Atten Percept Psychophys (2015) 77:2021–2036
DOI 10.3758/s13414-015-0897-2

Low attention impairs optimal incorporation of prior knowledge in perceptual decisions

Jorge Morales[1] · Guillermo Solovey[2,3,4,5] · Brian Maniscalco[2,6] · Dobromir Rahnev[2,7] · Floris P. de Lange[8] · Hakwan Lau[2,8,9]

Published online: 3 April 2015
© The Psychonomic Society, Inc. 2015

Abstract When visual attention is directed away from a stimulus, neural processing is weak and strength and precision of sensory data decreases. From a computational perspective, in such situations observers should give more weight to prior expectations in order to behave optimally during a discrimination task. Here we test a signal detection theoretic model that counter-intuitively predicts subjects will do just the opposite in a discrimination task with two stimuli, one attended and one unattended: when subjects are probed to discriminate the unattended stimulus, they rely less on prior information about the probed stimulus' identity. The model is in part inspired by recent findings that attention reduces trial-by-trial variability of the neuronal population response and that they use a common criterion for attended and unattended trials. In five different visual discrimination experiments, when attention was directed away from the target stimulus, subjects did not adjust their response bias in reaction to a change in stimulus presentation frequency despite being fully informed and despite the presence of performance feedback and monetary and social incentives. This indicates that subjects did not rely more on the priors under conditions of inattention as would be predicted by a Bayes-optimal observer model. These results inform and constrain future models of Bayesian inference in the human brain.

Keywords Attention: divided attention and inattention · Signal detection theory · Cognitive and attentional control · Ideal observer Bayesian models

Introduction

According to ideal observer Bayesian principles, when perceptual information or perceptual processing are weak, optimal observers should rely more on prior expectations in order to behave optimally (Feldman & Friston, 2010; Friston, 2005; Knill & Pouget, 2004; Kouider, de Gardelle, Sackur, &

Jorge Morales and Guillermo Solovey contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.3758/s13414-015-0897-2) contains supplementary material, which is available to authorized users.

✉ Jorge Morales
jorgemlg@gmail.com

[1] Department of Philosophy, Columbia University, 708 Philosophy Hall, 1150 Amsterdam Avenue, Mail Code: 4971, New York, NY 10027, USA

[2] Department of Psychology, Columbia University, New York, NY 10027, USA

[3] Instituto de Cálculo, FCEyN, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

[4] Laboratorio de Neurociencia Integrativa, Buenos Aires, Argentina

[5] CONICET, Buenos Aires, Argentina

[6] National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

[7] The Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA 94720, USA

[8] Donders Institute of Brain, Cognition and Behavior, Radboud University Nijmegen, 6500 HB Nijmgen, The Netherlands

[9] Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095, USA

Dupoux, 2010; Summerfield & de Lange, 2014; Summerfield & Egner, 2009; Summerfield & Koechlin, 2008). For instance, in a discrimination task, Bayes-optimal subjects capitalize on prior knowledge about the stimulus presentation probability to maximize the proportion of correct responses (Maddox & Bohil, 1998; Wyart, Nobre, & Summerfield, 2012). Such a strategy should be manifested in terms of a shift in subjects' response criteria, e.g., as strength or precision of sensory data decreases, such as for unattended stimuli (Carrasco, 2011), when stimulus presentation probability is biased the optimal discrimination criterion should become more liberal.

Here we test a signal detection theoretic model (Rahnev et al., 2011) that makes empirical predictions that may seem implausible from this commonsensical Bayesian perspective. According to our Common Criterion Model, in addition to increasing the strength of the perceptual signal (possibly by increasing activity in visual cortical areas that represent the attended location and suppressing cortical activity that represents unattended locations (Desimone & Duncan, 1995)), attention also reduces the trial-by-trial variability of the internal perceptual response (Bressler & Silver, 2010; Pestilli, Carrasco, Heeger, & Gardner, 2011; Wyart et al., 2012). Importantly, motivated by previous empirical evidence within a signal detection theoretic framework (Gorea & Sagi, 2000; Zak, Katkov, Gorea, & Sagi, 2012), the model also presupposes that subjects use a single common decision criterion for discriminating both attended and unattended stimuli. This model has gathered empirical support from studies using transcranial magnetic stimulation and neuroimaging (Rahnev, Bahdo, de Lange, & Lau, 2012a; Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2012b; Rounis, Maniscalco, Rothwell, Passingham, & Lau, 2010). The model makes the counterintuitive prediction that as long as there is uncertainty, when faced with a biased prior expectation of stimulus likelihood, human subjects will not be able to adjust their response bias in unattended trials as effectively as they will for attended trials.

We pit this model against the predictions of an Optimal Observer Model in a series of experiments in which subjects discriminated between left- and right-tilted gratings that were presented in cued (attended) or uncued (unattended or less attended) locations, one of which was probed in each trial. To anticipate, we confirmed the a priori prediction of our detection theoretic model repeatedly in five different experiments (plus one control experiment).

Materials and methods

Participants

Participants signed an informed-consent statement approved by the local institutional review board. Subjects had normal or corrected-to-normal vision and were naïve to the purpose of the experiments. They received a base compensation of US$10 per session plus a US$5 bonus if they had a higher percentage correct than the previous participant (whose performance was disclosed only at the end of the experiment). The first subject of each experiment competed against a mock subject who was given an arbitrary percentage of 80 % correct responses. This reward structure provided subjects with both monetary and social incentives to perform well in the task. A total of 79 subjects took part in this study. The mean age of participants was 23 years (range = 18–41 years).

Stimuli

Stimuli were generated using Psychophysics Toolbox (Brainard, 1997) in MATLAB (MathWorks, Natick, MA, USA) and were shown on an iMac monitor (24-in monitor size, 1920 × 1200 pixels, 60 Hz refresh rate). Stimuli were presented inside circular locations (5° diameter) on a gray background (111 cd/m^2), with circles' centers 5° away from fixation. Stimuli consisted of a noisy background composed of uniformly distributed intensity values with gratings on top (two cycles/degree) that could be tilted 45° (right) or 135° (left). Total contrast of stimuli was fixed at 30 % but the grating-to-noise ratio varied, thus making the gratings more or less visible.

General task design

Here we present the general design of our experiments. For details and variations, see the Experiment and task design section of each experiment.

Seated in a dimmed room (60 cm from the monitor), subjects were instructed to fixate at a central red square (0.4°) for 500 ms at the beginning of each trial. Subjects were instructed to attend to one of two possible locations pointed by a white (200 cd/m^2) pre-cue double-headed arrow presented for 500 ms, which randomly changed orientation every 40 trials. This blocking of attentional cueing reduced the cognitive demands on subjects and made it less likely that they could confuse the identity of cued stimulus. Stimuli were presented for 50 ms inside four circular locations (5° diameter) on a gray background (111 cd/m^2) on the four quadrants of the screen, with the center of each circle 5° away from fixation. This configuration was chosen in order to minimize eye movements and enhance central fixation, as in previous studies (Bahrami, Lavie, & Rees, 2007; Rahnev et al., 2011).

The tilt of the gratings was always the same in the two locations of each diagonal, thus the gratings in one diagonal can be considered as a single compound stimulus. The tilts of the gratings on different diagonals were always independent. After stimulus presentation, a response cue consisting of two white circles at the same location of the stimuli of one of the

two diagonals indicated the target stimulus for which subjects had to provide a perceptual judgment regarding the tilt (left or right) of the gratings. The response cue coincided with the pre-cue on 70 % of the trials (cued) and it selected the stimulus that was not pre-cued in the remaining 30 % of trials (uncued). After providing a perceptual judgment, subjects rated confidence on a 4-point scale (1 = not confident at all; 2 = barely confident; 3 = somewhat confident; 4 = highly confident). Perceptual judgments were provided with two fingers from the left hand on a normal keyboard, pressing one key for right tilt and a different key for left tilt. Confidence ratings were provided with four different keys with a different finger of the right hand for each key. Subjects received auditory feedback (a brief low-pitch beep) after an incorrect answer at the end of the trial and their overall performance (percentage correct) was shown to them every 40 trials. We note that the aim of collecting confidence ratings was to assess subjects' metacognitive capacity, but since they are irrelevant for the scope of this report regarding integration of prior information in perceptual decisions, we will not analyze them.

A session took approximately 1 hour to complete and consisted of 240 practice and 800 experimental trials. Two conditions were tested: in the unbiased condition, left and right tilts on the target stimuli were equally likely; in the biased condition, 70 % of the target stimuli were tilted to one side and 30 % to the other (i.e., 70 % of cued and 70 % of uncued trials were biased to one side and the remaining 30 % of each to the other side). The biased tilt was constant throughout the biased blocks in each experiment. Subjects were explicitly informed of the attentional and bias conditions, were reminded of them before every block, and had the chance to become familiar with the structure of the experiment during practice trials. The experimental trials were split into four blocks (two interleaved blocks for the unbiased and biased conditions, respectively; presentation order was counterbalanced across subjects). Participants had 15-second forced pauses every 40 trials, after which they were informed of their accumulated percentage correct. Subjects had self-paced pauses between blocks. Before the experiment, subjects completed 240 practice trials that followed the same structure of the main experiment except that they were split into six blocks: three per each type of bias. In the first four practice blocks stimuli were presented at higher grating-to-noise contrast ratio than in the actual experiment and in the last two blocks they had the same contrast ratio as in the actual experiment. Practice results were not analyzed.

