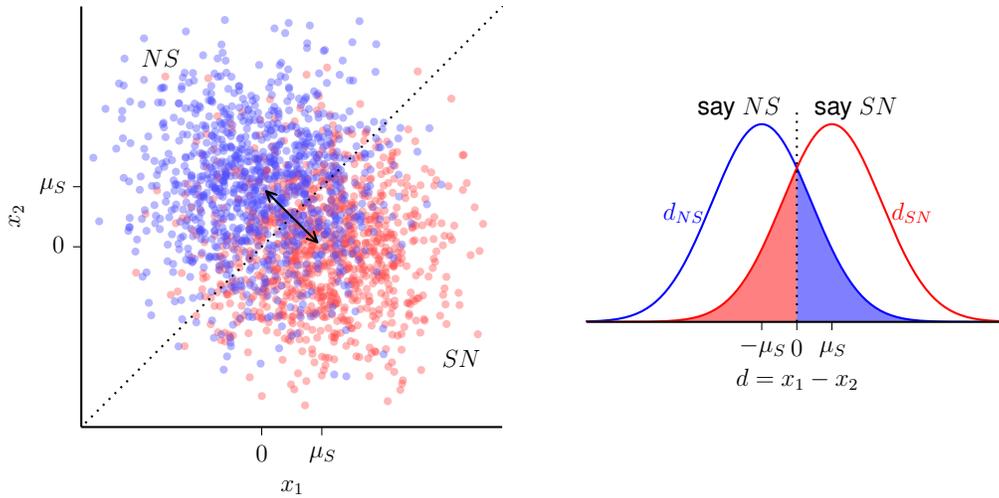


Figure 5: Intuition behind SDT models of performance in the 2AFC task. Each point (left) shows a simulated pair of sensory observations for an experimental trial. Red and blue points label the signal location, which is unknown to observers. Dashed lines show the observer’s decision threshold. The distributions of decision variables are shown (right), with incorrect decisions shown by shaded regions.

signal which is present on each trial. A Bayesian data analysis model of the yes/no task has been presented by [Lee \(2008\)](#), and [Vincent \(2015\)](#) reviews similarities and differences of Bayesian Brain models AFC and yes/no tasks.

Before modelling may commence, the 2AFC task needs to be described more formally. On each trial  $t$ , one signal item and one noise item will be displayed, with true feature values of  $\mu_S$  and  $\mu_N$ , respectively. The difficulty of the task will be related to the difference between signal and noise items, which we will call the signal intensity  $\Delta\mu = \mu_S - \mu_N$ . The stimuli on each trial can either contain a signal in the first position  $\langle SN \rangle$ , or in the second position  $\langle NS \rangle$ . The probability that the signal item will occur in location 1 ( $L_t = 1$ ) or 2 ( $L_t = 2$ ), is determined by a spatial prior probability such as  $\mathbf{p} = [0.5, 0.5]$ . If there was no uncertainty associated with making observations, then  $\mathbf{x} = \{\mu_S, \mu_N\}$ , or  $\mathbf{x} = \{\mu_N, \mu_S\}$ , however, we will assume there is normally distributed observation noise such that a single observation is described as  $x \sim \text{Normal}(\mu, 1/\sigma^2)$ , where  $\mu$  is either  $\mu_S$  or  $\mu_N$  depending upon the signal location. When the observation noise encoding precision  $1/\sigma^2$  is not infinite, i.e., the noise variance  $\sigma^2$  is greater than zero, the distribution

of observations  $\mathbf{x}$  (over many trials) may overlap (see Figure 5), and thus the performance of an observer to correctly indicate the target’s location will be less than 100%. This 2AFC task can be repeated in  $T$  trials over  $C$  signal level conditions such that we now have a list (vector) of signal intensity levels  $\Delta\boldsymbol{\mu} = \{\Delta\mu_1, \dots, \Delta\mu_C\}$ . Equations featuring bold face (such as  $\Delta\boldsymbol{\mu}$ ) refer to vectors over  $C$  stimulus intensity conditions, and non-bold face (such as  $\Delta\mu_n$ ) refers to a single signal intensity condition. To summarise,  $T$  trials are carried out at each of  $C$  signal intensity conditions, so a total of  $C \times T$  trials are conducted.

### 3. Model 1: SDT with Bayesian estimation

#### 3.1. Constructing the probabilistic generative model for the 2AFC task

While the experimenter may know the signal location,  $\langle SN \rangle$  or  $\langle NS \rangle$ , an observer does not. The proposed optimal way for an observer to decide (assuming equal encoding precision of signal and noise items) is to calculate a decision variable  $d = x_1 - x_2$ , and to respond that the target is in location 1 if this  $d > 0$ . This is equivalent to the observer responding to the location with the highest valued sensory observation, hence it being called the Max rule (Figure 5; Green and Swets, 1966; Wickens, 2002; Kingdom and Prins, 2009). The response threshold of zero is optimal under the assumption that the signal is equally likely to occur in location 1 and 2. Because of the experimenter’s knowledge, we can define  $x_S$  and  $x_N$  as the noisy sensory observation at the signal and noise locations, respectively. On each trial, the correct response will be given when  $x_S > x_N$  (or equivalently  $x_S - x_N \geq 0$ ), and so the probability of a correct response  $PC$  is simply the proportion of times that this occurs,  $PC = P(x_S - x_N \geq 0)$ . This can be calculated as

$$PC_c = \Phi\left(\frac{\Delta\mu_c}{\sqrt{2\sigma^2}}\right) \quad (2)$$

where  $\Phi$  is the cumulative standard normal distribution (Kingdom and Prins, 2009). This cumulative normal is a psychometric function.

