OPEN ACCESS

University of Dundee

Bayesian accounts of covert selective attention

Vincent, Benjamin

Published in:
Attention, Perception and Psychophysics

DOI:
10.3758/s13414-014-0830-0

Publication date:
2015

Document Version
Early version, also known as pre-print

Link to publication in Discovery Research Portal

Citation for published version (APA):
Vincent, B. T. (2015). Bayesian accounts of covert selective attention: a tutorial review. Attention, Perception and Psychophysics, 77(4), 1013-1032. DOI: 10.3758/s13414-014-0830-0

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 17. Feb. 2017

Author's accepted copy

Attention, Perception & Psychophysics manuscript No.
(will be inserted by the editor)

Bayesian accounts of covert selective attention: a tutorial review

Benjamin T. Vincent

Received: 02-Jun-2014 / Accepted: date

Abstract Decision making and optimal observer models offer an important theoretical approach to the study of covert selective attention. While their probabilistic formulation allows quantitative comparison to human performance, the models can be complex and their insights are not always immediately apparent. Part 1 establishes the theoretical appeal of the Bayesian approach, and introduces the way in which probabilistic approaches can be applied to covert search paradigms. Part 2 presents novel formulations of Bayesian models of 4 important covert attention paradigms, illustrating optimal observer predictions over a range of experimental manipulations. Graphical model notation is used to present models in an accessible way and Supplementary Code is provided to help bridge the gap between model theory and practical implementation. Part 3 reviews a large body of empirical and modelling evidence showing that many experimental phenomena in the domain of covert selective attention are a set of by-products. These effects emerge as the result of observers conducting Bayesian inference with noisy sensory observations, prior expectations, and knowledge of the generative structure of the stimulus environment.

Keywords Covert attention · Signal detection theory · Bayesian · optimal observer · Probabilistic graphical model

1 Introduction

Helmholtz (1856) is often credited as having provided the first experimental evidence of selective visual information processing. A prior decision by an observer to concentrate upon a specific peripheral location resulted in enhanced identification of briefly illuminated letters. Resolving how and why this, and related experimental effects, occur has not been a trivial matter. A vast array of experimental paradigms have since emerged to investigate different aspects of this visual information processing. On one end of the spectrum we have natural visual search taking place with multiple eye movements and natural scenes. These paradigms fully embrace the complexity of ongoing information processing of incoming sensory signals as the eyes move over time. However, if we wish to study the precise information processing mechanisms underlying an observer's behaviour, we must exclude uncontrolled variation in the nature of the information being processed by these mechanisms. The 'performance paradigm' achieves this by: short display durations, controlling for retinal stimulus location, and focussing upon performance measures with non-speeded response instructions. While this paradigm may miss many of the important challenges faced by observers in naturalistic stimulus and task environments, it is a necessary trade-off in order to study the information processing mechanisms.

A short stimulus display duration, typically in the order of 100ms, is central to this approach. This near eliminates the contribution from the serial process of eye movements Zelinsky and Sheinberg (1997). It also eliminates a speed accuracy trade-off in information accumulation time and performance that would occur if stimuli were presented until a response is made.

Another potential speed accuracy trade-off, in processing time, can occur in the more commonly used 'reaction time paradigm' (Wood and Jennings, 1976; Wickelgren, 1977). If observers respond as quickly as possible whilst keeping error rates low, it is possible that changes in reaction times across experimental conditions could reflect changes of response strategy, rather than of underlying information pro-

Supplementary material is available as well as downloadable Matlab code from `https://github.com/drbenvincent/BayesCovertAttention`.

B. T. Vincent
School of Psychology,
University of Dundee.
E-mail: b.t.vincent@dundee.ac.uk

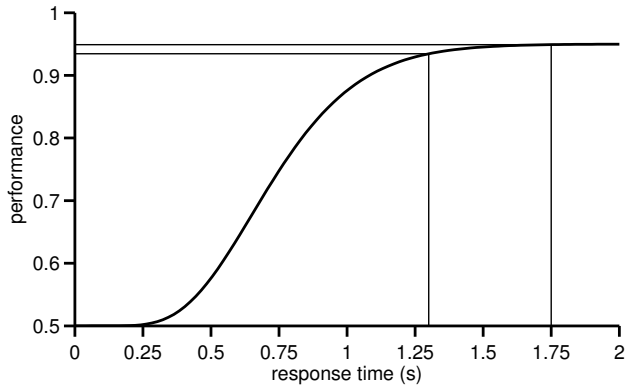

Fig. 1 A schematic speed-accuracy tradeoff function where performance increases with processing time before response. In the domain of high performance, large changes in choice reaction time could be due to a change in speed-accuracy strategy (rather than in the nature of information processing) undetectable in terms of performance changes.

cessing. This strategy change could be undetectable however, because large changes in reaction time can be associated with small changes in performance (see Figure 1). Wood and Jennings (1976) highlight the importance of establishing a complete speed accuracy trade-off function. Studies that do this show that information processing is best accounted for by parallel information processing mechanisms (McElree and Carrasco, 1999; Dosher et al., 2004), with serial processes being attributable to eye movements (Lu et al., 2010). The majority of studies examined here however employ the performance paradigm, where observers are instructed to maximise their performance, with this being the primary, or only, behavioural measure.

Due to the changes in photoreceptor sampling density over the retina, stimuli presented at different retinal eccentricities will be encoded with varying levels of precision, thus imparting differing amounts of information to an observer. If this is unconstrained over the course of a trial, then it is difficult to attribute experimental effects to information processing changes as opposed to these early sensory sampling changes (Kinchla, 1992). Using a circular array of stimuli with central fixation and brief display durations largely negates the major confound of retinal sampling density (Carrasco and Frieder, 1997).

Having established the rationale for the highly simplified experimental paradigm, we still have more work to do before embracing the details behind decision making approaches to covert selective attention. Namely, which of two very different forms of approach shall be taken and why?

1.1 Cause verses effect

We have at least two broad ways in which we may approach the issue of attention (James, 1890). Firstly, we may observe some behavioural phenomena, and then search for an internal mechanistic *cause* which produced those phenomena. Alternatively, we may look outwardly to the environment and ask why these behavioural *effects* occurred. This cause/effect distinction first highlighted by James, is rarely discussed directly, but more recent examinations show that it is crucial to address (and hopefully resolve or reconcile) these different approaches (James, 1890; Johnston and Dark, 1986; Fernandez-Duque and Johnson, 2002; Anderson, 2011; Krauzlis et al., 2014).

The causal approach, which could be mapped onto the algorithm or implementation levels of analysis of Marr (1982), proceeds broadly as follows: a) observe some behavioural effects, b) infer the existence of a mechanism which caused those effects, c) refine the proposed mechanism as more data are observed over time. In the present context, many researchers inferred the existence of a causal mechanism, called attention, to account for experimental phenomena. Over time, models of attention have been proposed and iteratively adjusted in the light of new evidence (e.g. Treisman and Gelade, 1980; Wolfe and Cave, 1989; Wolfe, 2007). While this class of account have proven extremely influential, it is important to remember that they carry this (sometimes implicit) assumption that attention exists as a causal mechanism, and as recently argued by B. Anderson (2011), this assumption is by no means universally accepted nor unproblematic.

Alternatively, we could examine the computational goal of observers (Marr, 1982), or take the related theory-level approach of J. Anderson (1990). This approach assumes that organisms are *adaptively rational* in that they try to optimise behaviour to suit goals within a particular environment, under the influence of constraints. This is conceptually very different from the mechanism-level approach. In this framework, potentially all behaviour is adaptive and our job as scientists is to propose what it is that organisms are optimising. Shaw and Shaw (1977) take this approach, arguing that viewing search behaviour as adapted, in some sense, by the evolutionary selection pressures in a competitive environment. Under this approach, as will become clear, we can reframe attention as being a set of experimental *effects* (Johnston and Dark, 1986; Anderson, 2011) that emerge as a by-product of our adaptively rational behaviour. This is a key conceptual difference to grasp if the theoretical implications of Bayesian accounts of attentional phenomena are to be fully appreciated. If we assume that behaviour is adapted to the environment, we must a) characterise the structure of the environment, b) define the behavioural goals of the observer, and then c) deduce the optimal behaviour.

In terms of (a), Anderson (1990) highlights that the structure of the external environment is easier to empirically measure compared to hypothesised internal cognitive mechanisms. In our case, the statistical structure of the environment in our simple experimental paradigms can be precisely

known and manipulated (see Figure 2). If we change the environment, then behaviour should alter in predictable ways, thus allowing the adaptive explanation of the behavioural observations to be experimentally tested [1] . In terms of (b) because the tasks of localisation or detection are so simple, we can assume that the behavioural goal of a motivated experimental observer is to maximise the proportion of correct decisions. If we accept this, then we are on our way to a theory-level explanation after conducting step (c), deducing some predicted behaviour in a variety of experimental situations. Traditionally this has been done by using signal detection theory (SDT) and deriving closed-form mathematical expressions to compute predicted performance levels.

1.2 Signal detection theory

Signal detection theory (Green and Swets, 1966) is an application of the more general statistical decision theory (Maloney and Zhang, 2010) and has been a powerful approach with which to model simple attentional tasks. It is conceptually simple, consisting of three main steps (Wickens, 2002). Firstly, it assumes that sensory evidence about a stimulus in the world can be represented by a single number, such that a stimulus display of 4 Gabors could be represented by 4 numbers. In practice, the sensory decomposition will consist of many sensory channels (such as size, contrast, spatial frequency, etc) but these are unmonitored due to their task irrelevance. Second, this sensory evidence is corrupted by stochastic noise. Third, the response decision is arrived at through applying a simple decision rule to the magnitude of sensory evidence. For example in yes/no detection, a yes response could be given if the highest-valued sensory measure exceeds a response threshold. Another aspect of the more general statistical decision theory, is the concept of a gain function. This specifies the gain or loss for each response, dependent upon the state of the world. This has been incorporated in some covert attentional studies (eg. Navalpakkam et al., 2009), but because the majority of studies reviewed here use symmetrical gain functions (e.g. the gain of a correct detection is equal to a correct rejection), we do not focus upon the role of rewards.

Application of SDT to covert visual search was pioneered by (Palmer et al., 1993), and has subsequently become a dominant explanation for a wide variety of experimental effects within this short display duration approach of studying attention (reviewed in Section 3, and see Verghese, 2001). While the approach is conceptually simple, calculating predicted behaviours can get somewhat technical, which perhaps subtly shifts the emphasis towards practical implementation and away from the theoretical implications of the models.

In some ways, the SDT and Bayesian models of covert attention are very similar. They are manifestations of statistical decision theory and bayesian decision theory, respectively. The key difference between these two versions of decision theory is that the latter models an observer's prior knowledge about the state of the world (Maloney and Zhang, 2010). For covert search tasks, both SDT and Bayesian models suggest a parallel, noise-limited mechanism, where cueing effects are caused by decision-level mechanisms (changes in response thresholds or priors) rather than cue-induced changes in sensory encoding precision (Palmer et al., 1993, 2000; Verghese, 2001).