Rationale for theoretical predictions

In our experiments, observers had to classify a binary stimulus as being S_1 or S_2. According to Signal Detection Theory (SDT) (Green & Swets, 1966; Macmillan & Creelman, 2005), when the stimulus is S_1, an internal perceptual response x is drawn in the brain of the observer with probability distribution $P(x/S_1)$. Similarly, x follows $P(x/S_2)$ when the stimulus is S_2. We assume that the distributions corresponding to S_1 and S_2 stimuli are Gaussian with equal variance (see Fig. 1a for an illustration). On every trial, x is compared with a decision criterion x_c such that if $x > x_c$, the observer's response is S_2; otherwise, the response is S_1. For simplicity, in this section we consider that $P(x/S_2)$ and $P(x/S_1)$ have means = $\pm\mu/2$ and variance = σ^2.

Bayes-optimal observer

A Bayes-optimal observer infers the most probable stimulus on every trial, given x, using Bayes rule. Conceptually, it calculates the decision variable $d(x) = \log\frac{P(S_2|x)}{P(S_1|x)}$ and responds S_1 if $d(x) < 0$ and S_2 if $d(x)>0$. Using Bayes rule, decision variable d becomes

$$d(x) = \log\frac{P(x|S_2)}{P(x|S_1)} + \log\frac{P(S_2)}{P(S_1)} \tag{1}$$

The decision rule can be expressed in terms of x. Replacing the probabilities $P(x|S_1)$ and $P(x|S_2)$ by the corresponding formulas and solving $d(x) = 0$, the Bayes-optimal criterion is:

$$x_c^{opt} = \frac{\sigma^2}{\mu}\log\frac{P(S_1)}{P(S_2)} \tag{2}$$

where $P(S_1)$ and $P(S_2)$ are a priori probabilities that the stimulus is S_1 and S_2, respectively. Observers using the decision criterion given by Eq. 2 effectively maximize the correct response probability, assuming μ and σ are constant. We assume they are constant because, unlike the observer's *subjective* decision criterion, they are determined by observers' perceptual systems (σ is not under subjects' control) and the task's structure (in our experiments, μ is determined by the stimulus' contrast) and thus can be considered *objective* variables.

Figure 1a shows $P(x|S_1)$, $P(x|S_2)$ and the location of the optimal criterion given by Eq. 2 (dashed vertical lines) assuming that S_1 is more probable than S_2 ($P(S_1):P(S_2) = 70:30$). In each panel, the distributions have different σ but the same μ, therefore sensitivity ($d' = \mu/\sigma$) is different. Note that optimal observers shift the decision criterion to account for biased prior probabilities. Also, the criterion is larger when sensitivity decreases. This makes sense, as an optimal observer should capitalize on prior knowledge about what stimulus is more probable, especially when sensory information is weak.

Signal Detection Theory offers different measures to quantify response bias. The most common is c, defined as the position of the criterion x_c relative to σ, i.e., $c = x_c/\sigma$. Therefore, using Eq. 2, the optimal c is:

$$c_{opt} = \frac{1}{d'}\log\frac{P(S_1)}{P(S_2)} \tag{3}$$

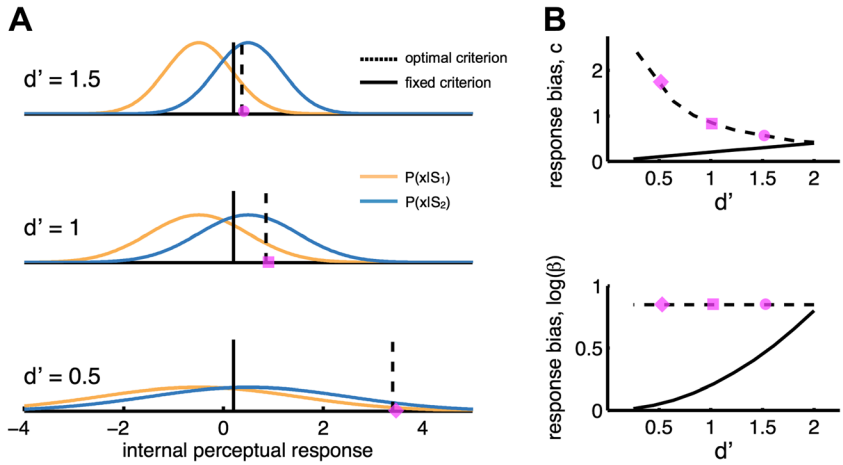

Fig. 1 Optimal (and suboptimal) decisions in a perceptual discrimination task using Signal Detection Theory (SDT). The figure illustrates the divergent behavior of an optimal observer and an observer with a fixed criterion, especially when sensitivity is low. In a discrimination task with binary stimuli (S_1 and S_2) where S_1 is more frequent than S_2 at a 70:30 ratio, SDT assumes that when a stimulus is presented to the observer, an internal perceptual response is generated in the brain of the observer with a Gaussian distribution: $P(x|S_1)$ when the stimulus is S_1 and $P(x|S_2)$ when the stimulus is S_2. If x is larger than a decision criterion, the observer's response is S_2; S_1 otherwise. **a** Each plot shows $P(x|S_1)$ (*orange*) and $P(x|S_2)$ (*blue*) for three different sensitivities (d'). Dashed vertical lines indicate the location of the optimal criterion (Eq. 1). Solid vertical lines indicate a fixed criterion. **b** Response bias measures c (optimal value given by Eq. 3; otherwise calculated using Eq. 6) and $\log\beta$ (optimal value given by Eq. 4; otherwise calculated using Eq. 7) as a function of sensitivity for hypothetical subjects using an optimal criterion (*dashed lines*) and those using a fixed criterion (*solid lines*). [All figures use colorblind safe colors following Brewer (2013)]

A second measure of bias is β (or $\log\beta$ for convenience), defined as the ratio between the probabilities $P(x|S_2)$ and $P(x|S_1)$ at the location of the criterion ($x = x_c$). The advantage of β is that the optimal β, as opposed to c, is independent of the observers' performance. In fact, at the location of the criterion ($d = 0$), Eq. 1 implies that:

$$\log\beta_{opt} = \log\frac{P(S_1)}{P(S_2)} \tag{4}$$

Figure 1b illustrates the relationship between optimal response bias and sensitivity. Whereas c_{opt} increases when sensitivity d' decreases, $\log\beta_{opt}$ is constant. The location of the optimal decision criterion when stimulus presentation prior probabilities of S_1 and S_2 are not equal, which is the case in our experiments, depends on sensitivity d' (as illustrated in Fig. 1a) and hence it varies across subjects and across attentional conditions. However, optimal $\log\beta$ is constant for all subjects regardless of sensitivity: it only depends on the stimuli frequencies. In our study, then, $\log\beta$ is a preferable measure for subjects' response bias because we can easily compare it to an optimal $\log\beta$ that is independent of d' and constant across subjects when we bias the stimulus presentation frequencies.

Fixed-criterion observer

Subjects setting a criterion different to the optimal given by Eq. 2 will perform worse than they potentially could. One suboptimal strategy worth considering is the one that our Common Criterion Model assumes: subjects use the same common decision criterion (x_c^{fixed}) for discriminating both attended and unattended stimuli, despite differences in sensitivity. An intuitive understanding of the model's predictions can be gained considering the fixed decision criterion in Fig. 1a (solid vertical lines). Using this suboptimal strategy, c and $\log\beta$ both increase when d' increases (Fig. 1b, solid line). More importantly, the worse the sensitivity is (weak visual information), the larger the divergence between a suboptimal and an ideal observer. In other words, there is a less efficient use of prior knowledge when visual evidence is weak. This suboptimal behavior is what our Common Criterion Model predicts will happen in unattended trials in our experiments (see Experiment 1 below for details on the model). Also note that with equal stimulus presentation probabilities (i.e., $P(S_1):P(S_2) = 50:50$), Eq. 4 implies that optimal $\log\beta = 0$, meaning that optimal observers place their criterion at the intersection of the S_1 and S_2 distributions ($x_c = 0$), regardless of their sensitivity d'.

Signal detection theoretic measures

We estimate sensitivity (d') with the standard signal detection theoretic formula:

$$d' = z(HR) - z(FAR) \tag{5}$$

where HR is the hit rate and FAR the false alarm rate, and the z transformation converts these rates to a z score (i.e., to standard deviation units). Response bias measures can also be expressed in terms of HR and FAR as:

$$c = -\frac{1}{2}[z(HR) + z(FAR)] \quad (6)$$

and

$$\log\beta = -\frac{1}{2}\left[z(HR)^2 - z(FAR)^2\right] \quad (7)$$

The latter is easily obtained because both measures of response bias are related through the expression $\log\beta = c \cdot d'$ (Green & Swets, 1966; Macmillan & Creelman, 2005).

Experiment 1: Smaller response bias shift in uncued trials compared to cued trials during biased stimulus presentation

Experiment and task design

Eighteen subjects (12 women) participated in this experiment. The grating contrast contribution to total contrast (i.e., 30 %) changed randomly across trials between low (5.7 %) and high (9 %) levels. Half of the cued and uncued trials in each bias condition had a low contrast-contributing grating and the other half had high contrast-contributing grating. Subjects completed one session (see Fig. 2 and Materials and methods section for details).