So far the relationships between these variables have been deterministic, but the next part of the model examines the probabilistic relationship between the underlying probability of a correct response  $PC$  and the actual proportion of correct responses  $\frac{k}{T}$ , for a given signal intensity condition. The variable  $k$  is the number of correct responses out of  $T$  trials for each signal level, and the vector of correct responses over each signal intensity condition

is  $\mathbf{k} = \{k_1, \dots, k_C\}$ . Because each trial is a Bernoulli trial (a biased coin flip, with  $PC$  being the probability of a correct response), the number of correct responses  $k$  will be Binomially distributed, so the generative model for this task is

$$k_c \sim \text{Binomial}(PC_c, T) \quad (3)$$

where  $c = 1, 2, \dots, C$ . The proportion of correct responses observed in an experiment  $\frac{k}{T}$  is merely a particular draw from a distribution of possible proportion correct values. An observed proportion correct could be consistent with a range of different underlying probabilities of responding correctly ( $PC$ ) and this must be inferred from the data, which will be done in Sections 3.3 and 3.4.

We will assume a uniform prior distribution  $P(\sigma^2)$  over a suitably large (albeit arbitrary) range from 0-1000.

$$\sigma^2 \sim \text{Uniform}(0, 1000) \quad (4)$$

This probabilistic generative model is represented in Figure 6. This form of graphical model representation is appealing both because they convey a lot of specific information about the model in a compact manner, but are also accessible and interpretable (Jordan, 2004; Lee and Wagenmakers, 2014). The large box in Figure 6 is called a plate, and represents a for-loop, showing that the nodes within the box are duplicated a number of times, in this case over  $C$  signal intensity conditions. Figure 7 provides a description of different types of nodes used in probabilistic generative models.

### 3.2. Using the model to generate behaviour

We can use the forward generative model to simulate behavioural observations. In our context, the observed data are the number of correct responses  $\mathbf{k}$  and the state of the world is our experimenter-defined parameters  $\{\Delta\mu, \sigma^2, T\}$ . Essentially this forward model allows us to calculate what pattern of data one would expect to see if these are the true parameters. If the forward model was deterministic then a given set of parameters would always give rise to the same number of correct trials. However, the generative model is stochastic and so the model will predict a distribution of number of correct responses for a given set of parameter values for each signal intensity level examined. We will use the ability of the generative model in this way in a later section, but here we take one sample from the distribution

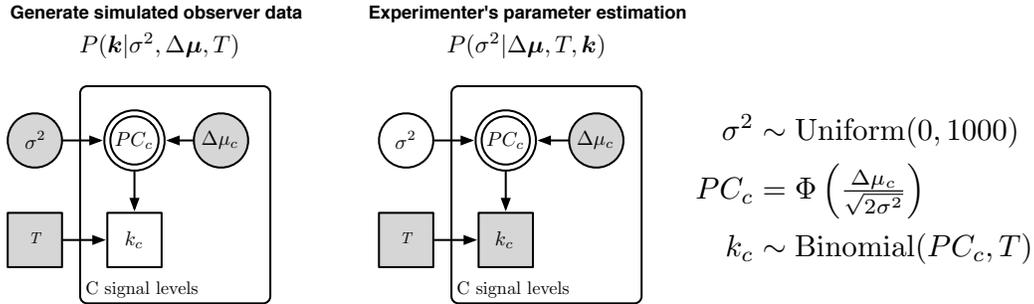


Figure 6: A representation of Model 1. A simple graphical model shown twice for generating simulated data (left) and for parameter estimation (middle). Ideal observers are assumed to have precise knowledge of the true signal and noise levels (thus  $\Delta\mu$ ), which were defined by the experimenter and so are observed variables in both the generative and parameter estimation steps.

|   |   |   |   |
|---|---|---|---|
| observed variables<br><i>(shaded nodes)</i>             |   |   | the node's value is probabilistically related to its parent nodes<br><i>(single border)</i> |
| unobserved, latent variables<br><i>(unshaded nodes)</i> |   |   |   |
|   |   |   | a deterministic variable, the function of parent variables<br><i>(double border)</i>        |
|   | Continuous variables<br><i>(circular nodes)</i> | Discrete variables<br><i>(square nodes)</i> |   |

Figure 7: Types of nodes used in graphical model diagrams. Nodes can represent observed vs. latent variables, continuous vs. discrete valued variables, or those which are probabilistically vs. deterministically related to their parent (input) nodes.

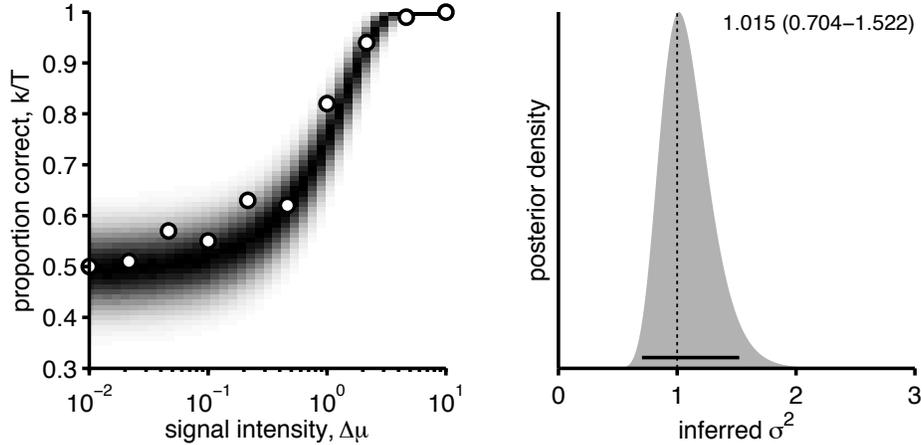


Figure 8: Inferences made with Model 1. The dataset of performances (points) were used to estimate a posterior distribution over the noise variance parameter  $\sigma^2 = 1$  (right). The posterior distribution of noise variance was used to generate a distribution of model predicted performance (left).

shown in Equation 3 for each signal intensity (i.e., simulate one realisation of an experiment). Drawing a sample from this distribution (repeating for each signal level) is very easy and can be done with built in functions which generate random numbers from the Binomial distribution.

Throughout this paper, simulated observers will have observation noise variance of  $\sigma^2 = 1$ . Having generated simulated response data from an observer (see Appendix A.2 for details), an empirical psychometric function can be plotted (points in Figure 8a).