However, SDT and Bayesian models of covert attentional effects are not always equivalent. Firstly, the Bayesian approach doesn't *necessarily* assume that stimuli are represented by a single number (such as in population coding, Zemel et al., 1998; Pouget et al., 2000). Ma (2012) points out that it is not just important to take a singular sensory measurement of stimuli, but also to estimate and represent the level of uncertainty associated with those sensory measurements on a trial-to-trial basis. A second difference is that while SDT models can result in a range of possible predictions depending upon different decision rules applied to a sensory axis, Bayesian (optimal observer) models make singular predictions (Eckstein, 2011, p.18) based upon an axis of posterior belief. Third, decision rules of SDT often apply to noisy sensory observations, whereas under the Bayesian approach, sensory information is always transformed into likelihoods, so the decision stage deals with probabilities of sensory measurements being caused by targets or distracters instead of the raw sensory measurement itself. Having said this, in summary, SDT can offer close approximations to Bayesian models (Nolte and Jaarsma, 1967) and it would be reasonable for SDT and Bayesian models to be thought of as similar in their theoretical approach in explaining attentional effects.

1.3 Bayesian observers

1.3.1 The Bayesian approach applied to our covert search tasks

One appeal of viewing observers as conducting Bayesian inference stems from a very basic assumption that the brain does not have direct access to the true state of the world but only to sensory measurements. The task of an observer is to make inferences about the world, based upon these sensory observations (Gregory, 1980; Pizlo, 2001). Probability

[1] This emphasis upon the role of the environment is also a key part of Gibson's ecological approach (Gibson, 1972). However, probabilistic approaches directly oppose Gibson's claim that the environment is sufficiently rich so as to be unambiguous. They are more in line with the constructivist approach that sensory observations of the environment are ambiguous, thus requiring inferences to be made about the state of the world (Helmholtz, 1856; Gregory, 1980).

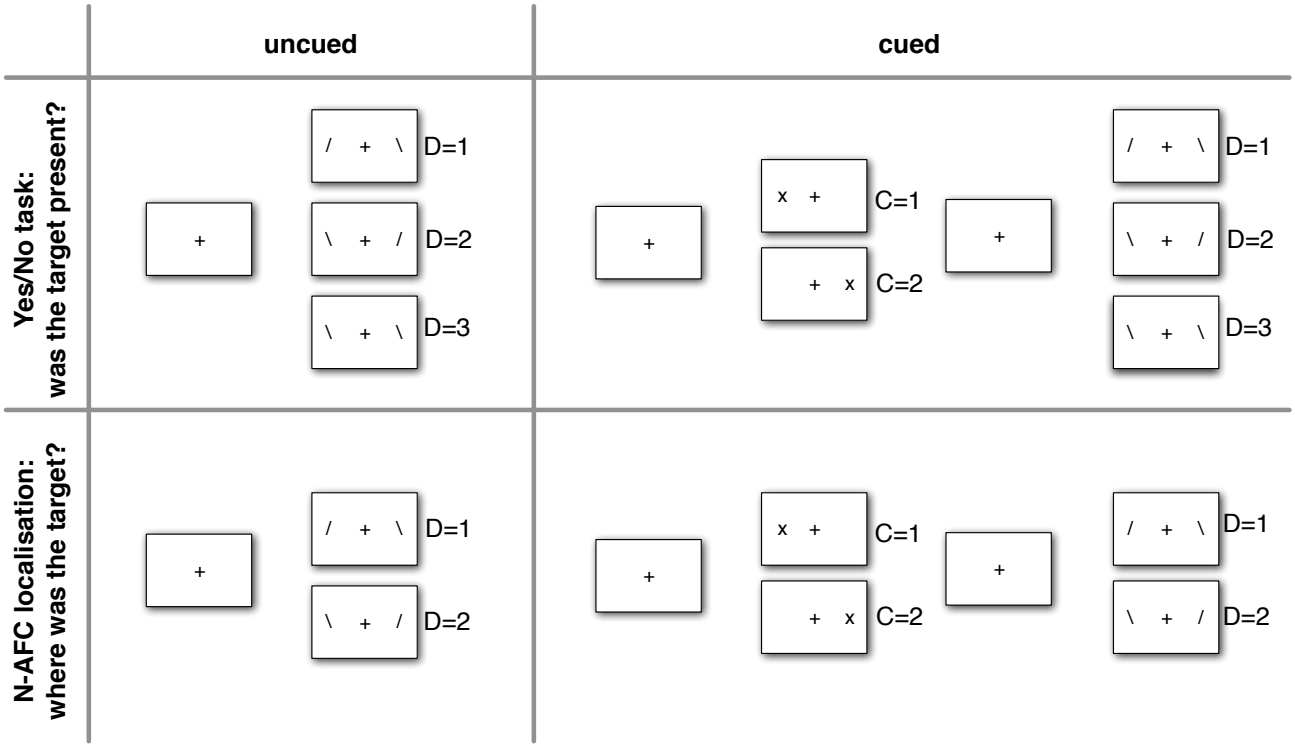

Fig. 2 Overview of the trial structures for both cued and uncued versions of the yes/no and localisation tasks. Examples are shown for $N = 2$ display locations, but straight-forwardly extend to higher set sizes. The stimulus display is represented by oriented bar items, targets are rotated clockwise from vertical (/), distractor items are rotated anti-clockwise from vertical (\). In spatial alternative forced choice with N display items (N-SAFC localisation), the observers task is to indicate the location of the target item, in yes/no the observer's task is to respond if the target is present or absent. In cued experiments, a short inter-stimulus interval occurs after the cue to ensure identical pre-stimulus visual transients.

theory provides a way of doing this, Baye's theorem shows us how to combine our prior expectations about the state of the world with our current sensory observations. A second appeal of Bayesian approaches is that by describing the generative structure and statistics of the environment, they fulfil an important aspect of Anderson's approach of adaptive rationality.

In the experiments considered, observers are asked to indicate either the location, or the presence or absence of a target item, and so the possible state of the world is conveniently limited to just a few possible display types (see Figure 2). For example, in a 4 spatial alternative forced choice (SAFC), where observers must indicate the location of a target item, there are only 4 possible display types (which we shall call D) corresponding to the true location of the target. In a yes/no task with 4 display items, there are now 5 possible display types due to the additional target absent display type.

The first step proposes that observers have a 'forward model' of how the true state of the world maps on to possible sensory observations $\mathbf{x}$ and represents an observer's internal mental model of the task [2]. This could also be called a causal model, or a generative model and could be summarised with the likelihood term $P(\mathbf{x}|D)$, the probability of the observed sensory data given a particular state of the world. Knowledge of the generative structure of the task could be imparted to the observer by verbal instruction or through experience of practice trials.

The second step involves the observer solving the inverse problem: that is, using their causal model in reverse working from observed sensory data to an inferred state of the world. This can be summarised as the posterior $P(D|\mathbf{x})$, and results not in a single most probable state of the world, but a distribution of belief over all possible states of the world (display types), constrained by the observed data. This second step, of solving the inverse problem is where Bayesian inference is used. Bayes' theorem shows that our beliefs about the world (each display type $1,\ldots,J$) can be updated in the light of new data,

$$P(D_i|\mathbf{x}) = \frac{P(\mathbf{x}|D_i)P(D_i)}{\sum_J P(\mathbf{x}|D_j)P(D_j)} \tag{1}$$

The mathematical definition of the forward model for a given experimental paradigm, and the steps used to conduct

[2] Bold symbols represent vectors, for example $\mathbf{x} = (x_1,\ldots,x_N)$ where N equals the number of display items. The display type on each trial D however only takes on one value where $D = \{1,\ldots,N\}$ for localisation, or $D = \{1,\ldots,N+1\}$ for the yes/no task.

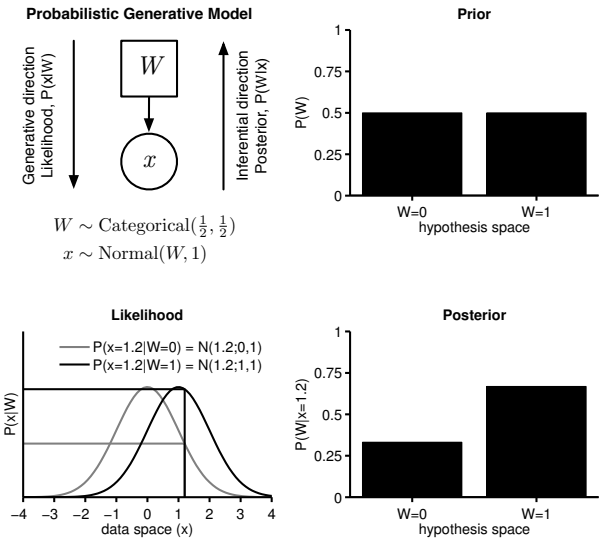

Fig. 3 Bayesian inference for a simple detection task with a single display item. The probabilistic generative model shows that the state of the world W can take on values of 0 (stimulus absent) or 1 (stimulus present), and an observer assumes these have equal prior probability. The model also defines the likelihood term $P(x|W)$, the probability of observing a particular value of x given a true state of the world W. The likelihood functions can be seen as neural tuning curves for stimulus absent (grey curve) or present (black curve). The likelihoods for a particular observations $x = 1.2$ are used in Bayes' Theorem to calculate the posterior, see text for details. We see that observing $x = 1.2$ has increased our belief that the target was present from 50% to 66.8%.

the Bayesian inference are a blessing and a curse. While the formal definition of the model offers all the advantages of a precise, unambiguous, and replicable quantitative model (Farrell and Lewandowski, 2010), it could arguably act as a barrier to understanding the core theoretical claims being made. This tutorial review attempts to avoid this issue as much as possible by using the expressive Graphical Modelling syntax (Jordan, 2004; Lee and Wagenmakers, 2014).

1.3.2 A worked example

Before describing how the Bayesian approach can be applied to the 4 covert search tasks in Figure 2 we work through a simple yes/no example (see Figure 3). Interested readers can work through this section in conjunction with the Matlab code `bayes101.m`. Observers are exposed to trials where either a single item is present or absent, and their task is to indicate which it is. The presence or absence can be thought of as the true state of the world W. Observers do not have direct access to the true state of the world however, only to a noisy sensory observation x.

This task and stimulus environment can be compactly represented by a probabilistic generative model (Figure 3, top left) as follows:

$$W \sim \text{Categorical}(\tfrac{1}{2}, \tfrac{1}{2}) \tag{2}$$

$$x \sim \text{Normal}(W, 1). \tag{3}$$

Equation 2 defines a uniform prior $P(W)$ over the two states of the world $W = \{0, 1\}$. Equation 3 is the likelihood function $P(x|W)$ and defines sensory observations to be normally distributed, centred upon the true state of the world with an observation noise variance of $\sigma^2 = 1$.

Step 1: Generate simulated data. We can use the probabilistic generative model to simulate a single trial, proceeding in the direction of the arrows shown in the model. First the state of the world is determined by sampling from the prior. In this case it is equivalent to tossing a fair coin, and the result was a signal present trial (W=1). While we as an experimenter know this, the simulated Bayesian observer does not. Next, a simulated sensory observation is made by sampling from the distribution $x \sim \text{Normal}(1, 1)$, and the result is $x = 1.2$.