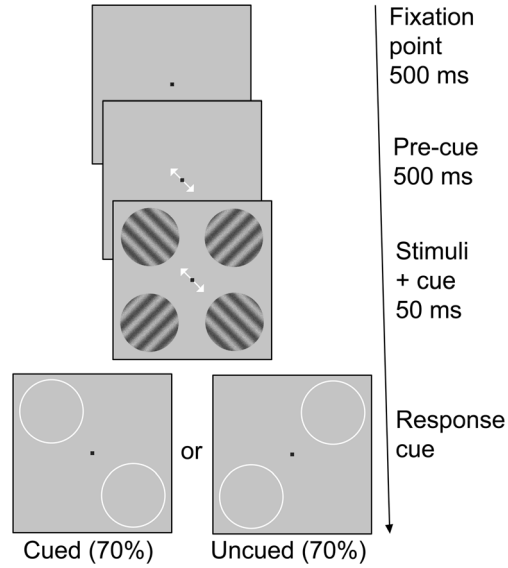

Fig. 2 Main task design. Each trial began with a fixation point followed by a pre-cue arrow indicating the diagonal to which subjects should attend. After stimulus presentation (same tilt for stimuli in the same diagonal), a response cue (*white circles*) indicated the diagonal for which subjects had to provide a perceptual judgment regarding the tilt (*left or right*) of the gratings. Subjects received auditory feedback after an incorrect answer. The response cue coincided with the pre-cue on 70 % of the trials (*cued*) and it was different in the remaining 30 % (*uncued*). In the unbiased condition (two 200-trial blocks), left and right tilts were equally likely. In the biased condition (two 200-trial blocks), 70 % of the response-cued stimuli were tilted to one side and 30 % to the other. This design was used for Experiments 1, 2, and 4 (see Materials and methods section for details)

Optimal observer and common criterion model predictions

In our experiments, on each trial, stimuli are presented at a cued and at an uncued location. Consequently, in the models, there are two internal response axes, one for the cued location and one for the uncued location.

Bayes-optimal observers are free to use two different criteria to make decisions in cued and uncued trials. In Fig. 3a we present the predictions of the Optimal Observer Model for Experiment 1 when the prior stimulus probabilities are biased. In this Experiment, the contrast was the same for cued and uncued locations, therefore in the models the difference between the mean of the distributions S_1 (i.e., left-tilted stimuli) and S_2 (i.e., right-tilted stimuli) is the same for cued and uncued distributions. Attention modulates the variance of the distributions. When observers pay attention, their objective performance is better. This is exemplified in the distributions plotted in Fig. 3a (top). In this example, at an attended location, $d' = 1.5$, and $d' = 0.5$ at an unattended location. The optimal location of criteria is given by Eq. 2. As shown in Fig. 3a (bottom), optimal observers shift their decision criterion in such a way that $\log\beta$ is the same for both decisions.

In contrast, the 'Common Criterion Model' assumes that observers use a common criterion for cued and uncued stimuli (Gorea & Sagi, 2000; Zak et al., 2012). As a consequence, it makes very different predictions (Fig. 3b). This assumption is informed by previous empirical psychophysical evidence (Gorea & Sagi, 2000). In this case, because d' is not the same for attended and unattended distributions, a common criterion implies that response bias cannot be optimal for both locations when stimuli presentation frequencies are biased. The prediction is that in attended trials, the response bias will be closer to optimal and, hence, further away from optimal in the unattended trials. This is the opposite of what Bayes-optimal observers would do: when vision is impaired (as during inattention), they shift response bias by a larger amount. In Fig. 3c we present the model predictions for the unbiased condition, which are the same for both the Optimal Observer and the Common Criterion Models. When stimulus presentation is unbiased, both models predict optimal response bias.

Results

As expected due to matched stimulation conditions, subjects showed a higher sensitivity (d') for discriminating the cued stimuli than for uncued stimuli (paired t tests: unbiased: T(17) = 4.37, P = 0.0004, CI = [0.62,1.78]; biased: T(17) =

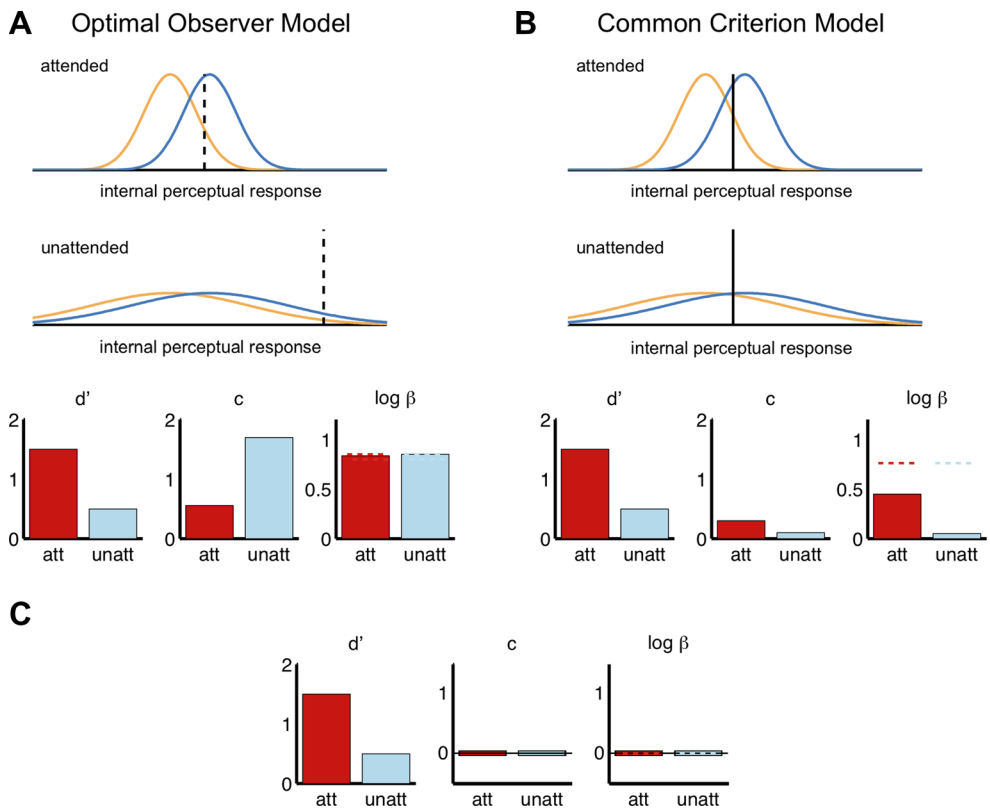

Fig. 3 Model predictions for Experiment 1. Biased Condition: **a** Optimal Observer Model. *Top*: Internal response distributions for stimuli presented in the attended and unattended locations of the visual field (orange: $P(x|S_1)$; blue: $P(x|S_2)$). The unattended internal response has a higher variance. In Experiment 1 the contrast at the attended location was the same as in the unattended location. To account for this, the distance between the mean values of the responses is unchanged with attention. The same means with a higher variance implies lower sensitivity (see Materials and methods). In the biased condition, the stimulus presentation probability was such that $P(S_1)$:$P(S_2)$ = 70:30. The optimal localization of the criterion (Eq. 1), i.e., the one that leads to maximum accuracy, is shown as vertical dashed lines. *Bottom*: Sensitivity (d') is worse in the unattended location. Response bias reflects the fact that an optimal subject uses prior information about the stimulus proportions and adjusts criterion accordingly: c is larger for unattended trials (Eq. 6) but $\log\beta$ is the same (Eq. 7). **b** Common Criterion Model. *Top*: Distributions are the same as in (**a**). However, this model assumes that subjects use the same criterion for both attended and unattended trials. *Bottom*: Sensitivity does not depend on criterion, therefore it is the same as in (**a**). Response bias (measured by c) is positive but in the unattended condition it is closer to zero (as opposed to the optimal observers predictions). In this model, response bias (measured by $\log\beta$) cannot be jointly optimal for both attentional conditions. Criterion shift induces a bias that is closer to optimal in the attended condition (for comparison, optimal $\log\beta$ are shown in dashed lines). Parameters of the shown distributions are arbitrary. **c** Unbiased Condition: Since sensitivity does not depend on response bias, the prediction for d' is the same as in the biased condition. When the stimulus presentation probability is unbiased (such that $P(S_1)$:$P(S_2)$ = 50:50), both the Optimal Observer and the Common Criterion Models make the same prediction regarding subjects' response bias: as measured by c and by $\log\beta$ subjects are optimal

4.05, P = 0.0008, CI = [0.54,1.70]; Fig. 4a). For hit and false alarm rates, see Fig. S1 in Supplemental Information.

Response bias measured by c, as expected by our model's prediction (Fig. 3b, c), was not significantly different between attentional conditions in either bias condition (paired t tests between cued and uncued conditions, unbiased: T(17) = 0.93, P = 0.36, CI = [−0.27,0.10]; biased: T(17) = 0.08, P = 0.93, CI = [−0.28,0.30]; Fig. 4b).