### 3.3. Parameter estimation with grid approximation

In a real experiment, a human’s encoding precision will not be known, and we as experimenters will have to solve the inverse problem to infer what this is. A common modelling approach is to find a single best fitting parameter value (also called a point estimate). However, in a Bayesian approach we estimate an entire distribution specifying how much we believe in a range of different possible parameter values, given the observed data. We can do this with the probabilistic generative model, we can calculate the joint probability of the model for a given set of data observations and parameter values. In our example this joint distribution is  $P(\sigma^2, \mathbf{k}, T, \Delta\boldsymbol{\mu})$ . We can compute this straightforwardly because we can break down (factorise) the

joint distribution into a series of simple conditional probability distributions (see [Appendix A.1](#)).

$$\begin{aligned}
 P(\sigma^2, \mathbf{k}, T, \Delta\boldsymbol{\mu}) &= P(\sigma^2) \times P(\mathbf{k}|\mathbf{PC}, T) \\
 &= \text{Uniform}(\sigma^2; 0, 1000) \\
 &\quad \times \prod_{c=1}^C \text{Binomial}(k_c; PC_c, T)
 \end{aligned} \tag{5}$$

The first expression is the prior probability for any given  $\sigma^2$  under consideration (Equation 4), and the second expression is the data likelihood for given values of  $\mathbf{k}, T$  and  $\sigma^2$  (Equation 3). The deterministic node  $PC$  in this example, does not have a probability associated with it, but is simply calculated (see Equation 2). This parameter estimation step can be implemented easily, see Supplementary Code.

Based upon the simulated dataset used throughout the paper (see [Appendix A.2](#)), Figure 8b shows the posterior distribution over  $\sigma^2$  calculated with grid approximation (as in Figure 3, left). The mode of this posterior distribution, which is the maximum a posteriori (MAP) estimate, is not exactly equal to the true value of  $\sigma^2 = 1$ , not because of any error but because the estimation is based upon only 100 trials worth of data for each signal level. The 95% credibility interval (see horizontal line in Figure 8b) overlaps with the true value, meaning that we believe that there is a 95% probability that the true value of  $\sigma^2$  is contained in this region. Running the same procedure for a different set of experimental data would result in a different posterior distribution of belief over the encoding precision. The choice of signal intensity levels will also affect how precisely we can estimate  $\sigma^2$ , for example, testing signal levels at floor or ceiling performance are less informative at localising the horizontal shift of the psychometric function, so will tell us less about  $\sigma^2$ .

Because the prior over  $\sigma^2$  is uniform, this parameter recovery step could be viewed as likelihood estimation. The posterior will be proportional to the likelihood and the mode of the posterior will equal the maximum likelihood estimate. Readers are referred to [Myung \(2003\)](#) for a tutorial on maximum likelihood estimation in general and to [Kingdom and Prins \(2009\)](#) and [Kuss et al. \(2005\)](#) for parameter estimation with parametrically defined psychometric functions.

### 3.4. Parameter estimation with MCMC sampling

Parameter estimation can also be achieved through MCMC sampling methods. Given that we can calculate the model posterior for a set of observed data and particular parameter values (Equation 5), the Supplementary code shows how the Metropolis Hastings algorithm can be used to generate MCMC samples from the posterior distribution. Conducting the parameter estimation with MCMC sampling results in very similar predictions to that achieved for grid approximation in Figure 8 (results shown in Supplementary Figure 1).

To test whether the parameter estimate was sensitive to the particular span of the uniform prior distribution over  $\sigma^2$ , the parameter estimation procedure was repeated multiple times with the uniform prior spanning the range 0 to 100, 1000, 10000, 100000. The resulting parameter estimate (mode and 95% CI of the posterior) was unaffected (not shown), so for all practical purposes the range of the prior distribution did not affect the inferences made.

In order to gain confidence that the samples returned from the MCMC algorithm accurately reflect the true posterior distribution, the general approach is to repeat this process multiple times, running multiple MCMC chains with different initial parameter values. Visualising MCMC chains (Supplementary Figure 4) showed that convergence was achieved quickly. Quantitative checks such as the  $\hat{R}$  statistic (e.g., Gelman and Rubin, 1992) can help confirm convergence to the posterior distribution. This was found to be the case for this model, with the statistic  $\hat{R} = 1.0$  (see Supplementary Figure 3).

Flegal et al. (2008) highlight that because inference with MCMC is approximate, there is a need to evaluate the accuracy of MCMC estimates. The standard error of 30 repeated MAP estimates of  $\sigma^2$  were calculated as a function of total number of MCMC samples (see Figure 9). We can see that, for this particular model and dataset, estimates are roughly centred on the value estimated with Model 1 (dashed line) and that we start to have confidence in the estimated MAP estimate of  $\sigma^2$  with a total of  $10^6$  or more total MCMC samples.

### 3.5. Model predictions

Having obtained a distribution of belief over parameter values consistent with the data, we also need to establish to what degree the model provides a good or a bad fit to the data. This can be done by using the generative process

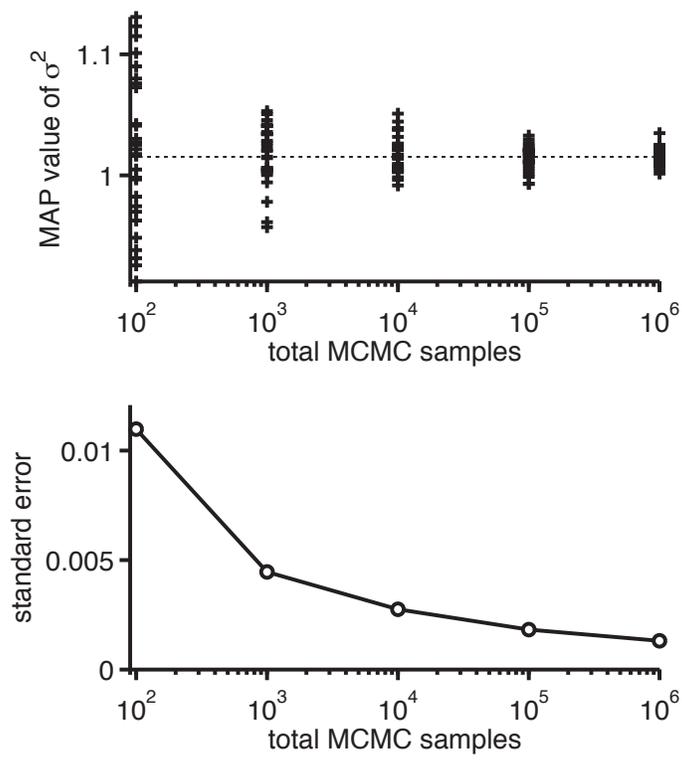


Figure 9: Uncertainty in posterior mode of MCMC-derived parameter estimation. Data points (top) show the posterior mode, conducted multiple times on the common dataset, as a function of total number of MCMC samples calculated (excluding burn-in period). The dashed line shows the estimated posterior mode, evaluated with grid approximation. In this case, a large number of MCMC samples need to be generated to obtain reliable inferences.