Step 2: The observer conducts inductive inference, proceeding from the observed value x to the state of the world W. Observers will do this using their model of the task and stimulus environment (ie. the generative model) which includes a prior, and the observed data. Observers do not just estimate the most likely state of the world, but a distribution of belief over each possible state of the world. In this example, this equates to having a degree of belief that the signal is present $(W = 1)$ or absent $(W = 0)$. The observer's prior over states of the world $P(W = 0) = 0.5$ and $P(W = 1) = 0.5$ are updated in the light of the observation $x = 1.2$ using Bayes' Theorem (Equation 1), which involves combining prior and likelihood. The likelihood (Equation 3) can be thought of as a neural tuning curve (Figure 3, bottom left), one representing what distribution of observations would be expected for signal absent trials, and another for signal present trials. Using this interpretation, the likelihood represents the activity of a neuron with a tuning curve matched to the stimuli expected for each possible state of the world (Zemel et al., 1998; Pouget et al., 2000). The posterior belief in each state of the world is calculated such that their belief is now updated compared to their prior (Figure 3, right). Because we only have two mutually exclusive states of the world, we can calculate the posterior probability of target presence, given the observation x, as

$$\begin{aligned} P(W = 1|x = 1.2) &= \frac{P(x = 1.2|W = 1) \times P(W = 0)}{P(x = 1.2|W = 0) \times P(W = 0) + P(x = 1.2|W = 1) \times P(W = 1)} \\ &= \frac{N(1.2; 1, 1) \times 0.5}{N(1.2; 0, 1) \times 0.5 + N(1.2; 1, 1) \times 0.5} \\ &= \frac{0.3910 \times 0.5}{0.1942 \times 0.5 + 0.3910 \times 0.5} \\ &= 0.6682. \end{aligned}$$

and target absence as $P(W = 0|x = 1.2) = 1 - 0.6682 = 0.3318$.

Step 3: Make a decision based upon the posterior belief. Unbiased observers will indicate the signal is present if $P(W = 1|x) > P(W = 0|x)$, which in this example trial would be the case as the observer believes there is a 66.8% probability that the signal was present.

In order to obtain predicted performance of this observer, many trials would be simulated where accuracy of the observer's decisions are evaluated. In this example, the noise variance σ^2 is a free parameter of the model which needs to be estimated from experimental data. This parameter estimation step is important in many of the modelling studies reviewed, but is not discussed here as it is not central to understanding the theoretical assertions of the approach.

1.3.3 Bayesian optimal observer models

A distinction can be made between the claim that observers conduct Bayesian inference, and that they do so optimally (Ma, 2012). Models of the latter type are Bayesian optimal observers (or ideal observers) and their utility lies in the comparison of human performance to a theoretical ideal. Discrepancies between human performance and this ideal, if there are any, provide clues to inspire further hypothesising (Geisler, 2011). Optimal observer models are therefore not necessarily put forward as complete hypotheses for how people act in the world, as they are highly customised to calculate best possible performance in specific situations. Many of the experimental phenomena reviewed in Section 3 are well described by optimal observer models. However, there are many ways in which observers can conduct Bayesian inference, but fall short of optimal performance (see Section 4.2), and a specific case study is highlighted in Section 3.2.1. The following section outlines Bayesian optimal observer models and their predictions in 4 simple covert attention paradigms shown in Figure 2.

2 Bayesian optimal observer models and predictions

The steps involved in the practical evaluation of the models presented below are outlined in the Supplementary Material. Matlab code is available to download from `https://github.com/drbenvincent/BayesCovertAttention`.

2.1 Inferences

Looking at the trial structures of the 4 experimental paradigms considered (Figure 2) we can see that these are not completely unrelated tasks. We can describe uncued yes/no and uncued localisation with a single probabilistic generative model (Figure 4, top), and we can describe cued yes/no and cued localisation with another model (Figure 4, bottom). In both cases the observer infers the display type. For localisation the observer infers which of N locations contains the target. In the yes/no task, the observer makes inferences about which of $N+1$ display type was shown. That is, was the target present $D = \{1, \ldots, N\}$, or absent $D = N+1$.

For the uncued tasks, the model (Figure 4, top) can be read in the forward generative direction as follows. On each trial a display type D_t is sampled from a prior distribution $\mathbf{p}$, that is, a display type is selected as the outcome of a biased roll of a dice. For example, with a set size of $N = 2$, this bias (or prior over display types) is $\mathbf{p} = [0.5, 0.5]$ for localisation, and $\mathbf{p} = [0.25, 0.25, 0.5]$ for yes/no. The display type then specifies the experimental stimuli, targets (with a feature value of 1) and distracters (feature value 0) and their locations. The observer then makes noise corrupted sensory observations $\mathbf{x}_t$ of the true stimulus. We assume this observation noise is normally distributed, centred on the true stimulus value, and with a specified variance. Because some features are encoded with greater sensory precision than others (eg. cardinal versus diagonally orientation stimuli), the variance of this observation noise is not assumed to be equal for targets σ_T^2 and distractors σ_D^2.

This generative model is then used in reverse to make inferences. Because the models here are more complex than the simple worked example in Section 1.3.2, it is challenging to concisely describe how inferences are made. Interested readers are directed to the Supplementary Code to get a more thorough insight, but it is possible to summarise the inference process as follows. Based upon the noisy sensory observations $\mathbf{x}$, the observer uses the probabilistic generative model to infer a posterior distribution of belief over display types D_t. The resulting posterior probability of belief over display types is then used to make a response decision, see next section.

The cued tasks are similar to the uncued tasks in that observers infer the display type, but now the cue provides a further source of information about the display type to the observer. A second probabilistic model (Figure 4, bottom) can be used to model both cued tasks. The only addition to the model is that the prior probability of each display type is updated on every trial $\mathbf{p}_t$, incorporating knowledge of the cue validity v and the observed location of the cue c_t. For example, if a 70% valid cue is observed in location 1 of 2, then the prior over the target location is $\mathbf{p}_t = [0.7, 0.3]$. The rest of the model is identical to the non-cued tasks.

Because these are Bayesian optimal observer models, the observer also has precise knowledge of observation noise variance for targets σ_T^2 and distracters σ_D^2, the prior probability $\mathbf{p}$ of each display type, and for cued tasks, the location of the cue c_t and the cue validity v.

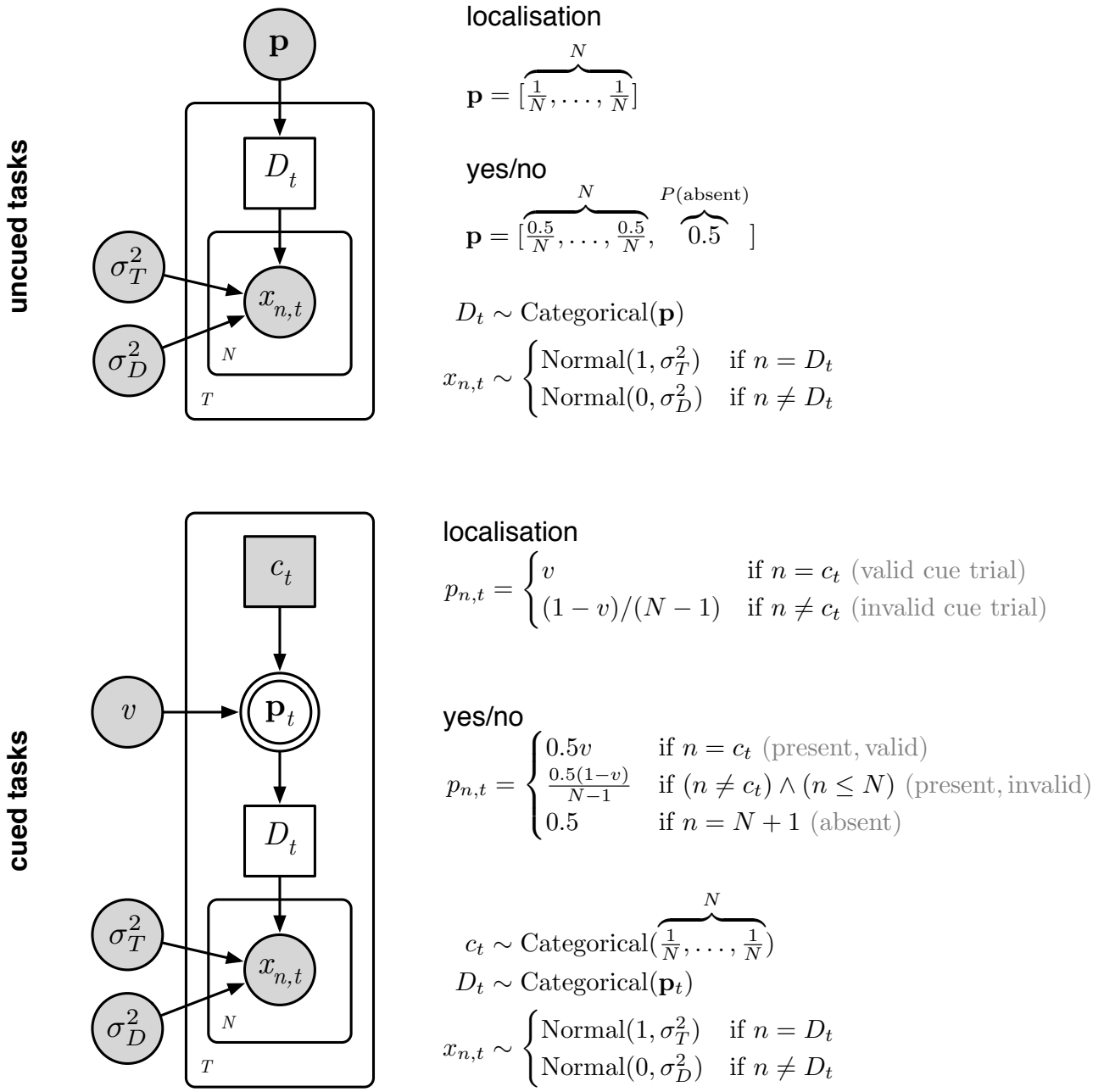

Fig. 4 Bayesian models of the yes/no and localisation tasks, for both uncued (top) and cued (bottom) variants. For the uncued tasks, display types on each trial D_t are sampled from a prior distribution $\mathbf{p}$. On each trial, noisy sensory observations $\mathbf{x}_t$ are made of N targets and distractors. For the cued tasks (bottom) the observer's prior over display types (target locations) are influenced on a trial-to-trial basis by the cue location c_t and the cue validity v. Circles represent continuous variables, squares represent discreet valued variables. Double bordered nodes represent deterministic relationships, otherwise the relationships between connected nodes are stochastic. Larger boxes (plates) represent for-loops over either N display items or T trials. Shaded nodes represent observations made by the optimal observer. See Supplementary Material for more detail.

2.2 Decisions

While the nature of the inferences made by observers in the yes/no and localisation tasks are the same, the way that an observer translates these into decisions varies depending upon the task. In the localisation task, after having inferred a posterior distribution of belief over display types (target location), the observer simply responds to the location with the greatest degree of belief, the posterior mode, also termed the maximum a posterior (MAP) estimate (see Figure 5, left).

The yes/no task requires the observer to indicate if the target was present or absent. It is straightforward to calculate a decision variable for this task from the posterior over display types by computing the probability that the target is

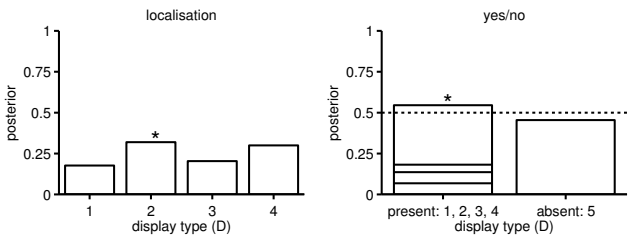

Fig. 5 Decision rules for each task. For localisation, the observer responds to the location with the highest posterior probability of containing the target. For yes/no, the observer responds yes if the probability of the target being present (the sum of display types 1 to N) is greater than 0.5.

present, $P(\text{present}) = 1 - P(\text{absent})$ where $D_t = N + 1$ represents a target absent display type (see Figure 5, right). The $P(\text{present})$ decision variable is used to calculate ROC curves describing an observer's performance in the next section. And hit rates and false alarm rates can also be computed if we assume the observer is unbiased, responding 'yes' if $P(\text{present}) > 0.5$.