In the unbiased condition, $\log\beta$ was not different from optimality for cued and uncued trials, i.e., optimal $\log\beta$ = 0 for a left-tilted and right-tilted stimulus presentation ratio of 50:50 (see Eq. 4 in Materials and methods section; t tests against value zero, cued: T(17) = 0.55; P = 0.59, CI = [−0.17,0.29]; uncued: T(17) =−0.03, P = 0.97, CI = [−0.14, 0.14]). When the stimulus probability changed from unbiased (i.e., the ratio of left-tilted to right-tilted stimuli being 50:50) to biased (in all our experiments the same ratio now being 70:30), the shift in subjects' response bias as measured by $\log\beta$ was farther from optimality in unattended trials than in attended trials (Fig. 4c). Critically, there was a significant interaction between attention (cued vs. uncued) and bias (biased vs. unbiased) (2 × 2 repeated measures ANOVA, F(1,17) = 7.26, P = 0.01)

Discussion

The shift in response bias measured by $\log\beta$ due to biased stimulus presentation was not numerically optimal on average in attended trials (as it is not unexpected based on previous

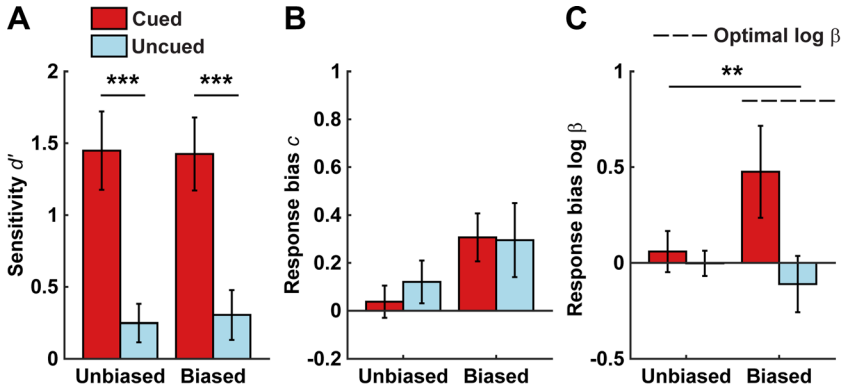

Fig. 4 Results for Experiment 1. **a** Detection sensitivity (d') differed between cued and uncued stimuli for both the unbiased and biased conditions. **b** Response bias measured by c. **c** In the unbiased condition, $\log\beta$ (log-likelihood ratio evaluated at the criterion) was near optimality (0) for cued and uncued trials. However, in the biased condition, $\log\beta$ shifted towards optimality only for cued but not for uncued trials. Solid bars indicate means and error bars SEM. $^{**}P \leq 0.01$ $^{***}P < 0.001$

literature (Gorea & Sagi, 2000; Green & Swets, 1966; Macmillan & Creelman, 2005; Rahnev et al., 2011)), but it was significant and in the right direction. On the other hand, the shift in uncued trials lagged behind the shift in cued trials, as if subjects were unable to adjust their response bias during inattention.

Note that the shift in response bias as measured by c is similar in the biased cued and uncued conditions (Fig. 4b). This, however, does not imply that subjects were equally close to optimality, since optimal c (not shown) is different for cued and uncued because d' was significantly different and it varied for every subject (see Eq. 3). These results, then, support a model in which the variance of the internal signal distributions is modulated by attention and subjects use a common criterion such that they are closer to optimality in cued trials than in uncued trials.

Experiment 2: Multiple sessions and higher contrast

Worried that the results in Experiment 1 were obtained only due to subjects' lack of familiarity with the task or extreme difficulty during unattended trials, in Experiment 2 we tested each subject for at least two 1-hour-long sessions (three subjects had four sessions), giving them ample opportunity to learn to behave more optimally. We also made the task somewhat easier during the unattended trials by increasing the mean contrast of both stimuli.

Experiment and task design

Each session was run in the exact same fashion as in Experiment 1. Participants (13 subjects; seven women), however, had two sessions, one with low and one with high contrast-contributing gratings (except for three participants who had four sessions, two with low and two with high contrasts). Statistics are presented for the subjects' average across sessions. Different sessions took place on different days within 1 week. Contrast levels were higher overall than in Experiment 1 to allow an increase in sensitivity for uncued trials. In Experiment 2 contribution of the grating contrast to total contrast was 9 % (low) and 12.3 % (high), instead of the lower levels in Experiment 1 of 5.7 % (low) and 9 % (high).

Results

Detection sensitivity (d') in cued and uncued trials differed in the unbiased and biased conditions (paired t tests, T(12) = 5.83, $P = 0.00008$, CI = [1.1,2.4] and T(12) = 6.35, $P = 0.00004$, CI = [1.15,2.35], respectively; see Fig. 5a). Overall discrimination sensitivity was higher than in Experiment 1 (Experiment 1 mean cued = 1.44, uncued = 0.28; Experiment 2 mean cued = 2.46, uncued = 0.71), reflecting the benefit from the extra training and the fact that we raised the contribution of grating contrast from a mean of 7.4 % in Experiment 1 to a mean of 10.7 % in this experiment. See Fig. S2 in the Supplemental Information for hit and false alarm rates.

Response bias as measured by c was different between attentional conditions in both bias conditions (paired t tests between cued and uncued conditions, unbiased: T(12) = 4.50, $P = 0.001$, CI = [0.18,0.51]; biased: T(12) = 2.33, $P = 0.04$, CI = [0.01,0.43]; Fig. 5b). Importantly, in the unbiased condition, $\log\beta$ was not different from optimality (0) for cued and uncued trials (t tests against value zero, T(12) =−0.06, $P = 0.96$, CI = [−0.48,0.46] and T(12) = 1.79, $P = 0.10$, CI = [−0.05,0.47], respectively). In the biased condition, despite the extra training, the higher overall sensitivity, and response bias c being higher on average for the unattended condition, $\log\beta$ shifted towards optimality (dashed line) in cued trials much more than it did in uncued trials. Importantly, as in the previous experiment, there was a significant interaction between attention and bias (2 × 2 repeated measures ANOVA, F(1,12) = 8.71, $P = 0.01$).

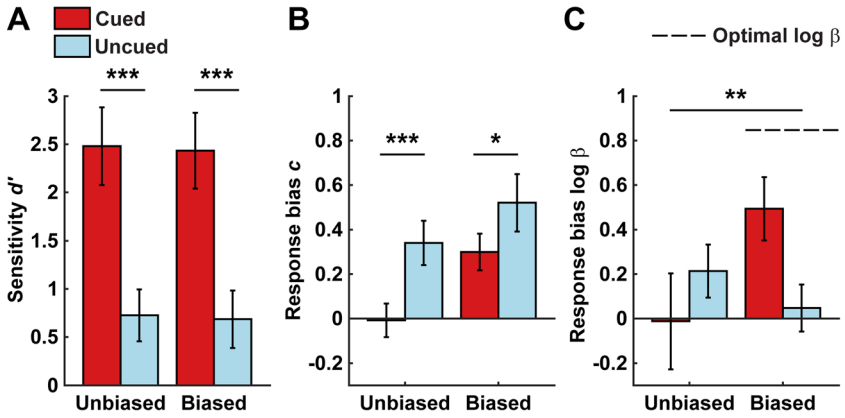

Fig. 5 Results for Experiment 2. **a** Similar task as in Experiment 1, with at least two sessions per subject and increased grating contrast. Overall sensitivity was increased and subjects showed higher sensitivity (d') to cued than to uncued stimuli. **b** Response bias measured by c. **c** In the unbiased condition, log β was near optimality for cued and uncued trials and not different from each other. In the biased condition, log β shifted towards optimality only in cued but not in the uncued trials. Solid bars indicate subjects' means and error bars SEM. $^*P < 0.05$ $^{**}P \leq 0.01$ $^{***}P \leq 0.001$

Discussion

Higher sensitivity and extended training did not change significantly the pattern of results we found in Experiment 1. The mean response bias as measured by c was higher in uncued trials (in both the unbiased and biased conditions). Note, however, that the shift was not enough for making log β approach optimality in the same proportion as in the cued trials. This indicates that, despite the larger average shift in c, they were further away from optimality in uncued trials. The Optimal Observer Model predicted that a higher response bias c during inattention should have led to an equally optimal log β for cued and uncued trials if subjects were being optimal. We did not see this in our data. Rather, as we explained above (Materials and methods; Figs. 1b and 3), response bias c, when distributions have different variances, is not particularly informative about how the two values of c compare to each other. Log β, on the other hand, is a much better measure for determining how close to optimality subjects' response bias is because optimal log β is the same regardless of variance or sensitivity. The results of this experiment, then, add evidence to the results from Experiment 1 to support our Common Criterion Model insofar as our model predicts that subjects would be able to shift their log β more in cued than in uncued trials.

Experiment 3: Low sensitivity and general inattention alone are not enough to prevent response bias shifting

The fact that discrimination sensitivity was relatively low for unattended stimuli in the first two experiments could be a source of concern. Subjects might have been frustrated and resolved to respond more or less randomly during unattended trials, i.e., in a 50/50 fashion. We note that sensitivity (d') for uncued trials was significantly positive in both Experiments 1 and 2, and in Experiment 2 it was within a range that is not atypical in psychophysics experiments (mean uncued $d' = 0.71$). Nevertheless, we performed Experiment 3 to test the hypothesis that it was the combination of attended and unattended trials in the same task rather than low d' or just a general lack of attention that prevented subjects from shifting their response bias adequately in reaction to changes in their prior expectation in the uncued trials. Recall, our model predicted that a common criterion for cued and uncued trials would prevent subjects' response bias from being equally close to optimality in both types of trial. If it were just inattention or just low sensitivity that drives the inability to shift response bias that we saw in Experiments 1 and 2, subjects should be unable to shift their criterion towards optimality in Experiment 3 as well.

Experiment and task design

There were ten participants in Experiment 3 (seven women). The experiment was carried out following the task design of Experiment 2, except for the following points. First, there was no attentional pre-cue/cue arrow before or during stimulus presentation, ensuring attention was equally divided between the stimuli in the two diagonals (Fig. 6). Circles randomly probed one of the two diagonals in half of the trials. Second, grating contribution to contrast was decreased to ensure lower sensitivity. Subjects had one session with low (3.7 %) and another with high contrast-contributing gratings (7.3 %). Two participants withdrew from the experiment after completing only one session (one subject had low contrast and the other had high contrast). We note that these two manipulations were meant to mimic the unattended conditions of the previous two experiments in terms of low sensitivity without having two different attentional conditions.