again as we did in Section 3.2. However rather than specifying an exact value for the observer’s encoding precision, we can use the posterior distribution over values consistent with the data that we obtained in the previous section. We can also use the model to make predictions of number of correct trials for additional signal intensity levels which were not used in the experiment,  $\widetilde{\Delta\mu}$ . This is done in a similar way to the steps in Section 3.2, just with the additional interpolated signal intensity levels,  $P(\widetilde{\mathbf{k}}|\widetilde{\Delta\mu}, \sigma^2, T, \mathbf{k})$ . Code to predict the outcome of one simulated experiment is also easy to numerically evaluate. By repeating this many times, the posterior predictive distribution will be approximated by these samples drawn, and a distribution of predicted number of correct trials will result (Figure 8a, shaded region).

#### 4. Model 2: A trial-to-trial SDT model

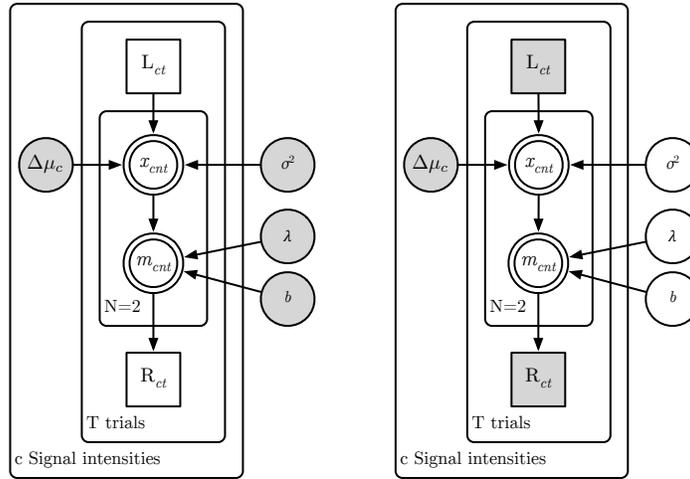
A trial-to-trial version of the SDT model can be constructed as a probabilistic generative model (see Figure 10). Trial-to-trial models of the 2AFC task have been investigated by DeCarlo (2012) and for the yes/no detection procedure (Ma et al., 2011; Mazzyar et al., 2012, 2013). One reason to create trial-to-trial models is that they more clearly represent the events occurring in an experiment, and are thus more interpretable. Another reason is that more specific predictions can be generated: rather than modelling averaged performance, we can calculate how a model observer would have responded given the very same sequence of signal locations that an actual experimental subject was exposed to.

The model in Figure 10 consists of a number of simple steps. On a given trial  $t$  for a given signal intensity condition  $c$ , a signal is present in location  $L$  (1 or 2) with equal probability. This gives rise to noise-corrupted sensory observations  $\mathbf{x}_{ct}$  which depend upon the observation noise variance  $\sigma^2$  and the signal intensity  $\Delta\mu_c$ . The observer is assumed to respond  $R$  that the signal is in location 1 or 2 with certain probabilities. This is governed by a probability vector  $\mathbf{m}$  which maps noisy sensory observations  $\mathbf{x}$  to response  $R$ . This probability vector incorporates both response biases and lapse rates.

Response bias is modelled as in DeCarlo (2012), meaning the observer responds to location 1, if  $x_1 - x_2 \geq b$ . If  $b > 0$ , then the observer has a response bias favouring location 2. Response biases are often ignored, especially with AFC tasks with more than 2 locations, not on any theoretical grounds, but because of difficulties in expressing and evaluating the model. This simplification is not without consequence, published parameter estimates can be

**Simulate experiment and  
participant responses**  
 $P(\mathbf{L}, \mathbf{R} | \sigma^2, \lambda, b, \Delta\mu)$

**Experimenter's parameter  
estimates**  
 $P(\sigma^2, \lambda, b | \mathbf{L}, \mathbf{R}, \Delta\mu)$



$$\begin{aligned}
 b &\sim \text{Normal}(0, 1/2000) \\
 \lambda &\sim \text{Beta}(1, 1) \\
 \sigma^2 &\sim \text{Uniform}(0, 1000) \\
 L_{ct} &\sim \text{Categorical}(0.5, 0.5)
 \end{aligned}$$

$$x_{cnt} \sim \begin{cases} \text{Normal}(\Delta\mu_c, 1/\sigma^2) & \text{if } n = L_{ct} \\ \text{Normal}(0, 1/\sigma^2) & \text{if } n \neq L_{ct} \end{cases}$$

$$\mathbf{m}_{ct} = \begin{cases} [1 - \frac{\lambda}{2}, \frac{\lambda}{2}] & \text{if } x_{c1t} - x_{c2t} > b \\ [\frac{\lambda}{2}, 1 - \frac{\lambda}{2}] & \text{if } x_{c1t} - x_{c2t} \leq b \end{cases}$$

$$R_t \sim \text{Categorical}(\mathbf{m}_{ct})$$

Figure 10: Model 2, a trial to trial Max-rule observer. The model describes the generative process giving rise to sensory observations and how these observations map onto observer responses. The model is first used to generate simulated signal locations and observer responses (left). This simulated data are then used to conduct parameter estimation and to generate model predictions (right).

substantially altered when response biases are accounted for, thus potentially affecting research conclusions (García-Pérez and Alcalá-Quintana, 2011).