2.3 Optimal observer predictions for the uncued yes/no task

Figure 6 shows predicted behaviour of a Bayesian optimal observer in the yes/no task. Technically, a Bayesian optimal observer does not have a free response threshold parameter (as described above, they respond 'yes' if $P(\text{present}) > 0.5$), but for purposes of illustration Figure 6(a) shows ROC curves if this threshold were to vary. Because we, as experimenters, know the true display types, then we can extract a distribution of decision variables for target present and target absent trials, and then simply compute the ROC curves from these target present/absent distributions of the decision variable. The plot shows the ROC curves improving (increasing in their area under curve, AUC) as the target distracter discriminability (d') increases.

The model was used to replicate set size effects, similar to Eckstein et al. (2000). Performance in terms of AUC was calculated as a function of set sizes, for a range of different target distracter distances (d'), Figure 6(b).

The model also demonstrates the search asymmetry effect in the form of predicted ROC curves for two detection searches with a set size of 2 (Figure 6c). The first is when targets have higher internal observation noise associated with them $\sigma_T^2 = 4, \sigma_D^2 = 1$. The second is when the identities are switched such that distractors now have the higher level of internal noise associated with them, $\sigma_T^2 = 1, \sigma_D^2 = 4$. Notice that performance is better (seen as higher AUC) when the distracters have higher encoding precision than targets. This is initially counter-intuitive, but it is a straight forward result due to the distracters contributing less noise to the decision variable compared to when distracters are encoded with higher precision. In summary, a Bayesian optimal observer account of search asymmetry effects is simply that different stimuli can be encoded in our visual systems with different levels of precision.

2.4 Optimal observer predictions for the cued yes/no task

The yes/no task has also been examined in conjunction with a cue (eg. Shimozaki et al., 2003). Figure 7 shows predicted cuing effects (hit rate advantage for the cued versus uncued locations) in a range of situations. The cue has the effect of updating an observer's prior belief about the upcoming target location. For a set size of 2, if the cue is predictive of the cued location ($v > 0.5$) then the observer has an increased belief that the target will occur at the cued location, and a performance benefit is conferred (Figure 7, left). In non-Bayesian terms one might typically read the assertion that 'attention is allocated to the cued location' but attention in this sense is often ill-defined. When the cues are counter-predictive of the target location, then the performance benefit is conferred to the uncued location (negative cuing effect). This means that an optimal observer should decreased their degree of belief that the target will occur at the cued location, and increase it at the uncued location (eg. Eckstein et al., 2004; Vincent, 2011a). The beneficial effect of a cue is also dependent upon the noise variance (d') and the set size (Figure 7, middle). The peak cueing effect increases with set size, as the cue conveys greater amounts of information to the observer when the number of possible target locations are high (also see Figure 7, right).

2.5 Optimal observer predictions for localisation and cued localisation tasks

Figure 8 (thick lines) shows predicted performance in the localisation task for set sizes of 2, and 4. In each case, the spatial prior of where targets appear was manipulated. The probability of the target occurring in location 1 was manipulated in 9 conditions between 0% to 100%, with equal probability of the target occurring in the remaining locations. In other words, the amount of prior information available to the optimal observer was varied. The predicted performance is intuitive. Firstly performance was higher for higher d' values (achieved by manipulation of internal observation noise, σ^2). Secondly, we can see that the lowest performance occurs when the targets are uniformly distributed, that is, where the observer has no prior knowledge of the upcoming target location. This model provides a good account of a spatial probability manipulation (see Sections 3.2.1 and 4.2). Figure 8 (thin lines) shows the predicted performance in the exogenous cued localisation task for set sizes of 2, and 4. Note, that these predictions are identical to that of the endogenous spatial probability manipulation (Figure 8, thick

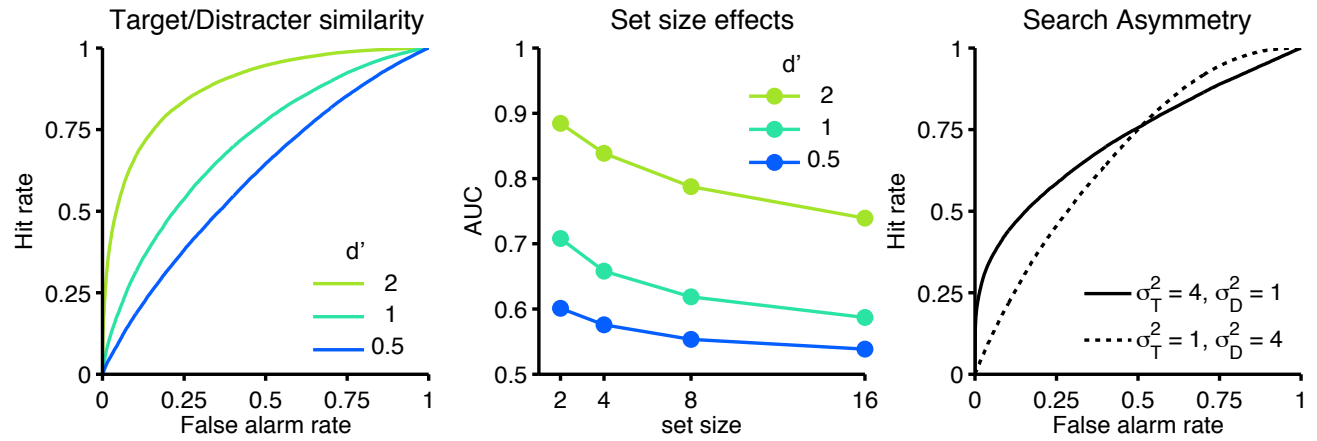

Fig. 6 Predictions of a Bayesian optimal observer model for the yes/no task. ROC curves are shown (left) for a range of d' values with a set size of 2. Set size effects are predicted (middle) for the same d' values. Search asymmetry effects are shown (right) for a set size of 2.

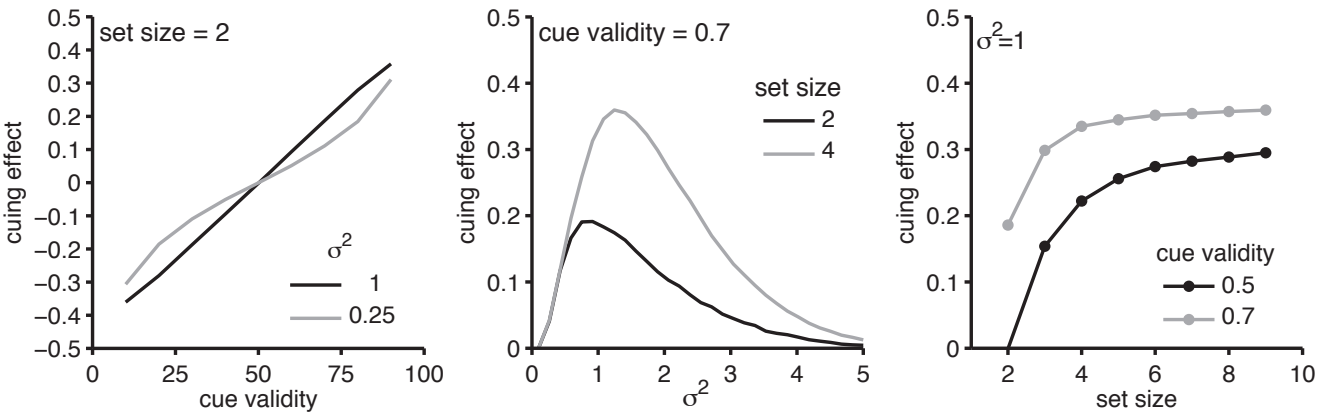

Fig. 7 Predictions for the cued yes/no task. The cueing effect ($HR_{valid} - HR_{invalid}$) is shown as a function of cue validity (left), d' (middle), and set size (right). Negative cueing effects mean an advantage to non-cued locations.

lines). This also mirrors the predictions made by (Vincent, 2011a) and provides a reasonable account of human performance (but see Sections 3.2.2 and 4.2).

3 Explanations of attentional phenomena

Having been introduced to Bayesian concepts and seen specific optimal observer models applied to 4 attentional tasks, we are in a position to generalise to the wider range of attentional effects observed in the domain of visual selective information processing with briefly displayed stimuli. While different models are formulated to account for each specific experimental task, these are all realisations of one core theoretical claim which could be described as: *Attentional phenomena are by-products of conducting inference about the state of the world.* We can use this approach to categorise a wide range of attentional phenomena, and I will present a brief, selective review of stimulus-based, and belief-based phenomena. One could also argue that a class of reward-based phenomena also exist, but these are not discussed here.

3.1 Stimulus-based phenomena

Many of what could be thought of as stimulus-based phenomena (set size effects, conjunction searches, and search asymmetries) were key experimental effects used as evidence to support well known 2-stage serial-parallel models such as Feature Integration Theory (Treisman and Gelade, 1980) and Guided Search (Wolfe, 2007). However, SDT and Bayesian approaches showed that a 1-stage, purely parallel (noise-limited) mechanism provide good accounts of these effects within the simplified performance paradigm.

3.1.1 Set size effects

As the number of display items increase, the performance at detecting presence or absence of a target amongst distractors decreases. Palmer et al. (1993) examined set size effects in 2IFC and yes/no detection tasks. Their stimuli were horizontal lines, distracters were shorter, and target lines were longer. However, rather than plotting how performance decreases as set size increases, they plot the amount of sen-

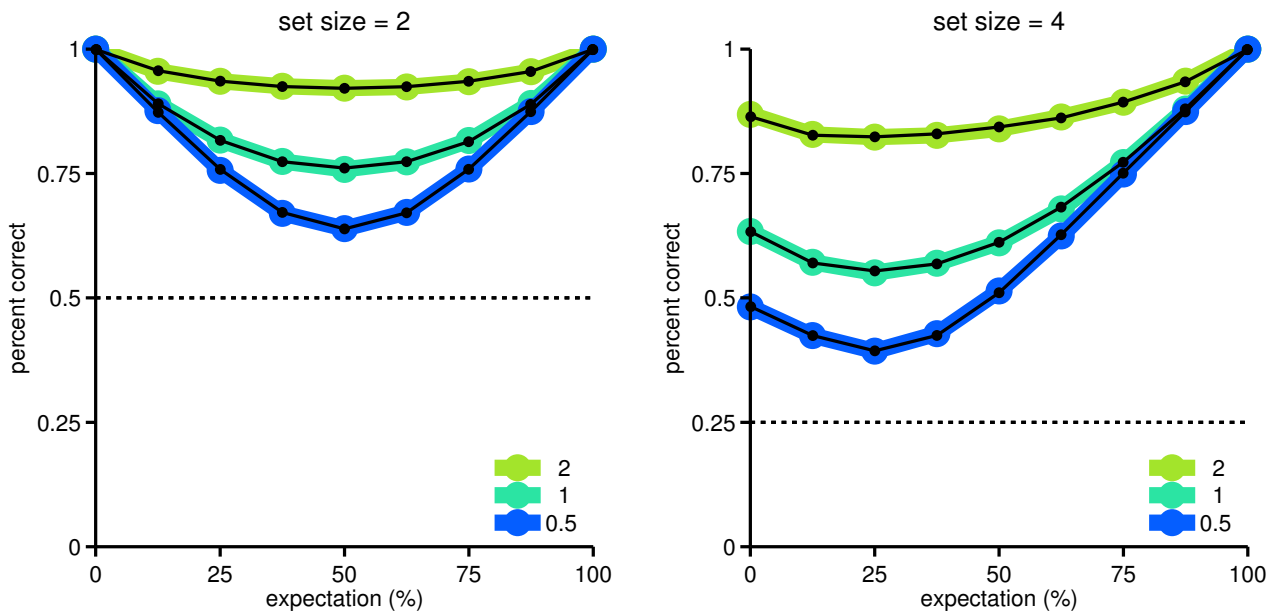

Fig. 8 Predicted proportion correct responses of a Bayesian optimal observer for the 2-SAFC (left) and 4-SAFC (right) task with spatial prior manipulation (thick lines) and corresponding *cued* localisation tasks (thin lines). In the spatial localisation task the expectation refers to the probability that a target will occur in the manipulated location, which the observer has been informed of. In the cued localisation task, expectation refers to the cue validity. Line colours represent different d' values, see legend. Dashed lines show chance performance levels.

sory evidence required to maintain a threshold performance level. They found that the amount of evidence (difference between target and distracter line lengths) increased roughly linearly (on a log-log axis of set size v.s. threshold) with a slope of 0.25 (for detection) and 0.31 for 2-interval-forced-choice (2IFC). However, using this approach allowed them to predict that these slopes (not intercepts) should be constant regardless of the stimuli used. This strong prediction matched human performance both in the 1993 paper and also for many (but not all) stimuli, such as luminance increments and the colour and size of blobs, in a follow up study (Palmer, 1994).