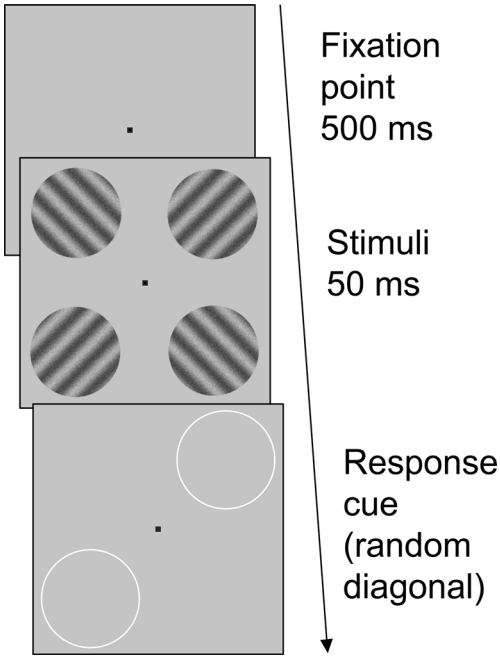

Fig. 6 Task design for Experiment 3. The task in Experiment 3 was similar to the task in Experiment 2 except that there was no pre-cue/cue arrow and contrast was lower. Subjects' attention was equally divided between both diagonals and response circles were equally likely to be on either diagonal. See Experiment and task design of Experiment 3 for details

Results

Detection sensitivity (d') was lower (mean = 0.63) than uncued d' in Experiment 2 (mean = 0.71). Sensitivity in Experiment 3 was higher than in uncued d' from Experiment 1 (mean = 0.28), but it was still less than half of cued d' (mean = 1.44). Sensitivity was not different between conditions (paired t test, T(9) = 1.1, P = 0.30, CI = [−0.06,0.16]; Fig. 7a). For hit and false alarm rates, see Fig. S3 in the Supplemental Information.

Response bias as measured by c was different between conditions (paired t test: T(9) = 6.9, P = 0.0001, CI = [0.38,0.76]; Fig. 7b). Like in the unbiased conditions from previous experiments, in this experiment subjects displayed an optimal response bias measured by $\log\beta$ during the unbiased condition (t test against value zero, T(9) = 0.47, P = 0.65, CI = [−0.11,0.07]; Fig. 7c). However, unlike the uncued conditions in Experiments 1 and 2, in this experiment subjects were able to shift their response bias significantly towards optimality (dashed line) in the biased condition (Fig. 7c). Importantly, response bias in the biased condition was significantly different from the unbiased condition (paired t test, T(9) = 3.65, P = 0.005, CI = [0.10,0.42]).

Discussion

The shift in $\log\beta$ was achieved even when discrimination sensitivity was at a low level, comparable to that in the uncued condition in Experiments 1 and 2. This finding is critical because a significant shift in $\log\beta$ despite such a low d' implies that subjects were free to incorporate the priors about stimulus identity into their response bias. This suggests that, consistent with our Common Criterion Model (Fig. 3b), it is a common criterion for an attended and an unattended sets of distributions, rather than low sensitivity overall, that led to suboptimal behavior.

Experiment 4: Subjects fail to shift their response bias during inattention even when performance is matched

Another concern about the results in the first two experiments is that the different response bias shift was driven by the different performance in cued and uncued trials. Furthermore, despite the results of Experiment 3, a remaining concern could be that subjects behave erratically during uncued trials when these are mixed with cued trials, like in the tasks in Experiments 1 and 2. To rule out these possibilities we conducted Experiments 4A and 4B, in which we tried to match sensitivity between the cued and uncued trials by presenting a higher grating contrast in the uncued location.

Experiment and task design

Experiments 4A and 4B were similar in structure and design to Experiment 1. Eleven participants in Experiment 4A (three women) and sixteen in Experiment 4B (seven women) had one session only. Like in previous experiments, the biased tilt was constant throughout the biased blocks and it was counterbalanced across subjects. The only difference was that contrast for cued and uncued trials was different. In Experiments 4A and 4B we aimed to match sensitivity (d') between cued and uncued trials using the QUEST threshold determination procedure (Watson & Pelli, 1983) to find for each subject the grating contribution to total contrast that would produce about 75 % correct responses for both the cued and uncued stimuli. In Experiment 4A we used 120 practice trials for QUEST; 70 % of these were cued and were used to determine the grating contrast contribution to total contrast for cued stimuli, while the remaining 30 % of trials (uncued) were used to determine the grating contrast contribution of uncued stimuli. Halfway through the experiment, i.e., after one block of each condition, we updated this ratio for the cued or uncued stimuli (or both) to force d' for both type of trials to be as similar as possible. We lowered or raised the grating-to-noise contrast ratio using the psychometric function obtained during QUEST in order to decrease d' if it was higher than 1 or raise it if it was lower than 1. Mean grating contrast contribution during the first half were: cued = 15.1 % ±9.2; uncued =

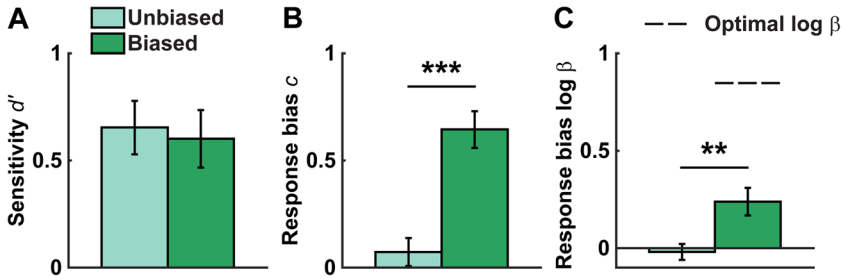

Fig. 7 Results for Experiment 3. **a** Detection sensitivity (d') was lower than uncued d' in Experiment 2 and it was not different between conditions. **b** Response bias measured by c. **c** In the unbiased condition, log β was not significantly different from optimality (*dashed lined*), but in the biased condition, despite the relatively low d', log β shifted towards optimality (*dashed line*) significantly like in the cued conditions of the previous experiments. The two log β values were different from each other. Solid bars indicate subjects' means and error bars indicate SEM. ** $P < 0.01$ ***$P < 0.001$

23.7 % ± 11.3, and during the second half: cued = 11.9 % ±8.1; uncued = 23.3 % ± 13.1.

In Experiment 4B, we used 80 practice trials for QUEST, 70 % of which were cued and were used to determine the grating contrast contribution for cued stimuli. The uncued grating contrast contribution was equal to $e^{(\log(x)+k)}$, where x is the grating contrast contribution obtained with QUEST for cued stimuli and the constant $k = 0.7$ was determined by obtaining the mean difference between the natural logarithm of the grating contrast contributions to total contrast for cued and uncued stimuli of participants in Experiment 4A. We did not readjust the ratios halfway through the experiment. Mean grating contrast contribution for cued stimuli were: cued = 35.9 % ± 28.3; uncued = 46.9 % ± 31.

Figure 8 illustrates the predictions of the models when d' is matched. Although the predictions are essentially the same as for Experiment 1 (Fig. 3), it is important to show that it is the difference in variance, and not sensitivity, that drives the main effect predicted by our Common Criterion Model.

Results

In Experiments 4A and 4B we observed a similar pattern of results. In the biased condition, log β shifted towards optimality only in the cued trials (paired t tests, Experiment 4A: T(10) = 2.63, $P = 0.02$, CI = [0.07,0.91]; Experiment 4B: T(15) = 2.85, $P = 0.01$, CI = [0.08,0.54]). Overall, sensitivity was not significantly different between the cued and uncued conditions in either Experiment 4A (paired t test, unbiased: $P = 0.22$; biased: $P = 0.11$) or 4B (paired t test, unbiased: $P = 0.39$; biased: $P = 0.41$). However, d' matching was not always perfect for every subject, most likely because subjects had more opportunity to learn how to improve their sensitivity in the attended trials, which were presented more frequently.

The motivation of these experiments was to test whether the same pattern of response bias shifts that we observed in the previous experiments would be preserved when sensitivity in cued and uncued trials was matched. For this reason, to achieve better sensitivity matching, we combined data from the best-matched subjects (total $n = 14$) of Experiment 4A ($n = 6$) and Experiment 4B ($n = 8$) to maintain similar power. The absolute difference between cued and uncued d' was taken as a measure of matching. We ranked subjects accordingly and half of the subjects of each experiment with the smallest differences in unbiased and biased conditions were included in the joint analysis presented in Fig. 9. Sensitivity (d') for this subset of the data was matched between cued and uncued trials (paired t tests, unbiased, T(13) = 0.61, $P = 0.55$, CI=[−0.40,0.72]; biased T(13) = 0.58, $P = 0.57$, CI=[−0.41, 0.72]; Fig. 9a). See Figure S4 in the Supplemental Information for hit and false alarm rates.

Even when we imposed these conservative constraints to the data analysis, ensuring that we only considered subjects whose d' was reasonably matched (unbiased condition mean difference between cued and uncued sensitivity = 0.16, biased condition mean difference = 0.15), we still found the same response patterns. Response bias c was not different between attentional conditions in either bias condition (paired t tests between cued and uncued conditions, unbiased: T(13) = 0.19, $P = 0.85$, CI = [−0.19,0.23]; biased: T(13) = 1.91, $P = 0.08$, CI = [−0.01,0.22]; Fig. 9b). In the unbiased condition, cued and uncued trials were near optimal (t test against value zero, T(13) = 0.56, $P = 0.59$, CI = [−0.19,0.32]; T(13) = 1.21, $P = 0.25$, CI = [−0.12,0.43], respectively), and subjects still struggled to shift their response bias significantly towards optimality only in uncued trials (Fig. 9c). Importantly, as in all the previous experiments, we found an interaction between attention and bias: 2×2 repeated measures ANOVA, F(1,13) = 4.5, $P = 0.05$.