The model also incorporates lapse rates. Response errors are modelled by responses being randomly selected on a small proportion  $\lambda$  of trials. This is also achieved through the deterministic node  $\mathbf{m}$  implementing the max decision rule. The value of the node  $\mathbf{m}_{st}$  will be a pair of numbers representing location 1 and 2,  $[1 - \frac{\lambda}{2}, \frac{\lambda}{2}]$  or  $[\frac{\lambda}{2}, 1 - \frac{\lambda}{2}]$  depending upon which location contains the highest valued sensory observation. For example, if according to the Max rule the signal is decided to be in location 1, then for a lapse rate  $\lambda = 0.05$  the probability of responding to location 1 and 2 is  $[0.975, 0.025]$ .

#### 4.1. Formulating the model

Beyond very simple models it can become counterproductive to manually generate code to evaluate the model’s joint posterior and to implement an MCMC sampling algorithm. Instead, we use one of a number of advanced MCMC sampling packages available. The JAGS software package (Just Another Gibbs Sampler; Plummer, 2003) was used to formulate the graphical model and to conduct the inference. JAGS is well-developed, free, and available for Mac, PC, and Linux platforms. The probabilistic generative model (Figure 10) is translated into a JAGS model specification (see Supplementary Material and Supplementary Code). For more examples, readers are referred to the JAGS user manual (Plummer, 2003), and the books by Lee and Wagenmakers (2014) and Lunn et al. (2012).

#### 4.2. Model 2 results

For this model, the experimenter is aware of the true signal locations  $\mathbf{L}$  used in the experiment, the observer’s responses  $\mathbf{R}$ , and the signal intensities  $\Delta\boldsymbol{\mu}$ . The experimenter must infer the latent parameters  $\sigma^2$ ,  $\lambda$ , and  $b$ . Figure 11 (right) shows posterior distributions over the noise variance, lapse rate, and bias parameters. The histograms show the marginal distributions for each variable, and we can see that for this dataset the 95% credibility interval of the marginal posterior distributions clearly includes the true values, indicating successful recovery of all parameter values. One of the advantages of this Bayesian analysis however is we have a full distribution over parameter values, and so we can examine whether there are any trends between parameters (see the density plots).

Figure 11 (left) shows the observer’s performance data (points), the model’s posterior predictive distribution (similar to Figure 8 left), but we also have a

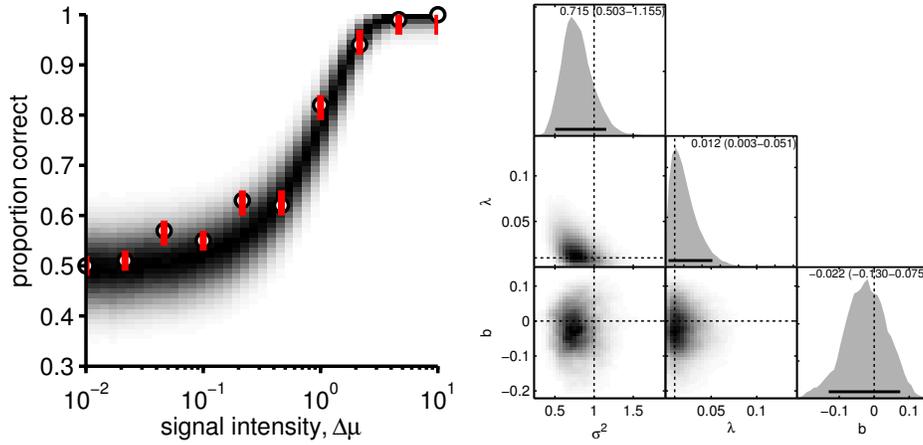
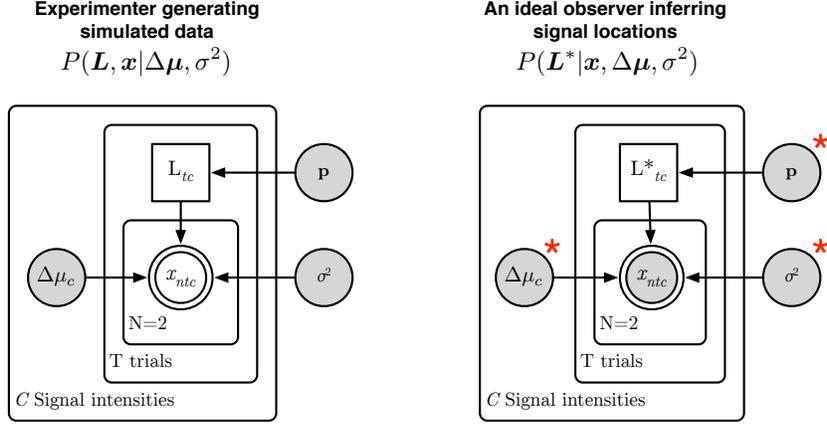


Figure 11: Inferences made with Model 2. The model predictions are shown by the posterior predictive distribution (grayscale intensity, left). Red 95% confidence intervals show a second posterior predictive distribution, see text. Parameter estimation for the noise variance, lapse rate, and decision bias parameters are shown (right).

second posterior predictive distribution (red lines). This latter set of model predictions answer the question: what responses would the model predict given knowledge of location of the signal ( $L$ ) and the response ( $R$ ) on each trial? We can see that these predictions are more precise (narrower 95% credibility intervals). Models that make more specific predictions are appealing as it gives a greater chance that experimental data will conflict with the predictions, so the model can be tested more stringently.