Control of stimulus-based factors is an important issue when studying information processing, and two important issues were addressed by Palmer et al. (1993). Firstly, are set size effects due to internal attentional factors or are they simply by-products of the stimuli or our sensory sampling of them? This was tested by seeing whether set size effects persisted even when sensory factors were controlled for in their methodological procedure: the 'performance paradigm' outlined in the introduction. Even with this paradigm it was still possible that the different numbers of displayed stimuli (display set size) could form a non-attentional contribution to set size effects, and so they compared these results to what they termed a 'relevant set size' manipulation (see Figure 9). The number of displayed stimuli remains constant, and set size is manipulated by use of bounding-box cues determining the possible number and locations of relevant stimuli on that trial or block. They found no difference between a relevant set size and a display set size manipulation, and because the former can only be interpreted as an attentional effect, they conclude that display set size manipulations are also attentional (not sensory) in origin. Their second question was to determine if these effects are caused by sensory- or decision-level mechanisms, or both (also see Section 3.2.2). By comparing model fits to data, they found a decision-based explanation could account for the results of their Experiments 1 and 2. That is, their set size effects could be accounted for purely by considering that additional display items contribute noise to the sensory signals being considered as either targets or distractors. The more display items, the higher the chance that one particular noisy observation will be mistaken for a target (false alarm).

The generality of this explanation was established by follow up studies. Palmer et al. (2000) considered a wider range of SDT models, finding that a) optimal observers, b) maximum of outputs, and c) maximum of differences models all provided good accounts of their experimental effects, including that of set size. SDT explanations were also able to account for observer's performance in a wider range of experimental tasks (Cameron et al., 2004). A 2-target paradigm was used, where targets could be either $+15°$ or $-15°$ Gabors. Their tasks asked, which of two targets occurred (identification), whether either target appeared (detection), and identification of a spatial location of either target (localisation). Their SDT models could provide good accounts for human set size effects under these additional tasks. One twist on the set size effect, is that in oddity search (when the target

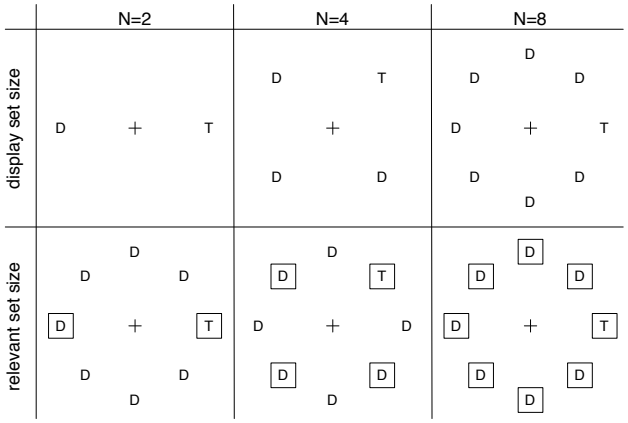

Fig. 9 Schematic plots of display set size and relevant set size manipulations for a localisation task. T and D represent target and distractor stimuli, respectively.

is defined as being different from distractors, but the feature properties of targets and distractors are unknown in advance) then the set size effect is either very shallow or flat. Schoonveld et al. (2007) showed, in a 2AFC task (target in group 1 or group 2) the shallow set size effect was simply a by-product of conducting inference with the observed stimuli in the context of this particular task structure, no other mechanisms were required to account for the effects.

In summary, the set size effect can be understood fairly intuitively. Taking yes/no detection of a target as an example, observer's responses of target presence/absence, is determined by an inference based upon N noisy sensory observations. As the number of display items decrease, then the number of items that could potentially be confused for a target decreases giving rise to more accurate responses and higher levels of performance. Therefore, we have a consistent information processing mechanism which makes inferences based on a particular set size. The change in performance as a function of set size can then be attributable only to the experimentally determined set size, and so the set size effect is a by-product of increasing the number of stimuli being processed.

3.1.2 Distracter heterogeneity effects

It is rare, in naturalistic situations, that a target could be present amongst a set of entirely uniform distractor items, normally these distractor items vary. To study the effects of this heterogeneity, additional *external* noise (feature jitter) is often added to distracters. While previous studies had demonstrated a clear cost of increased distracter heterogeneity (e.g. Duncan and Humphreys, 1989), only later did the effects receive quantitative treatment and support from SDT models (Palmer et al., 2000). Distracters were vertical lines, and the orientation offset of a target required to achieve a threshold performance was determined. When switching to a noise condition where distracters had feature jitter ($\sigma = 4^\circ$), targets then had to be offset further from vertical to achieve the same level of performance. Palmer et al. (2000) found that optimal observer (and other SDT) models could quantitatively account for this increased sensory evidence required over a range of set sizes.

In a yes/no detection task, some initial evidence showed performance was explicable by Bayesian optimal use of sensory information (Vincent et al., 2009). Distractor heterogeneity was manipulated on a block-wise basis. In this experiment, the targets were Gabors oriented 0° from vertical, with no external feature noise. Distracter orientations were sampled from a Normal distribution with the same mean orientation as the target, but external feature jitter was manipulated. As distractor feature jitter was increased, target detection performance *increased*. Initially this may sound in conflict with the results of Palmer et al. (2000) where adding distracter jitter decreased performance (thus requiring greater feature separation between the target and distracters), but is merely due to a difference in task (see Figure 10). In both cases, performance decreases as feature overlap between targets and distracters increase, as there is an increased chance for distracters to be confused for a target (false alarm) for example. This is powerful as the approach can account for how distractor heterogeneity can both increase and decrease performance in different situations. What matters is not distractor heterogeneity as such, but the degree of stimulus overlap between targets and distracters. A Bayesian model was able to provide a good account of how performance increased with distractor noise, as well as the shapes of the underlying ROC curves (Vincent et al., 2009). However, despite the claims of this model being optimal, it had some limitations in that it only made locally (not globally) optimal decisions. Stronger evidence was provided by Ma et al. (2011). Targets were defined by orientation, but stimulus reliability was manipulated (by item contrast) on a trial-to-trial basis. This meant that the observer was faced with a set of distractors whose variability was uncertain from one trial to the next. Their globally optimal Bayesian observer provided a good account of human performance, and provided strong support for the idea that the reliability of sensory information is continuously assessed.

In summary, distractor heterogeneity impacts performance as a direct result of observers making Bayesian inferences about the display type where an *external* source of uncertainty is added to distractors.

3.1.3 Search asymmetry effects

Search asymmetry effects occur when the search for a target item A amongst distractors B gives rise to a different level of performance than searching for a B target amongst A distractors. The Bayesian explanation of search asymmetry effects

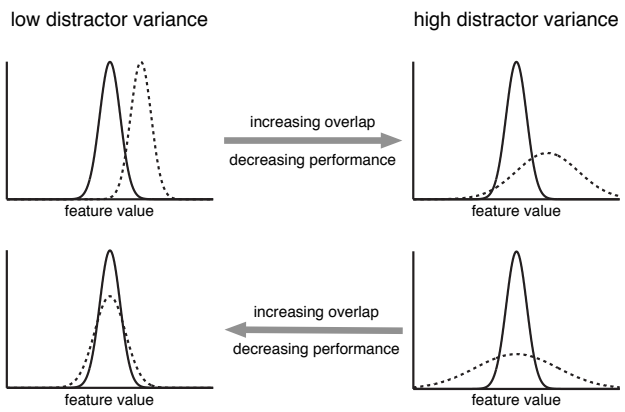

Fig. 10 Increasing distractor heterogeneity can both decrease performance (Palmer et al., 2000, top) and increase performance (Vincent et al., 2009, bottom), depending upon the stimuli. The distributions of targets (sold lines) and distractors (dashed lines) reflect both internal observation noise and external, experimenter added feature jitter. The results are due to the differential effect of heterogeneity upon target and distracter overlap.

is near-identical to that of distracter heterogeneity effects, in that there is differential sensory uncertainty associated with targets and distractors. Except that search asymmetry effects represent an *internal* source of uncertainty difference associated with different stimuli. The notion that search asymmetries could be accounted for by differences in the sensory uncertainty associated with display items A and B was operationalised by Palmer et al. (1993). The magnitude of the asymmetry effect should then relate to how far the sigma ratio (σ_A/σ_B) deviates from 1. For example, search for a tilted line amongst vertical lines is easier than the converse because there is a lower chance that one of the vertical lines (with lower associated sensory noise) will be mistaken for a tilted target.

Initial evidence in a standard RT paradigm search was provided by Carrasco et al. (1998) using oriented line stimuli. They also propose that asymmetry effects can be accounted for by a single parallel mechanism which processes sensory information, where the tuning bandwidths is greater for tilted lines. Simple cells of the primary visual cortex could provide a plausible neural basis for this, both because of the number of cells tuned to cardinal directions and because of their narrower tuning bandwidth (Li et al., 2003). Dosher et al. (2004) used a speed accuracy tradeoff paradigm, and their modelling work supported a parallel mechanism underlying search asymmetry effects. Further empirical and modelling (Bayesian and SDT) results confirmed this sigma ratio (differential uncertainty) explanation in a short display duration performance paradigm (Vincent, 2011b; Bruce and Tsotsos, 2011). In summary, search asymmetry effects are the result of conducting Bayesian inference upon sensory observations of stimuli A and B, where the level of *internal* noise (or encoding precision) is not the same for each item.

3.1.4 Conjunction search effects

The phenomena discussed up to this point relate to simple feature search, where targets and distracters take on values along a single dimension such as orientation or contrast. One very small step toward a more realistic stimulus environment is to consider what happens when targets and distracters are defined by combinations of features. Conjunction search tasks examine this case, where targets are now defined as the combination of two particular feature values (such as a red square) where distracters take on only one of those properties (so there are distractors that can be either red circles, or green squares). The basic effect of defining targets by combinations of features is to lower the performance of observers, as compared to searches for each individual feature search. From a SDT approach, the intuition for this effect is that the d' of a conjunction search will be worse by a factor of $\sqrt{2}$ (assuming statistical independence of the feature dimensions) because the uncertain sensory observations are being projected onto a decision axis combining information from 2 feature dimensions. Put a different way, for a correct detection to occur the stochastic noise in a conjunction search could potentially make the target appear to look like a distractor not just in one dimension, but in two.