Discussion

In these experiments we found that subjects did not fail to shift their log β due to low sensitivity or due to being frustrated during uncued trials, since their sensitivity was equally good as in cued trials and the constant feedback makes it

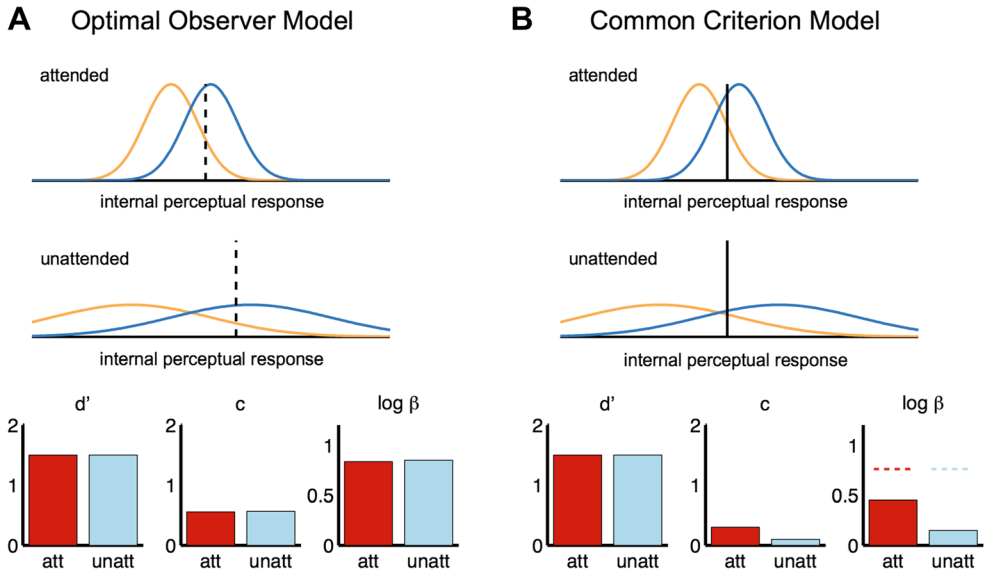

Fig. 8 Model predictions for Biased Condition in Experiments 4A and 4B. This is the same as Fig. 3 but in Experiments 4A and 4B the contrast of the unattended location was titrated to equate sensitivity (see Experiment and Task Design). This was accounted for in the models increasing the distance between the distributions $P(x|S_1)$ and $P(x|S_2)$ to compensate for the larger variance in the unattended condition. Parameters of the distributions are arbitrary

implausible to think they were not aware of their good performance. On the other hand, our model's assumptions (i.e., a common criterion for cued and uncued trials and the reduction of variance by attention), account for the observed results as predicted a priori (Fig. 8).

Experiment 5: The effect is preserved with a simpler display

While the diagonal arrangement of the stimuli has been successfully used before with the purpose of minimizing eye movements (Bahrami et al., 2007; Rahnev et al., 2011), having multiple items in the display could have made the task unnecessarily difficult. In order to dispel this worry, in Experiment 5 we used a simpler display.

Experiment and task design

In Experiment 5 we used a similar experimental setup to the one used in Experiment 1, except that participants (n = 11, three women) saw just two stimuli, instead of four, 5° away of fixation on opposite horizontal directions. A single-headed arrow pointed to just one of them, thus minimizing items on the screen as well as subjects' attentional effort (Fig. 10). Accordingly, the probe was just one circle that overlapped with the spatial location of one of the stimuli. Attentional and biasing format were just as in Experiment 1,

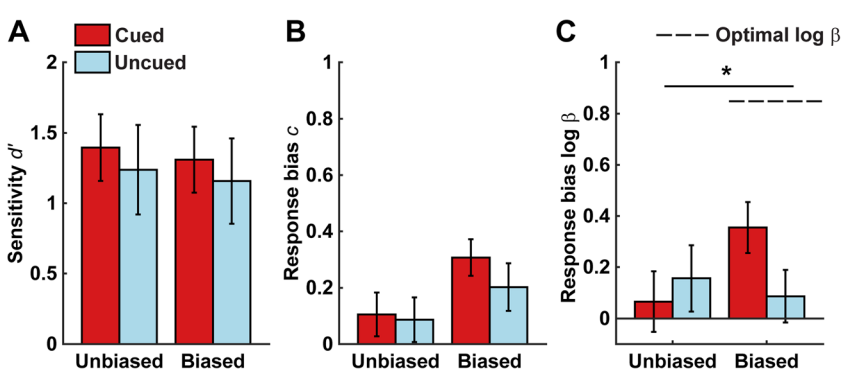

Fig. 9 Results for Experiments 4A and 4B combined. **a** Data from subjects (n = 14) with best-matched sensitivity were analyzed (see Experiment and Task Design for Experiment 4 for details). Sensitivity (d') was matched between cued and uncued trials. **b** Response bias measured by c. **c** In the unbiased condition, log β for cued and uncued trials was near optimality. In the biased condition, in spite of having similar sensitivity, log β shifted towards optimality in the cued and not in the uncued trials. Solid bars indicate means and error bars SEM. *$P \leq 0.05$

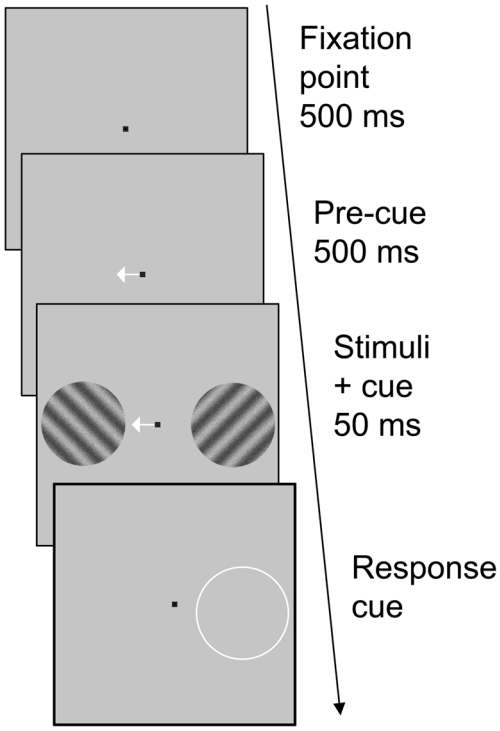

Fig. 10 Task design for Experiment 5. This is similar to the task in Experiment 1, but with two single stimuli on the horizontal of fixation. A single contrast was used throughout a unique session for all the stimuli. An uncued trial is depicted

counterbalancing left and right bias across subjects. The same grating contribution to total contrast (9 %; median value used in Experiments 1 and 2) was used throughout a single session for all stimuli.

Results

Overall discrimination sensitivity differed, as expected given the different attentional conditions, in both unbiased (paired *t* test, T(10) = 4.23, *P* = 0.002, CI = [0.99,3.2]) and biased conditions (T(10) = 4.7, *P* = 0.0008, CI = [1.2,3.4]; Fig. 11a). Again, we were able to find the response bias shifting patters from the previous experiments using this minimal display.

Response bias *c* was not significantly different between attentional conditions in either bias condition (paired *t* tests between cued and uncued conditions, unbiased: T(10) = 0.73, *P* = 0.48, CI = [−0.31,0.15]; biased: T(10) = 1.6, *P* = 0.14, CI = [−0.69,0.11]; Fig. 11b). In the unbiased condition, log β was optimal for cued and uncued trials as predicted by our model (*t* tests against value zero, T(10) = −0.29, *P* = 0.77, CI = [−0.60, 0.46] and T(10) = −0.67, *P* = 0.52, CI = [−0.40,0.21], respectively). In the biased condition, log β was optimal only in cued but not in the uncued trials and, importantly, we were able to find an interaction between attention and bias: 2×2 repeated measures ANOVA, F(1,10) = 10.59, *P* = 0.02; Fig. 11c. See Figure S5 in the Supplemental Information for hit and false alarm rates.

Further data analysis

A potential worry about all the experimental results is that stimulus adaptation could be a confound. In the biased condition, one stimulus orientation appears more frequently than the other. Adaptation reduces stimulus salience, but attention can overcome adaptation restoring contrast sensitivity (Pestilli, Viera, & Carrasco, 2007). Uncued trials, then, would have been more affected by adaptation and possibly prevented subjects from incorporating priors into their responses in a way not related to our model. A similar worry is that subjects could have tried to balance their responses in uncued trials, thereby avoiding the most frequent response in the most frequent and salient trials (i.e., cued trials) and subsequently reducing the (more optimal) shift in their response bias (Akaishi, Umeda, Nagase, & Sakai, 2014).

We first note that the structure of our experiments makes both possibilities unlikely to account for our results. Briefly presented stimuli, like the ones we used, usually do not generate substantial adaptation (Patterson, Wissig, & Kohn, 2013). Further, the orientation of the gratings changed randomly every trial in cued and uncued trials alike, minimizing the probability of stimulus adaptation. Also, cued and uncued trials were interleaved, making it unlikely that subjects would avoid the more frequent response (in the cued trials) trying to somehow balance their responses in the uncued trials. The more frequent tilt was proportionally split between cued and uncued targets such that 70 % of cued and 70 % of uncued targets were tilted to the same side, making "the most frequent response" proportionally the same in both conditions. Furthermore, subjects received trial-by-trial feedback and the same results were obtained when the uncued stimuli were discriminated just as well as the cued ones in Experiments 4A and 4B, completely ruling out the possibility of explaining our results by subjects responding randomly during uncued trials.