### 5. Model 3: A Bayesian optimal observer model

Model 3 is a Bayesian optimal observer of the 2AFC task (see Figure 12). While SDT and Bayesian approaches are similar, a number of differences can be identified (Ma, 2012). Firstly, while the SDT model can have a spatial bias term which will achieve the same effect, it is not an explicit prior over the signal location. It mirrors the top half of Model 2 in that it is just a generative model of what gives rise to the observer’s observations on each trial. The Bayesian optimal observer is then assumed to use this generative model of the task to conduct inferences over the state of the world (signal location) given the sensory observations. Optimal observers are also assumed to have perfect knowledge of other key variables (Figure 12, red asterisks),



$$\begin{aligned}
 L_{tc} &\sim \text{Categorical}(\mathbf{p}) \\
 \sigma^2 &\sim \text{Uniform}(0, 1000) \\
 x_{cnt} &\sim \begin{cases} \text{Normal}(\Delta\mu_c, 1/\sigma^2) & \text{if } n = L_{ct} \\ \text{Normal}(0, 1/\sigma^2) & \text{if } n \neq L_{ct} \end{cases}
 \end{aligned}$$

Figure 12: Model 3, a Bayesian ideal observer. A dataset of locations and sensory observations can be simulated (left). The model can then be used to infer the observers’ inferences about the true state of the world  $\mathbf{L}^*$  (right). Red asterisks represent quantities known precisely by a theoretical ideal observer. A real human observer would have uncertainty over those quantities, imparted by experimental task instruction or previous experience.

here, the signal intensities  $\Delta\boldsymbol{\mu}$ , their own observation noise variance  $\sigma^2$ , and the probability of a signal appearing in each location, eg.  $\mathbf{p} = [0.5, 0.5]$ .

Predicted behaviour of an observer was calculated in two steps. The first generates a dataset of true signal location  $\mathbf{L}$  and noisy observations  $\mathbf{x}$  (Figure 12, left). In the second stage, the model now switches meaning to the observer’s internal mental model of what generated their observations. This can be used to make inferences about the true stimulus location  $\mathbf{L}^*$  given the sensory data  $\mathbf{x}$  (Figure 12, right). On one trial,  $L^*$  is the observer’s posterior belief of the signal location, and observer is assumed to respond to the most probable location,

$$R = \underset{n}{\operatorname{argmax}} L_n^*. \quad (6)$$

Figure 13 shows predicted performance of an optimal observer and a suboptimal observer. The optimal observer has the correct, unbiased prior belief that the signal will appear in location 1 and 2,  $\mathbf{p} = [0.5, 0.5]$ . However,

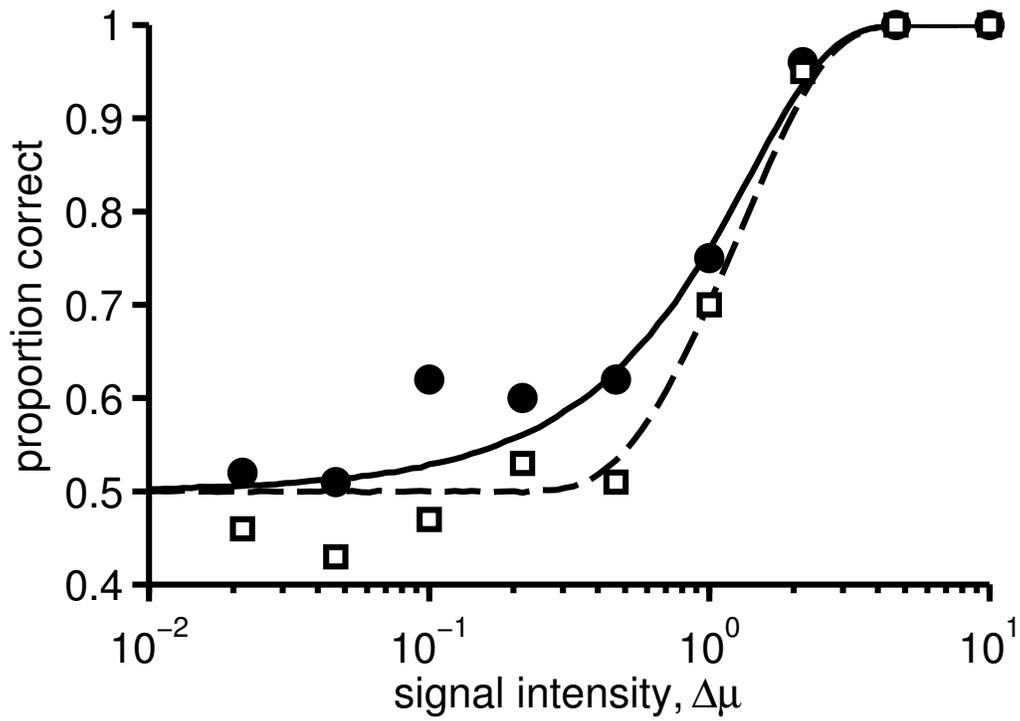


Figure 13: Predicted psychometric functions of Model 3. Points represent performance calculated with MCMC methods. Filled black circles and solid line correspond to an unbiased observer, squares and dashed lines correspond to a biased observer who has the incorrect belief that targets have a 75:25% spatial prior.

if the observer’s prior over signal location is incorrect ( $\mathbf{p} = [0.75, 0.25]$ ) then it will have diminished performance until the signal intensity provides so much information about the signal’s location that the effect of the incorrect prior is overcome.

Just as with Models 1 and 2, Model 3 can be used to make inferences about an observer’s  $\sigma^2$  (see Supplementary Figure 7). However, because of the form of the model, it was not as straightforward to calculate this posterior distribution over  $\sigma^2$ . For example, the model does not allow simultaneous observation of the true signal locations and observer responses  $R$ . There is in fact no node in the model for  $R$ , this has to be calculated outside of the JAGS model because evaluating Equation 6 requires access to all of the MCMC samples describing the posterior over  $L^*$ . The solution used to calculate the posterior distribution over  $\sigma^2$  (see Supplementary Figure 7) was as follows. Grid approximation over many values of  $\sigma^2$  was conducted. For each value, the posterior was evaluated with a Binomial likelihood, a discrete uniform prior over  $k$ , and a continuous uniform prior over  $\sigma^2$ :  $P(k_c | \Delta\mu_c, \sigma^2, T) = \text{Uniform}(k_c; 0, 100) \times \text{Uniform}(\sigma^2; 0, 1000) \times \prod_{c=1}^C \text{Binomial}(k_c; PC_c, T)$ . Where  $PC_c = f(\sigma^2, \Delta, \mu_c)$  was calculated by assessing the performance of the observer over  $10^6$  simulated trials (steps 1 and 2 demonstrated in Figure 12). This process was found to be too computationally demanding to conduct with JAGS, and so the evaluation of  $PC_c$  used manually written code instead of JAGS (see Supplementary Code).