The SDT approach was extended from single-dimension feature search to multiple feature conjunction search by Eckstein (1998). A 2IFC task was used to map performance as a function of set size. This performance curve was high for each individual feature search in isolation, but the performance curve was decreased in the conjunction search condition. Predictions of SDT models provided a much better accounts of human search performance as compared to serial, and hybrid noisy serial models. Eckstein et al. (2000) replicated effects for feature and conjunction but test the account further in disjunction (e.g. targets red circles, distractors green squares) and triple conjunction displays. While a serial model could be rejected, it was unclear which of two possible SDT decision rules provided the best fit of the performance data across 3 subjects.

One of the powerful aspects of the parallel SDT models is that performance as a function of set size can be predicted for both individual features searches, and the conjunction search. Further, these d' parameters used for conjunction search predictions are not free parameters, but are determined from each separate feature search. There is nothing different about information processing of stimuli with multiple feature properties, the change in performance simply reflects parallel information processing of uncertain sensory data.

3.2 Expectation- or Belief-based phenomena

If we wish to learn about internal information processing underlying attentional effects then it is important to exclude uncontrolled external stimulus-based factors from consideration. When this is done in the performance paradigm, the experimental effects that I have described as 'stimulus-based' show that no internal attentional mechanism is required to account for the data. Instead they can be seen as by-products of experimentally manipulating stimulus characteristics. This places the locus of these effects externally, into the environment. But there *are* attentional phenomena influenced by internal processes, namely an observer's beliefs about the state of the world.

3.2.1 Spatial probability effects

We live in a highly structured world where objects are not uniformly distributed, so it would seem plausible to assume that we can learn and utilise spatial distributions of where targets are more likely to occur. But do we learn such spatial distributions optimally and is this combined with visual cues of the target's location? Promising early evidence came from Shaw and Shaw (1977) who used a spatial probability manipulation in a task requiring recognition of a letter stimulus. Letters could appear close to the fovea (1°) in one of 8 locations. In a uniform condition the letter had an equal probability of appearing in each location and the display duration was such that identification performance was approximately 68%. In a non-uniform condition, the location of stimuli was determined by a spatial prior distribution which the subjects had become familiar with in practice sessions. In this non-uniform condition where some locations had a much greater and some had a much lower probability of containing the target, identification performance increased to around 71%. Interestingly, the identification performance in the high probability regions was higher ($\sim 80\%$) than in the low probability regions ($\sim 35\%$). Their model, not framed in SDT or Bayesian terms, suggested that the distribution of search resources was proportional to the prior probability distribution for each condition. In other words, observers were sensitive to the environmental statistics governing target location.

Further evidence to suggest we utilise spatial prior probability distributions was provided by Druker and Anderson (2010) using a choice reaction time measure in the judgement of the colour of a single dot. Their first spatial probability distribution was of a mixture of a uniform distribution across the display and a strong 2D Gaussian distribution to the side of central fixation. Reaction times were faster to the high probability side of the screen, and also increased as a function of distance from the center of the high probability region. These effects were not attributable to retinal eccentricity, nor speed accuracy tradeoffs. While this provided further evidence for use of spatial prior expectations, without formal modelling of the RT data it is not possible to address the question of how optimally observers were learning or utilising the spatial priors.

Evidence that observers do near-optimally utilise target location probability was provided by Vincent (2011a). In one endogenous cuing condition, observers indicated which of 4 locations contained a target amongst 3 distractors. The spatial prior distribution was altered such that one spatial location (which the observer was informed of) had a certain probability of containing the target, while it was uniformly distributed amongst the remaining 3 locations. The performance of observers in this 4SAFC task matched the predictions of a Bayesian optimal observer (see Figure 8, thick lines). This provided strong evidence that people were combining (in a Bayesian manner) their spatial prior expectations and their uncertain sensory observations of the targets and distracters. However, inspection of slight deviations between the predicted and actual performance showed that observers had probability biases. In low probability conditions (where a location was chosen to have a lower than chance level of occurring) observers acted as if they overestimated the spatial prior of the target occurring at that location. In the high probability conditions, they acted as if they underestimated the probability. This pattern of probability bias has been extensively observed and is the same pattern that Prospect Theory describes (Kahneman and Tversky, 1979). So while the results of Vincent (2011a) show that observers are combining observations with their spatial expectations, there exist non-normatively rational biases in what those expectations are (see Section 4.2).

3.2.2 Spatial cuing effects

In the SDT framework, the two possible ways in which a cue could affect the ability to localise a target is through a sensory- or decision-level mechanism. The sensory-level explanation (also termed signal enhancement) is that observers have a finite set of sensory resources, and the effect of the cue is to reallocate those resources such that the d' sensitivity (or signal-to-noise ratios) are changed in favour of the cued location. Formal modelling of the sensory-level explanation in terms of resources was provided by Eckstein et al. (2009). The alternative, but not necessarily mutually exclusive explanation, is that the cue has its affects at a later decision-level stage (also termed noise reduction, uncertainty reduction, response criterion shifts, or updated prior expectations). Cues reduce the uncertainty about the upcoming target location by updating a spatial prior belief of where the target may occur, given the information imparted by the cue (see Figure 4, right). For example, with a 100% valid cue, uncued locations are expected to have a 0% probability of containing the target and any stimulus-based information at

these locations only contribute noise to the decision process. This noise can be removed or decreased, enhancing performance, by down-weighting sensory contributions from these uncued locations.

While SDT models may be considered as ambivalent between these two explanations, Bayesian optimal observer models are more constrained and would not predict any changes in sensory encoding precision (although, see Mazyar et al., 2012, 2013, for effects of *set size* upon encoding precision). This is theoretically important because this prediction is a direct consequence of the statistical structure of the stimulus environment. Figure 4 shows generative models of tasks, which observers putatively use (as an internal mental model) as the basis for making inferences of the target's location given the cue location and the noisy sensory stimuli. There is nothing in the generative structure of the cued localisation task linking the cue location to the standard deviation of sensory noise, therefore the encoding precision of stimuli is expected to be statistically independent from the cue location.

But what does the behavioural evidence show in terms of the short display duration performance paradigm? There certainly is support that signal enhancement (encoding precision effects) occur under some circumstances (Bashinski and Bacharach, 1980; Müller and Humphreys, 1991; Downing, 1988). However, the conditions under which these occur seem to be limited to studies which use backward masks (Smith, 2000). It was also found that there is no capacity limit to these effects, as sensitivity increases have been observed for multiple locations simultaneously (Solomon, 2004). Therefore, while sensitivity changes can and do occur, Solomon suggests this could be due to a non-attentional process. Instead, the balance of evidence seems to favour a decision-level locus as a robust explanation for cuing effects (Müller and Findlay, 1987; Palmer et al., 1993; Palmer, 1994; Eckstein et al., 2002; Shimozaki et al., 2003; Eckstein et al., 2004; Gould et al., 2007; Vincent, 2011a; Shimozaki et al., 2012; Eckstein et al., 2013).

How do these decision-level accounts work in detail, from a Bayesian optimal observer perspective? Put simply, according to the Bayesian optimal observer approach, cuing effects are the result of an updated internal prior belief of where a target may occur (see Figure 4, bottom). We could say the sequence of events are as follows: An observer has a degree of belief that a target could be in 1 of N locations, thus we have N hypotheses. At the beginning of a trial, we may assume that an observer has no information about where the target may occur, and their prior expectations of each hypothesis being true is uniform. When the cue appears, the observer updates their prior beliefs, given knowledge of the cue validity. And when the stimuli appear, the prior belief is combined with the likelihood of each hypothesis. This likelihood can be thought of as how consistent all of the stimuli are with the hypotheses that the target is present in each location. One way to summarise how this combination-step works is that the sensory information is weighted by the prior belief. However, it is not the noisy sensory information itself which is weighted (as in Kinchla, 1977; Kinchla et al., 1995, and SDT models), but it is the likelihood of the sensory data which is combined with the prior belief (Shimozaki et al., 2003; Vincent et al., 2009).

In contrast to what one may predict from the findings from 'attentional capture' it is clear that observer's weightings (prior beliefs) are not drawn reflexively to cues, but utilise the information provided by the cue. For example, cued locations are ignored (weighted at zero) when the cues are 100% invalid (Eckstein et al., 2004). If the cue validity is greater than $1/N$ then the prior belief at the cued location will increase, and vice-versa. This is predicted in Figure 8. If belief in the target's location was always increased in a cued location, even when the cue validity indicates this is less likely, then performance would decrease when cue validities are counter-predictive. However, this is not the case, observers utilise the information imparted by the cue to update their beliefs (Eckstein et al., 2004; Vincent, 2011a).

There is reasonable evidence that the specific cueing effects seen in these highly simplified paradigms may well be functionally explicable by a decision level change in prior beliefs. But these SDT and Bayesian models are simple and in no way capture the complexity of the neural mechanisms underlying the behaviour of observers. The more detailed neural mechanisms involved in attention are perhaps better left to other classes of models such as perceptual template models (see Lu and Dosher, 1999, 1998, 2014; Dosher and Lu, 2000; Carrasco, 2011), neural population coding models (Pouget et al., 2000; Ma et al., 2006; Beck et al., 2008; Borji and Itti, 2014), and predictive coding (Rao, 2005; Spratling, 2008).

3.2.3 Target prevalence

The majority of yes/no studies have utilised a target prevalence of 50%, however many interesting real world searches involve rare targets, such as prohibited items in airport baggage screening. Knowing whether human search performance exhibits biases (harming their performance compared to optimal) would be of practical importance (Wolfe and Kenner, 2005; Mitroff and Biggs, 2013). SDT predicts that as targets become rarer, an observer's performance (ROC curve and d') should remain constant, but where they position themselves on this curve (their response criterion) should become more conservative in order to maximise performance. A more natural way to express this in a Bayesian manner is that: decreasing target prevalence leads observers to require more visual evidence to overcome their elevated prior expectation of target absence. Studies have broadly found this

to be the case, an observer's response criterion shifts in a more conservative direction, leading to a decreased hit rate. In other task domains, and in the absence of reward manipulations, this shift in response criterion is near-optimal (Maddox, 2002; Kubovy and Healy, 1977; Healy and Kubovy, 1981). There is also some evidence from a covert yes/no detection task, that human observers quickly learn to optimally place their response criterion so as to maximise rewards (Navalpakkam et al., 2009).

4 Discussion

4.1 Bayesian models: under-constrained and weakly falsifiable?

Bayesian approaches to understanding human behaviour at a wide variety of levels show great promise. While Bayesian approaches are in one sense very simple, they can be complex when a theoretical explanation is distilled down into a specific model to account for a given phenomenon. This complexity, as well as the demand for some slight conceptual shifts (e.g. effect versus cause, subject versus objective probability) quite naturally leads to skepticism towards the enthusiastic claims being made. Bowers and Davis (2012) claimed that Bayesian models have so many degrees of freedom (free parameters, specification of prior, likelihood, and utility functions) that they can account for any pattern of data. In the context of Bayesian models of covert selective attention, this claim seems rather ill-founded. Many of these models have exceptionally low degrees of freedom and almost no room for the experimenter to alter their model to fit the data.