However, to further rule out both possibilities we performed an analysis of stimulus adaptation and response compensation (Fig. 12). For measuring adaptation, we calculated the conditional probability (CP1) of each trial's response being equal to the most frequent stimulus given that the preceding stimulus at the same location was the same as the most frequent stimulus. Then, we subtracted from CP1 the conditional probability (CP2) of the response being equal to the most frequent stimulus given that the preceding stimulus on the same location was different from the most frequent stimulus. Formally,

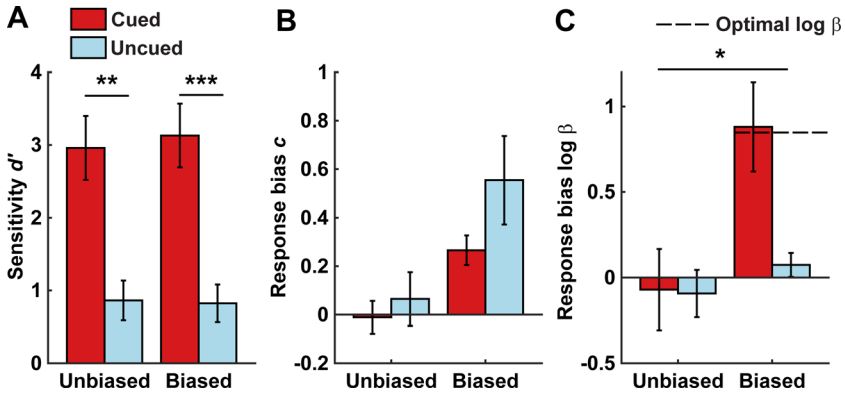

Fig. 11 Results for Experiment 5. **a** Detection sensitivity (d') differed in the unbiased and biased conditions. **b** Response bias measured by c. **c** In the unbiased condition, log β was optimal for cued and uncued trials and not different from each other. In the biased condition, log β was optimal only in cued but not in the uncued trials. Solid bars indicate means and error bars SEM. $*P < 0.05$ $**P < 0.01$ $***P < 0.001$

$$AdaptationIndex = P(R_n = S_1|S_{n-1} = S_1) - P(R_n = S_1|S_{n-1} = S_2) \quad (8)$$

where R_n is the response in a given trial, S_1 is the most and S_2 the least frequent stimulus, respectively, and S_{n-1} is the identity of the stimulus in the same location in the preceding trial.

A similar analysis was performed for response compensation, except that instead of S_{n-1}, we used R_{n-1}, that is, subjects' response in the preceding trial:

$$CompensationIndex = P(R_n = S_1|R_{n-1} = S_1) - P(R_n = S_1|R_{n-1} = S_2) \quad (9)$$

If there was no adaptation or if there was no response compensation, the difference of the two relevant conditional probabilities should equal zero. A positive Adaptation Index value would mean there is positive priming and a negative value would mean there is repulsion. As expected for this kind of briefly presented stimuli, the Adaptation Index was not

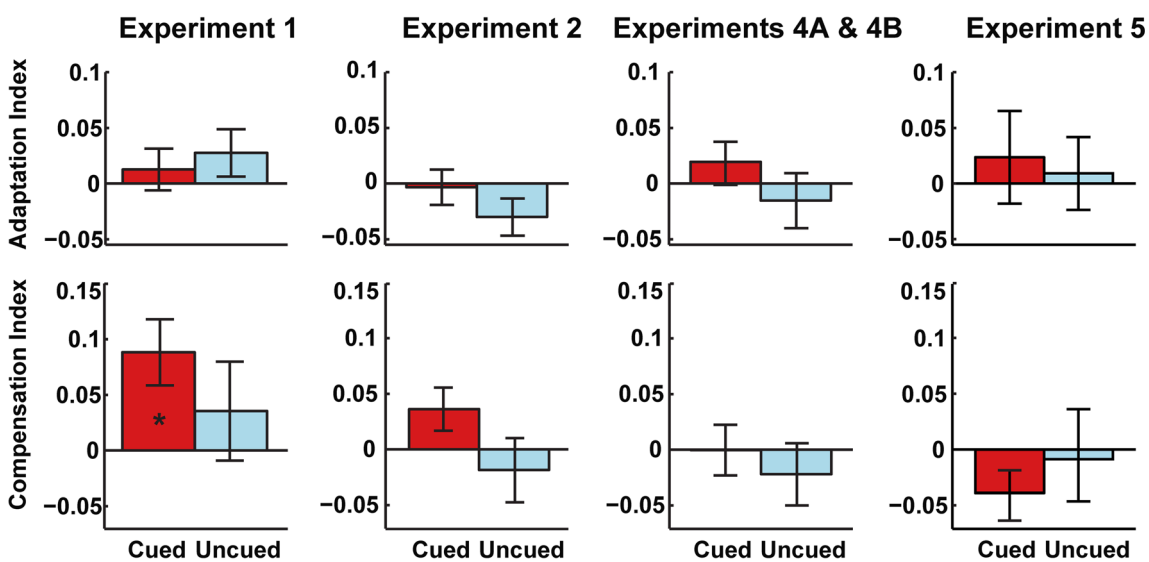

Fig. 12 Analysis of stimulus adaptation and response compensation in the biased conditions. We estimated stimulus adaptation by comparing subjects response on a given trial depending on whether the most frequent stimulus was presented in the previous trial or not. If there was no adaptation the Adaptation Index should equal zero (see Further data analysis section for details). Adaptation Index was not significantly different from zero in the biased condition of all experiments and, importantly, it did not differ between the cued and uncued conditions. A similar analysis was performed for response compensation. Instead of the identity of the stimulus in the previous trial we used the previous response to estimate the degree to which subjects compensate for giving a more frequent response. The only case where Compensation Index was different from zero ($P = 0.01$) was for the cued condition in Experiment 1. However, the difference with the uncued condition was not significant ($P = 0.09$). Therefore, adaptation and response compensation cannot account for the lack of shift in response bias during uncued trials in any of the experiments. Solid bars indicate means and error bars SEM. $*P = 0.01$

significantly different from zero in the biased condition of any of the experiments ($P > 0.05$ in all Experiments) and, importantly, it did not differ between the cued and uncued conditions ($P > 0.05$ in all Experiments) (Fig. 12, top). Similarly, a positive Compensation Index would mean subjects were biased to repeat the response equal to the most frequent stimulus, while a negative value would mean there was some tendency to avoid the response equal to the most frequent stimulus. Again, as expected, subjects were not biased in any direction in all the uncued conditions, ruling out alternative explanations of our results based on subjects just guessing in uncued trials (Fig. 12, bottom). For the cued condition in Experiment 1 subjects deviated from zero bias ($P = 0.1$), sticking to the most frequent response in a mild way (Compensation Index mean was less than 0.1 and, in any case, positive). Further, the difference with the uncued condition was not significant ($P = 0.09$). Summarizing, in all the experiments the cued and uncued conditions were not significantly different from each other by either the Adaptation or the Compensation Index measures, eliminating the worry that the lack of shift in response bias during uncued trials can be explained by stimulus adaptation or response compensation.

General discussion

Taken together, the results from these six experiments confirm the counter-intuitive predictions of our Common Criterion Model. These predictions (Figs. 3 and 8) are counter-intuitive because at first glance they appear to defy the commonsensical Bayesian principle that during uncertainty, prior stimulus probability should have a bigger impact on optimal observers' perceptual decisions when the sensory signal is weak. We by no means claim that all specific Bayesian models are falsified by the findings we present here—rather, we hope our results can inform the arbitration among these models.

Within the scope of perception and action during uncertainty, many studies have shown that human subjects can perform psychophysical tasks in a near-optimal fashion (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Ernst & Banks, 2002; Knill & Pouget, 2004; Körding & Wolpert, 2006; Trommershäuser, Maloney, & Landy, 2008), particularly in the context of combining prior expectation with sensory information (Körding & Wolpert, 2004). Importantly, however, it also has been shown that when subjects had to learn the category structure and the base-rates of a task through continuous sampling, they started the task behaving suboptimally (Maddox & Bohil, 1998). Yet, they became optimal after a few hundred trials (possibly due to the lack of attentional manipulations and to stimulus presentation above threshold), which is something our subjects were unable to do, in some cases even after thousands of trials. This points to the possibility that asking subjects to keep track during a single task of both attended and unattended targets may pose a unique constraint to the perceptual decision system.

Our predictions rest on a simple signal detection theoretic model with two assumptions: that attention reduces the variability of internal perceptual response, and that subjects use a common decision criterion to classify both the attended and unattended stimuli. The assumption of variance reduction is supported by recent research (Bressler & Silver, 2010; Pestilli et al., 2011; Wyart et al., 2012). From a computational point of view, attention has recently been interpreted as facilitating inferring the level of uncertainty or precision during hierarchical perception, thus minimizing the error prediction signal (Feldman & Friston, 2010). At a neuronal circuitry level, it has recently been found that attention reduces the pairwise noise correlation between neurons (Cohen & Maunsell, 2009; Mitchell, Sundberg, & Reynolds, 2009). If we assume that the population response is somewhat monotonically related to the average spike rate of individual neurons, reduction of pairwise noise correlation would mean a less variable population response, because reduced pairwise noise correlation could mean noise is averaged out more effectively (Zohary, Shadlen, & Newsome, 1994).