## 6. Discussion

This tutorial has provided an introduction to a range of important concepts underlying the evaluation of Bayesian explanations of perception. Some of these, such as simulating data, parameter estimation, and calculating model predictions are important regardless of whether or not one uses Bayesian methods of evaluation, or if a model claims that perceptions involve Bayesian inference. More in depth treatments of these concepts with non-Bayesian and Bayesian approaches are provided by [Lewandowski and Farrell \(2011\)](#) and [Lee and Wagenmakers \(2014\)](#), respectively. The review also explored issues relating to SDT and Bayesian explanations of perception with the important 2AFC task as a case study. The first model introduced the SDT account of the task and demonstrated important aspects of Bayesian model evaluation. The second model which was equivalent to the first (except for the inclusion of bias and lapse rate terms), showed that describing the events on a trial-to-

trial basis, leading from noisy stimuli through information processing to behavioural response was more intuitive. This second form of model is perhaps most useful in providing a template for modelling other perceptual tasks. Finally, a Bayesian ideal observer more clearly demonstrated some of the more important theoretical claims underlying Bayesian explanations of perceptual phenomena. Namely that observers conduct inference about the state of the world, based upon priors over world states, noisy and sometimes ambiguous sensory information, and an internal probabilistic generative model of the world.

The modeller has a number of options when it comes to the practical evaluation of the models. Both grid approximation and MCMC sampling require the joint probability distribution of the model to be evaluated (demonstrated with Model 1). For more complex models it is perhaps more convenient to use purpose built MCMC sampling software to do this task. However the current algorithms and software implementations do have limitations, it can become computationally demanding to evaluate large models with many nodes. For example, using JAGS it was not possible to calculate the optimal observer’s performance with more than a few thousand simulated trials (Figure 13, points), but it was perfectly possible to construct hand-written code to evaluate performance over millions of simulated trials (Figure 13, lines). No doubt, these issues will soon disappear given the pace of development in MCMC sampling algorithms. One of the key limitations of grid approximation, however, is that it becomes too computationally demanding as the number of model parameters increase. More than 3-4 parameters becomes infeasible to conduct on current desktop computers. However, some extra care is required with MCMC approaches. Figure 9 clearly demonstrated the approximate nature of the inference being conducted by the MCMC approach. Given that research conclusions drawn from data rest on MCMC-derived parameter estimates, it would seem prudent to have confidence in the accuracy of the MCMC estimate (Flegal et al., 2008, and Figure 9). Secondly, one needs to confirm convergence of the MCMC chains using both visual and quantitative checks such as the  $\hat{R}$  statistic (Gelman and Rubin, 1992). Third, care is also needed in situations where MCMC chains contain autocorrelation, such as in the parameters  $b$  and  $\lambda$  in Model 2. While some advocate discarding some chain values to decrease this autocorrelation, others recommend against this (Link and Eaton, 2011).

Bayesian explanations of perceptual phenomena have been rising in popularity, in part this is because they allow direct and quantitative testing of the

compelling constructivist approach (Gregory, 1980). To determine the extent to which this Bayesian Brain hypothesis can provide insights into perception will require that a broad range of researchers can not only understand the theoretical claims made by Bayesian models of perception, but also have the ability to construct and test the limits of such explanations. Visual depiction of probabilistic generative models aid in conveying these theoretical claims, avoiding reader attrition sometimes caused by dense mathematical descriptions alone (Fawcett and Higginson, 2012). Rather than this necessitating a more verbose, less efficient, less accurate model description (Fernandes, 2012), generative model diagrams offer accessibility alongside the compact accuracy of mathematics. Use of MCMC sampling algorithms also ‘black boxes’ some of the practical evaluation processes, which again allows focus to be placed on the models and their theoretical claims. It also lowers the barrier to entry such that more people can engage in Bayesian modelling. If this is combined with emerging good practice for making research code publicly available (Peng, 2011; Morin et al., 2012; Ince et al., 2012; Kubilius, 2014) then the barriers to entry are lowered further. The more researchers probing the limits of Bayesian explanations of perceptual phenomena the better.

## Appendix A. Appendix

### *Appendix A.1. Bayesian Networks*

The probabilistic generative models considered in this paper are Bayesian networks, where variables are defined (deterministically or probabilistically) in terms of their inputs (or parents). Put formally, Bayesian Networks (also known as belief networks) are a subclass of probabilistic generative models where the joint distribution between all of the model variables can be expressed as a function of their parents (Barber, 2012),

$$P(x_1, \dots, x_N) = \prod_{i=1}^N P(x_i | \text{pa}(x_i)). \quad (\text{A.1})$$

Where  $\text{pa}(x_i)$  represents the parents of the variable  $x_i$ . If a variable  $x_i$  has no parents, then  $P(x_i | \text{pa}(x_i))$  is represented by a prior distribution over values of node  $x_i$ . Bayesian networks form directed acyclic graphs in that the relationship between variables proceeds in one direction, and no circular loops can be drawn through the network.

### *Appendix A.2. The simulated dataset*

Model 2 was used to create a common dataset of responses in a simulated 2AFC experiment. There were 10 signal intensities, logarithmically spaced between 0.01 and 10. The true parameter values were  $\sigma^2 = 1$ ,  $T = 100$ ,  $b = 0$ ,  $\lambda = 0.01$ , and  $\mathbf{p} = [0.5, 0.5]$ . For each signal intensity, 100 simulated trials were run. The raw data was transformed from correct or incorrect on each trial, to proportion correct  $\frac{k}{T}$  for use with Model 1 which does not model individual trials. The proportion correct responses are shown as points in Figures 8 and 11. This dataset is comparable to one we may obtain from a real experiment with the exception that we do not know the true values of the latent parameters  $\sigma^2$ ,  $b$ ,  $\lambda$ , or  $\mathbf{p}$ .

### **Acknowledgements**

I am grateful to Daniel Baker, Shane Lindsey, Keith May, Tom Wallis, Britt Anderson, Alastair Clarke, Anuenue Baker-Kukona, Ben Tatler, Kirsty Miller, and Karl Smith-Byrne for providing instructive comments during the preparation of this paper.