Taking the cued localisation task as an example we can run through each aspect of the model, with the criticisms in mind. The structure of the generative model has to reflect the actual experimental task, there is no degree of freedom here. Due to this being an optimal observer model, the cue validity parameter v is fixed as being equal to what was used in the actual experiment. The parameter governing the variance of the internal noise σ^2 *is* a free parameter, the value of which can be estimated from the data (not demonstrated here). The graphical model shows that there is only a single parameter, not one for every condition, and so the effect of changing this parameter is to influence the level of performance (see Figure 8). There is no way that this model can predict a fundamentally different pattern of results, it will always predict lowest performance when observers have uniform expectation of a target's location (expectation levels of $1/N$). Could the data have conflicted with the predictions of the model? Yes, it was entirely feasible that human observers did not behave in this way. A very plausible hypothesis before observing the data would have been that a counter-predictive cue would lead subjects to reflexively (and incorrectly) allocate prior belief to the counter-predictive cue location.

Was there leeway in how the likelihood was described? The likelihoods are the relationships between a child node and its parents in the graphical models. In many of these cases the relationships are determined by the task structure, so there is no flexibility in many of these cases. The only likelihood of relevance to this point is how internal noisy observations are Normally distributed about the true stimulus location. It is true that there is leeway here, the specification of this noise as being Normally distributed is an educated guess. While a t-distribution could have been used for example, it is a very clearly stated part of the model and it is up to the authors to convince reviewers and readers that these modelling decisions are reasonable.

Was there leeway in describing the priors? Because this model is a hypothetical Bayesian optimal observer, it is assumed to completely believe an experimenter's instructions of the cue validity and that targets are uniformly distributed. The observer's prior distribution of target location was equal to the actual prior distribution governing the target's location. So there was no leeway for this optimal observer in terms of specifying its prior. The notion that priors can be chosen such that the model predictions account for the different patterns of data seems unrealistic in anything but highly simplified examples, or in complex multi-parameter models. Specification of priors, in general, can allow for some modelling leeway, but just as with any modelling approach if a particular prior distribution is required to account for data then this can either be justified through argumentation or by additional experiments.

In summary, the same process of examining free parameters and modelling leeway can be walked through with many of the SDT and Bayesian models cited here. While SDT models have some flexibility, for example in terms of decision rules, this has been the focus of explicit investigation (e.g Baldassi and Verghese, 2002) rather than picking the best on an ad hoc basis. In general there is very little scope (with even less for Bayesian optimal observers) to adjust models, parameters, or priors to fit the data.

4.2 Bayes and optimality

If optimal observer predictions match behavioural observations then we may be justified in concluding that people are Bayesian and optimal, for a given task. Many of the studies reviewed here fall into this category. However, despite the assertion of Bowers and Davis (2012), advocates of the Bayesian approach are not solely fixated upon optimality (Griffiths et al., 2012): one can be Bayesian and suboptimal (Ma, 2012). But what can be concluded when we find significant discrepancy between optimal observer predictions and behavioural data? I consider three possibilities.

People are neither optimal nor Bayesian. One of the strengths of optimal observer modelling is that the fairly restricted range of predictions means that there is ample opportunity to observe disconfirmatory experimental evidence. This could mean that people are neither optimal nor Bayesian. This does not imply that optimal observer modelling serves no purpose: it could be seen to be the start of a process, representing the best possible performance obtainable. Deviations from this baseline performance level can then be used to generate and test further hypotheses about why this sub-optimality occurs (Geisler, 2011). In order to accept the possibility that people are neither optimal nor Bayesian, the following two possibilities would have to be ruled out.

People are Bayesian, but suboptimal. There are many ways we can be Bayesian (combine prior knowledge and current sensory evidence using Baye's equation) and suboptimal. One possibility is that the Bayesian computations are suboptimal because incorrect generative models are being used by observers. Beck et al. (2012) suggest suboptimal inference is inevitable, especially in complex tasks such as object recognition where the full specification of the generative model (the physics of light interacting with surfaces) is impossible due to its complexity. Alternatively, there could be limitations upon the ability to learn and represent complex prior distributions (Acerbi et al., 2014).

Poeople are Bayesian, suboptimal for an experiment, but optimal for the real world. Optimal observer models are very specific models intended to derive the best possible performance in a given task. As such, they tend to make restrictive assumptions that are unlikely to be valid when applied to real people. I consider two examples of strong assumptions that are unlikely to be valid for human observers.

Assumption 1: An optimal observer's prior beliefs are assumed to be fixed, certain and accurate. The assumption that an observer's beliefs are fixed is also an oversimplification. Droll et al. (2009) examined how human observer's beliefs changed over time as they learnt cue validity. If an optimal observer was correctly informed that a precue has 70% validity, then it is optimal for the observer to completely believe this instruction and represent this precise knowledge as $v = 0.7$. However, in the real world, where experimenters can make mistakes or deliberately mislead observers, then it seems unwise to specify complete and total belief in the experimenter's instructions (see Fennell and Baddeley, 2012). Therefore, it would be unrealistic to assume that human observers would use this approach. Evidence for this was provided by the exogenous cueing condition in Vincent (2011a). Observer's acted as though they exhibited biases in how they mapped experimenter defined cue validity into an internal degree of belief. These biases were in line with those observed in higher-level decision making tasks, described by Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Instead, an observer who treated the experimenter's instructions of cue validity as another source of uncertain information, would be expected to be suboptimal in the narrow confines of the experiment, but more robust and adaptable to the real world. Such observers could represent their belief in cue validity as a distribution, rather than a precise value, such as $v \sim \text{Beta}(1+7b, 1+3b)$ for example, where $b \geq 0$ with higher values representing greater belief in the task instruction. However, Martins (2006) and Fennell and Baddeley (2012) make promising proposals along these lines.

Assumption 2: Experimental trials are assumed to be independent events. In these simplistic experiments, each trial is an independent event, that is the presence or absence of a target on a trial is unrelated to its presence or absence on the previous trial. Because this is true for these particular experiments, an optimal observer's generative model should reflect this fact. An optimal observer in this context will not display any trial-to-trial effects. However, there is abundant evidence that people do exhibit such effects in a range of experimental task domains (reviewed by Mozer et al., 2007). There is also an accumulating body of work suggesting that these sequential effects are not just by-products of an arbitrary mechanism, but that they reflect an observer's adaptation to the temporal statistics of a task (Yu and Cohen, 2008; Wilder et al., 2009; Green et al., 2010; Vincent, 2012; Jones et al., 2013; Schüür et al., 2013).

4.3 Beyond the performance paradigm

The theory that observers are conducting Bayesian inference about the state of the world based upon sensory observations, prior beliefs, and a generative model is well supported. However, the highly simplified performance paradigm which has enabled the theoretical assertion to be assessed by relatively simple models has its limitations. The short duration of an unchanging stimulus provides experimental control over how much information about the state of the world is imparted to the observer, but it is far removed from naturalistic behaviour. Will the Bayesian concepts established in this simplified situation extend to more naturalistic settings? There are promising signs that the Bayesian approach can provide insight in these situations.

A key limitation of many of the models described in this review is that they predict performance, and not reaction times. One way that combined reaction time and performance predictions are made is through the use of sequential sampling models (Smith and Ratcliff, 2004), which include the drift-diffusion (e.g. Ratcliff and McKoon, 2008), the LATER (Carpenter and Williams, 1995) and linear ballistic accumulator models (Brown and Heathcote, 2008). They examine how noisy sensory information is integrated over time to give rise to a perceptual decision or an eye movement (e.g. Smith and Ratcliff, 2009; Ludwig, 2012). But are these

temporal accumulation models Bayesian? It has been known that drift-diffusion models implement optimal decision making in two-choice decisions, but it was only recently that this specific equivalence was made explicit through the use of a generative model (Bitzer et al., 2014). This is an active area of research, and clearly an interesting one in establishing the extent of the insights that can be provided by the Bayesian approach.

Are the results from these simple covert perceptual decision making tasks (often with button press responses) able to drive overt saccadic behaviour? Firstly, there is evidence that saccadic behaviour (with a saccade latency measure) is sensitive to the statistical structure of the environment, observers can learn a spatial prior of target occurrence (Carpenter and Williams, 1995). Eye movements to localise a target also utilise information imparted by a precue (Shimozaki et al., 2012), although not necessarily optimally. This updating of expectations also extends beyond first order spatial statistics (a spatial prior), people are able to learn and use second order (sequential) statistics to update their expectations of a target's location (Vincent, 2012). Observers are also able to make saccades based upon prior knowledge combined with uncertain sensory information (Liston and Stone, 2008), a key component of demonstrating Bayesian processes.

Can the Bayesian approach provide insight into ongoing multi-fixation search? One approach of multiple-fixation search has been based around observers making Bayesian inferences about the state of the world, but to explore different decision/fixation policies (Najemnik and Geisler, 2005; Verghese et al., 2007; Najemnik and Geisler, 2008; Zhang and Eckstein, 2010). Other work has cast doubt on the optimality of saccadic decisions (Morvan and Maloney, 2012), showing that they do not obey normative axioms of rationality (Zhang et al., 2010). The added complexity of multiple-fixation search as compared to the covert performance paradigm is opening up a rich set of questions around how Bayesian and how optimal people may be.

4.4 Summary

Some claim that attention simply does not exist as a causal mechanism at all (Anderson, 2011). What we can be reasonably sure of is that for *these* tasks, we can clearly view covert selective attention as being a set of experimental effects. A wide range of precisely specified quantitative models have been proposed to account for different phenomena. No SDT or Bayesian models provide categorically poor explanations of behaviour in this domain of short-display duration covert tasks. All of these models are based on specific, refutable, information processing mechanisms, and many studies compare multiple models, with parallel, 1-stage, Bayesian noise-limited explanations being favoured over serial, 2-stage, resource-limited non-Bayesian explanations. Bayesian approaches place emphasis upon the statistical structure of the environment, and thus are synergistic with the approach of adaptive rationality (Anderson, 1990) which allows us to ask *why* these effects occur, not just *what* mechanisms caused those effects. Attentional effects are not just due to the environment however, this review has emphasised the locus of these effects as both stimulus-based and internal belief-based. In all cases examined we can see these experimental effects as being a set of by-products of conducting Bayesian inference in an uncertain world. We need not invoke additional attentional causes or mechanisms to explain *these* covert effects. Given a generative model of the environment, our prior beliefs and our noise-corrupted sensory observations, we conduct the inferences demanded by the experimental tasks. Our internal causal models may or may not precisely match the structure of an actual experiment, and our subjective beliefs may not be entirely accurate. And so in some covert search situations we may be close to optimal, in others we may not be, but it appears that we are still Bayesian.

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.

Anderson, B. (2011). There is no Such Thing as Attention. *Frontiers in psychology*, 2:1–8.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Psychology Press.

Baldassi, S. and Verghese, P. (2002). Comparing integration rules in visual search. *Journal of Vision*, 2(8):559–570.

Bashinski, H. S. and Bacharach, V. R. (1980). Enhancement of Perceptual Sensitivity as the Result of Selectively Attending to Spatial Locations. *Perception & Psychophysics*, 28(3):241–248.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., Latham, P. E., and Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, 60(6):1142–1152.

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.

Bitzer, S., Park, H., Blankenburg, F., and Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers In Human Neuroscience*, 8:102.

Borji, A. and Itti, L. (2014). Optimal attentional modulation of a neural population. *Frontiers in Computational Neuroscience*.

Bowers, J. S. and Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3):389–414.

Brown, S. D. and Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3):153–178.

Bruce, N. and Tsotsos, J. K. (2011). Visual representation determines search difficulty: explaining visual search asymmetries. *Frontiers in Computational Neuroscience*, 5:1–10.