### **References**

- Acerbi, L., Vijayakumar, S., and Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS Computational Biology*, 10(6):e1003661.
- Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.
- Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., and Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, 74(1):30–39.
- Bowers, J. S. and Davis, C. J. (2012a). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3):389–414.
- Bowers, J. S. and Davis, C. J. (2012b). Is that what Bayesians believe? Reply to Griffiths, Chater, Norris, and Pouget (2012). *Psychological Bulletin*, 138(3):423–426.

- Colombo, M. and Seriès, P. (2012). Bayes in the Brain—On Bayesian Modelling in Neuroscience. *The British Journal for the Philosophy of Science*.
- Cum, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49(4):327–335.
- DeCarlo, L. T. (2012). On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and bayesian approaches to estimation. *Journal of Mathematical Psychology*, (56):196–207.
- Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N., editors (2007). *Bayesian Brain*. Probabilistic Approaches to Neural Coding. MIT Press.
- Farrell, S. and Lewandowski, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science*, 19(5):329–335.
- Fawcett, T. W. and Higginson, A. D. (2012). Heavy use of equations impedes communication among biologists. *Proceedings of the National Academy of Sciences*, pages 1–5.
- Fennell, J. and Baddeley, R. J. (2012). Uncertainty plus prior equals rational bias: an intuitive Bayesian probability weighting function. *Psychological Review*, 119(4):878–887.
- Fernandes, A. D. (2012). No evidence that equations cause impeded communication among biologists. *Proceedings of the National Academy of Sciences*, 109(45):E3057–author reply E3058–9.
- Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov Chain Monte Carlo: Can We Trust the Third Significant Figure? *Statistical Science*, 23(2):250–260.
- Frith, C. (2013). *Making up the Mind*. How the Brain Creates Our Mental World. John Wiley & Sons.
- García-Pérez, M. A. and Alcalá-Quintana, R. (2011). Interval bias in 2AFC detection tasks: sorting out the artifacts. *Attention, Perception & Psychophysics*, 73(7):2332–2352.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*.

- Geisler, W. S. (2003). Ideal observer analysis. In Chalupa, L. and Werner, J., editors, *The Visual Neurosciences*, pages 825–837. The visual neurosciences, Boston.
- Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. 51(7):771–781.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition*. CRC Press.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472.
- Gibson, J. J. (2002). A theory of direct visual perception. In Nöe, A. and Thompson, E., editors, *Vision and Mind: Selected Readings in the Philosophy of Perception*, pages 77–89. MIT Press.
- Glymour, C. N. (2001). *The Mind’s Arrows*. Bayes Nets and Graphical Causal Models in Psychology. MIT Press.
- Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics*. Peninsula Publishing, Los Altos.
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 290(1038):181–197.
- Griffiths, T. L., Chater, N., Norris, D., and Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychological Bulletin*, 138(3):415–422.
- Helmholtz (1856). *Treatise on Physiological Optics*. The Optical Society of America.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.
- Ince, D. C., Hatton, L., and Graham-Cumming, J. (2012). The case for open computer programs. *Nature*, 482(7386):485–488.
- Jordan, M. I. (2004). Graphical Models. *Statistical Science*, 19(1):140–155.

- Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55:271–304.
- Kingdom, F. A. A. and Prins, N. (2009). *Psychophysics. A Practical Introduction*. Academic Press.
- Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719.
- Knill, D. C. and Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge University Press.
- Kording, K. P. (2007). Decision Theory: What “Should” the Nervous System Do? *Science*, 318(5850):606–610.
- Kruschke, J. K. (2011). Bayesian models of mind, psychometric models, and data analytic models. <http://doingbayesiandataanalysis.blogspot.co.uk/2011/10/bayesian-models-of-mind-psychometric.html>. Last accessed on Nov 21, 2014.
- Kruschke, J. K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, 2nd edition.
- Kubilius, J. (2014). Sharing code. *i-Perception*, 5:75–78.
- Kuss, M., Jakel, F., and Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, 5(5):8–8.
- Lee, M. D. (2008). BayesSDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods*, 40(2):450–456.
- Lee, M. D. and Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- Lewandowski, S. and Farrell, S. (2011). *Computational Modeling in Cognition: principals and practice*. SAGE Publications.
- Link, W. A. and Eaton, M. J. (2011). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3(1):112–115.

- Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. J. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. CRC Press.
- Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences*, 16(10):511–518.
- Ma, W. J., Navalpakkam, V., Beck, J. M., Berg, R. v. d., and Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, 14(6):783–790.
- Ma, W. J., van den Berg, R., Vogel, M., and Josic, K. (2012). Optimal inference of sameness. *Proceedings of the National Academy of Sciences*, 109(8):3178–3183.
- MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, UK.
- Maloney, L. (2002). Statistical decision theory and biological vision. *Perception and the physical world: Psychological and . . .*
- Maloney, L. and Zhang, H. (2010). Decision-theoretic models of visual perception and action. 50(23):2362–2374.
- Mazyar, H., van den Berg, R., and Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision*, 12(6):1–16.
- Mazyar, H., van den Berg, R., Seilheimer, R. L., and Ma, W. J. (2013). Independence is elusive: Set size effects on encoding precision in visual search. *Journal of Vision*, 13(5):1–14.
- Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., and Baker, D. (2012). Shining light into black boxes. *Science*.
- Myung (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):11–11.
- North, D. W. (1968). A tutorial introduction to decision theory. *IEEE Transactions on Systems Science and Cybernetics*, 4(3):200–210.
- Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060):1226–1227.

- Pizlo, Z. (2001). Perception viewed as an inverse problem. 41(24):3145–3161.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, pages 20–22.
- Sims, C. R., Jacobs, R. A., and Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, 119(4):807–830.
- Tarantola, A. (2004). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics, Philadelphia.
- Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, 2(8):492–494.
- Vincent, B. T. (2011). Covert visual search: prior beliefs are optimally combined with sensory evidence. *Journal of Vision*, 11(13):25.
- Vincent, B. T. (2015). Bayesian accounts of covert selective attention: a tutorial review . *Attention, Perception & Psychophysics*.
- Wickens, T. (2002). *Elementary signal detection theory*. Oxford University Press, Oxford.