Cameron, E. L., Tai, J. C., Eckstein, M., and Carrasco, M. (2004). Signal detection theory applied to three visual search tasks–identification, yes/no detection and localization. *Spatial Vision*, 17(4-5):295–325.

Carpenter, R. H. S. and Williams, M. L. L. (1995). Neural Computation of Log Likelihood in Control of Saccadic Eye-Movements. *Nature*, 377(6544):59–62.

Carrasco, M. (2011). Visual attention: the past 25 years. 51(13):1484–1525.

Carrasco, M. and Frieder, K. S. (1997). Cortical Magnification Neutralizes the Eccentricity Effect in Visual Search. 37(1):63–82.

Carrasco, M., McLean, T., Katz, S., and Frieder, K. S. (1998). Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation and spatial frequency.

Dosher, B. A. and Lu, Z.-L. (2000). Mechanisms of perceptual attention in precuing of location. 40(10-12):1269–1292.

Dosher, B. A., Lu, Z.-L., and Han, S. (2004). Parallel Processing in Visual Search Asymmetry. *Journal of Experimental psychology: Human Perception and Performance*, 30(1):3–27.

Downing, C. J. (1988). Expectancy and visual-spatial attention: effects on perceptual quality. *Journal of Experimental psychology: Human Perception and Performance*, 14(2):188–202.

Droll, J. A., Abbey, C. K., and Eckstein, M. (2009). Learning cue validity through performance feedback. *Journal of Vision*, 9(2):18.1–22.

Druker, M. and Anderson, B. (2010). Spatial probability aids visual stimulus discrimination. *Frontiers In Human Neuroscience*, 4:1–10.

Duncan, J. and Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3):433–458.

Eckstein, M. (1998). The Lower Visual Search Efficiency for Conjunctions Is Due to Noise and not Serial Attentional Processing. *Psychological science*, 9(2):111–118.

Eckstein, M. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5):14–14.

Eckstein, M., Mack, S. C., Liston, D., Bogush, L., Menzel, R., and Krauzlis, R. J. (2013). Rethinking human visual attention: spatial cueing effects and optimality of decisions by honeybees, monkeys and humans. 85:5–19.

Eckstein, M., Peterson, M. F., Pham, B. T., and Droll, J. A. (2009). Statistical decision theory to relate neurons to behavior in the study of covert visual attention. 49(10):1097–1128.

Eckstein, M., Pham, B. T., and Shimozaki, S. S. (2004). The footprints of visual attention during search with 100% valid and 100% invalid cues. 44(12):1193–1207.

Eckstein, M., Shimozaki, S. S., and Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, 2(1):25–45.

Eckstein, M., Thomas, J. P., Palmer, J., and Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, 62(3):425–451.

Farrell, S. and Lewandowski, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science*, 19(5):329–335.

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.

Fernandez-Duque, D. and Johnson, M. L. (2002). Cause and effect theories of attention: The role of conceptual metaphors. *Review of General Psychology*, 6(2):153–165.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. 51(7):771–781.

Gibson, J. J. (1972). A theory of direct visual perception. In Royce, J. and Rozeboom, W., editors, *The Psychology of Knowing*, pages 77–89. MIT Press, New York.

Gould, I. C., Wolfgang, B. J., and Smith, P. L. (2007). Spatial uncertainty explains exogenous and endogenous attentional cuing effects in visual signal detection. *Journal of Vision*, 7(13):1–17.

Green, C. S., Benson, C., Kersten, D., and Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proceedings of the National Academy of Sciences*.

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.

Healy, A. and Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental*

Psychology: Human Learning and Memory, 7(5):344.

Helmholtz (1856). *Treatise on Physiological Optics*. The Optical Society of America.

James, W. (1890). *The principles of psychology*. Dover, New York.

Johnston, W. A. and Dark, V. J. (1986). Selective attention. *Annual Review of Psychology*, 37:43–75.

Jones, M., Mozer, M. C., Curran, T., and Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, 120(3):628–666.

Jordan, M. I. (2004). Graphical Models. *Statistical Science*, 19(1):140–155.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292.

Kinchla, R. A. (1977). The role of structural redundancy in the perception of visual targets. *Attention, Perception & Psychophysics*, 22(1):19–30.

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology*, 43:711–742.

Kinchla, R. A., Chen, Z. Z., and Evert, D. D. (1995). Precue effects in visual search: data or resource limited? *Perception & Psychophysics*, 57(4):441–450.

Krauzlis, R. J., Bollimunta, A., Arcizet, F., and Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, 18(9):457–464.

Kubovy, M. and Healy, A. (1977). The decision rule in probabilistic categorization: What it is and how it is learned. 106(4):427.

Lee, M. D. and Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.

Li, B., Peterson, M. R., and Freeman, R. D. (2003). Oblique effect: a neural basis in the visual cortex. *Journal of Neurophysiology*, 90(1):204–217.

Liston, D. and Stone, L. S. (2008). Effects of prior information and reward on oculomotor and perceptual choices. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(51):13866–13875.

Lu, Z.-L. and Dosher, B. A. (1998). External noise distinguishes attention mechanisms. 38(9):1183–1198.

Lu, Z.-L. and Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 16(3):764–778.

Lu, Z.-L. and Dosher, B. A. (2014). *Visual Psychophysics: From Laboratory to Theory*. MIT Press, Cambridge, Mass.

Lu, Z.-L., Dosher, B. A., and Han, S. (2010). Information-limited parallel processing in difficult heterogeneous covert visual search. *Journal of Experimental psychology: Human Perception and Performance*, 36(5):1128–1144.

Ludwig, C. J. H. (2012). Saccadic decision-making. In Liversedge, S. P. and Everling, S., editors, *The Oxford Handbook of Eye Movements*, pages 425–437. OUP, Oxford.

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences*, 16(10):511–518.

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–1438.

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.

Maddox, W. T. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. *Journal of the experimental analysis of behavior*, 78(3):567–595.

Maloney, L. and Zhang, H. (2010). Decision-theoretic models of visual perception and action. 50(23):2362–2374.

Marr, D. (1982). *Vision*. A Computational Investigation Into the Human Representation and Processing of Visual Information. MIT Press, Cambridge, Mass.

Martins, A. C. R. (2006). Probability biases as Bayesian inference. *Judgment and Decision Making*, 1(2):108–117.

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.

McElree, B. and Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental psychology: Human Perception and Performance*, 25(6):1517–1539.

Mitroff, S. R. and Biggs, A. T. (2013). The Ultra-Rare-Item Effect: Visual Search for Exceedingly Rare Items Is Highly Susceptible to Error. *Psychological science*.

Morvan, C. and Maloney, L. (2012). Human visual search does not maximize the post-saccadic probability of identifying targets. *PLoS Computational Biology*, 8(2):e1002342.

Mozer, M. C., Kinoshita, S., and Shettel, M. (2007). Sequential dependencies in human behavior offer insights into cognitive control. In Gray, W. D., editor, *Integrated Models of Cognitive Systems*. Oxford University Press, Oxford.

Müller, H. J. and Findlay, J. M. (1987). Sensitivity and criterion effects in the spatial cuing of visual attention. *Perception & Psychophysics*, 42(4):383–399.

Müller, H. J. and Humphreys, G. W. (1991). Luminance-increment detection: Capacity-limited or not? *Journal of Experimental psychology: Human Perception and Perfor-*

mance, 17(1):107–124.

Najemnik, J. and Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–391.

Najemnik, J. and Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3):4–14.

Navalpakkam, V., Koch, C., and Perona, P. (2009). Homo economicus in visual search. *Journal of Vision*, 9(1):1–16.

Nolte, L. W. and Jaarsma, D. (1967). More on the detection of one of M orthogonal signals. *Journal of the Optical Society of America A*, 41(2):497–505.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. 34:1703–1721.

Palmer, J., Ames, C. T., and Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental psychology: Human Perception and Performance*, 19(1):108–130.

Palmer, J., Verghese, P., Pavel, M. M., and Pavel, M. (2000). The psychophysics of visual search. 40(10-12):1227–1268.

Pizlo, Z. (2001). Perception viewed as an inverse problem. 41(24):3145–3161.

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–132.

Rao, R. P. N. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport*, 16(16):1843–1848.

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922.

Schoonveld, W. A., Shimozaki, S. S., and Eckstein, M. (2007). Optimal observer model of single-fixation oddity search predicts a shallow set-size function. *Journal of Vision*, 7(10):1.1–16.

Schüür, F., Tam, B., and Maloney, L. (2013). Learning Patterns in Noise: Environmental Statistics Explain the Sequential Effect. In *CogSci*.

Shaw, M. L. and Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance*, 3(2):201–211.

Shimozaki, S. S., Eckstein, M., and Abbey, C. K. (2003). Comparison of two weighted integration models for the cueing task: linear and likelihood. *Journal of Vision*, 3(3):209–229.

Shimozaki, S. S., Schoonveld, W. A., and Eckstein, M. (2012). A unified bayesian observer analysis for set size and cueing effects on perceptual decisions and saccades. *Journal of Vision*, 12(6):1–26.

Smith, P. L. (2000). Attention and luminance detection: effects of cues, masks, and pedestals. *Journal of Experimental psychology: Human Perception and Performance*, 26(4):1401–1420.

Smith, P. L. and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3):161–168.

Smith, P. L. and Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, 116(2):283–317.

Solomon, J. A. (2004). The effect of spatial cues on visual sensitivity. 44(12):1209–1216.

Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision research*, 48(12):1391–1408.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, 31:523–535.

Verghese, P., Renninger, L., and Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*.

Vincent, B. T. (2011a). Covert visual search: prior beliefs are optimally combined with sensory evidence. *Journal of Vision*, 11(13):25.

Vincent, B. T. (2011b). Search asymmetries: parallel processing of uncertain sensory information. 51(15):1741–1750.

Vincent, B. T. (2012). How do we use the past to predict the future in oculomotor search? 74:93–101.

Vincent, B. T., Baddeley, R. J., Troscianko, T., and Gilchrist, I. D. (2009). Optimal feature integration in visual search. *Journal of Vision*, 9(5):15–15.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1):67–85.

Wickens, T. (2002). Elementary signal detection theory. Oxford University Press, Oxford.

Wilder, M. H., Mozer, M. C., and Jones, M. (2009). Sequential effects reflect parallel learning of multiple environmental regularities. *Advances in neural ...*, pages 2053–2061.

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In Gray, W., editor, *Integrated Models of Cognitive Systems*, pages 99–119. Integrated models of cognitive systems.

Wolfe, J. M. and Cave, K. (1989). Guided search: An alternative to the feature integration model for visual search. *J Exp Psychol*, 15:419–433.

Wolfe, J. M. and Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, 435(7041):439–440.

Wood, C. C. and Jennings, J. R. (1976). Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. *Perception & Psychophysics*, 19(1):92–102.

Yu, A. J. and Cohen, J. D. (2008). Sequential effects: superstition or rational behavior? In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 21*. Advances in Neural Information Processing Systems 21.

Zelinsky, G. J. and Sheinberg, D. L. (1997). Eye movements during parallel–serial visual search. *Journal of Experimental psychology: Human Perception and Performance*, 23(1):244–262.

Zemel, R. S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, 10(2):403–430.

Zhang, H., Morvan, C., and Maloney, L. (2010). Gambling in the visual periphery: a conjoint-measurement analysis of human ability to judge visual uncertainty. *PLoS Computational Biology*, 6(12):e1001023.

Zhang, S. and Eckstein, M. (2010). Evolution and Optimality of Similar Neural Mechanisms for Perception and Action during Search. *PLoS Computational Biology*, 6(9):e1000930.