The other assumption of a common criterion is also supported by previous empirical work within signal detection theoretic (Gorea & Sagi, 2000) and neurophysiological frameworks (Fetsch, Kiani, Newsome, & Shadlen, 2014). Nevertheless, in a recent study, Qamar et al. (2013) found that subjects are able to incorporate trial-to-trial knowledge of different levels of sensory uncertainty when setting their decision boundaries, with some subjects even performing almost optimally. However, it is important to note that in our experiments subjects experienced two stimuli with different degrees of noise (attention) in the same trial, while in Qamar et al.'s experiments subjects saw a single stimulus in every trial and uncertainty changed from trial-to-trial (due to contrast variations). The simultaneous presentation of an attended and an unattended stimulus is a crucial part of our task design, as our model's assumption concerns a common criterion of two simultaneous sets of distributions of which subjects have to keep track.

In terms of neural mechanisms, one way to conceptualize how decision criteria are implemented is to consider the feed-forward connectivity from sensory to decisional regions in the frontal and parietal cortices. For instance, in a discrimination task, a biased criterion towards stimulus S_1 means that for the same level of sensory activation, the decision area is more likely to produce an S_1 response than an S_2 response, which in turn would mean the feed-forward connectivity is strong. This could be implemented by means of more excitable neurons in the decision areas when the discrimination criterion is biased. Alternatively, sensory information processing could be biased by altering the starting point of the evidence accumulation process (Mulder, Wagenmakers, Ratcliff, Boekel, &

Forstmann, 2012) or its drift rate (van Ravenzwaaij, Mulder, Tuerlinckx, & Wagenmakers, 2012). If discrimination criterion is to be understood in one of these ways, in a context where the stimuli are mixed along a single experiment such that the decision circuitry has to read out potentially from either the cued or the uncued sensory representations, the same criterion (or two very similar criteria) will be applied to both. This is because the decision circuitry may not have the resolution to distinguish between the cued and uncued representations by significantly changing the respective feed-forward connectivity by the time subjects are asked to respond. When decision neurons are highly excitable, they may be excitable at very similar rates for both cued and uncued stimuli.

There are several limitations to our model, as it is intended to be simple. The notion of a single, common decision criterion can be further specified in order to avoid potential oversimplifications. It is perhaps more biologically realistic to say that the two decision criteria for the cued and uncued stimuli are very similar, rather than being constrained to be identical (Zak et al., 2012). The identity assumption should be considered a computational convenience of the model and also a tool of higher theoretical value since it allowed us to make empirical predictions that would have been impossible assuming two independent criteria. While it is possible that the brain applies different criteria for cued and uncued trials, it is biologically unrealistic that the firing thresholds implementing the criteria are extremely different during the same trial.

A related question arises regarding the factors that determine the common criterion's value. In our experiments, the common criterion was placed on or short of the optimal response bias value for attended trials. This agrees with previous evidence that response bias tends to be conservative (Gorea & Sagi, 2000; Green & Swets, 1966; Macmillan & Creelman, 2005; Rahnev et al., 2011). However, it is still open why the criterion was closer to the attended optimal response bias value and not, for example, midway between the optimal attended and unattended optimal response biases. One possibility is that since attended stimuli were probed more often than unattended ones they were given preference when setting the common criterion. This pre-eminence of attended stimuli is ecologically valid too, since humans and other animals tend to solve problems and control their behavior guided by attention, and only infrequently they have to react to unattended stimuli in the environment. We emphasize that this is a speculative account that requires further evidential support.

Another concern is that our model's assumption that the internal response distributions are Gaussian might not hold. Recent reports show that in attentional tasks with multiple items the precision of the internal representation itself could be a random variable, perhaps due to fluctuations in attention (Mazyar, van den Berg, & Ma, 2012; Mazyar, van den Berg, Seilheimer, & Ma, 2013; van den Berg, Shin, Chou, George, & Ma, 2012). Nevertheless, as can be inferred from Fig. 1, as long as the distributions remain roughly similar for cued and uncued stimuli, the same qualitative predictions would continue to hold. Our model's predictions depend critically on a reduction in variance by attention paired with a common decision criterion but not on the specific shape of the distributions. Further, Experiment 5 suggests that our main finding cannot be solely attributable to alterations in the shape of the distributions of the internal representations because reducing the number of displayed stimuli (which should have lead to changes in the internal representations) did not affect the response bias patterns seen throughout our experiments.

Conclusions

Here we put to a test a signal detection theoretic model that parsimoniously predicted a priori that subjects would not optimally incorporate the priors into their perceptual judgments in unattended trials during a detection task where attended and unattended trials were interleaved. These results suggest that the commonsensical notion of optimal criterion setting may not apply to situations as common as inattention. Future modeling efforts should shed light on the cognitive constraints faced by the observer in dealing with situations where stimuli have different statistical properties or where there is different variability of internal responses, which may account for this interesting pattern of behavior that may seem "suboptimal" from the outset.

Acknowledgments This work is partially supported by a grant from the Templeton Foundation (6–40689). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Ai Koizumi, Megan Peters, and Alex White for comments on an earlier version of this manuscript, and David Heeger for suggesting that we reward the subjects financially to motivate them to behave optimally.

Conflict of interests The authors declare no competing financial interests.

References

Akaishi, R., Umeda, K., Nagase, A., & Sakai, K. (2014). Autonomous mechanism of internal choice estimate underlies decision inertia. *Neuron, 81*(1), 195–206. doi:10.1016/j.neuron.2013.10.018

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Current Biology, 17*(6), 509–513.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*(4), 700–765. doi:10.1037/0033-295X.113.4.700

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. doi:10.1163/156856897X00357

Bressler, D. W., & Silver, M. A. (2010). Spatial attention improves reliability of fMRI retinotopic mapping signals in occipital and parietal cortex. *NeuroImage, 53*(2), 526–533. doi:10.1016/j.neuroimage.2010.06.063

Brewer, C. A. (Ed.). (2013). *Color Brewer 2.0*. Retrieved January 30, 2015, from http://colorbrewer2.org

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. doi:10.1016/j.visres.2011.04.012

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience, 12*(12), 1594–1600. doi:10.1038/nn.2439

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433. doi:10.1038/415429a

Feldman, H., & Friston, K. J. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, *4*. doi:10.3389/fnhum.2010.00215

Fetsch, C. R., Kiani, R., Newsome, W. T., & Shadlen, M. N. (2014). Effects of cortical microstimulationon confidence in a perceptual decision. *Neuron, 83*(4), 797–804. doi:10.1016/j.neuron.2014.07.011

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 360*(1456), 815–836. doi:10.1098/rstb.2005.1622

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 12380–12384. doi:10.1073/pnas.97.22.12380

Green, D., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (pp. 1–455). New York: Wiley.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719. doi:10.1016/j.tins.2004.10.007

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences, 14*(7), 301–307. doi:10.1016/j.tics.2010.04.006

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*(6971), 244–247. doi:10.1038/nature02169

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10*(7), 319–326. doi:10.1016/j.tics.2006.05.003

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory* (Secondth ed.). Mahwah: Lawrence Erlbaum.

Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology Learning Memory and Cognition, 24*(6), 1459–1482.

Mazyar, H., van den Berg, R., & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision, 12*(6), 10. doi:10.1167/12.6.10

Mazyar, H., van den Berg, R., Seilheimer, R. L., & Ma, W. J. (2013). Independence is elusive: set size effects on encoding precision in visual search. *Journal of Vision*, *13*(5). doi:10.1167/13.5.8

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron, 63*(6), 879–888. doi:10.1016/j.neuron.2009.09.013

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience, 32*(7), 2335–2343. doi:10.1523/JNEUROSCI.4156-11.2012

Patterson, C. A., Wissig, S. C., & Kohn, A. (2013). Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. *The Journal of Neuroscience, 33*(2), 532–543. doi:10.1523/JNEUROSCI. 3345-12.2013

Pestilli, F., Carrasco, M., Heeger, D. J., & Gardner, J. L. (2011). Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. *Neuron, 72*(5), 832–846. doi:10.1016/j.neuron.2011.09.025

Pestilli, F., Viera, G., & Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? *Journal of Vision, 7*(7), 1–12. doi:10.1167/7.7.9

Qamar, A. T., Cotton, R. J., George, R. G., Beck, J. M., Prezhdo, E., Laudano, A., Tolias, A. S., & Ma, W. J. (2013). Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proceedings of the National Academy of Sciences of the United States of America, 110*(50), 20332–20337.

Rahnev, D. A., Bahdo, L., de Lange, F. P., & Lau, H. (2012a). Prestimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *Journal of Neurophysiology, 108*(5), 1529–1536. doi:10.1152/jn.00184.2012

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012b). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology, 107*(6), 1556–1563. doi:10.1152/jn.00985.2011

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience, 14*(12), 1513–1515. doi:10.1038/nn.2948

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience, 1*(3), 165–175. doi:10.1080/17588921003632529

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, 1–12. doi:10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409. doi:10.1016/j.tics.2009.06.003

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron, 59*(2), 336–347. doi:10.1016/j.neuron.2008.05.021

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12*(8), 291–297. doi:10.1016/j.tics.2008.04.010

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America, 109*(22), 8780–8785. doi:10.1073/pnas.1117465109

van Ravenzwaaij, D., Mulder, M. J., Tuerlinckx, F., & Wagenmakers, E.-J. (2012). Do the Dynamics of Prior Information Depend on Task Context? An Analysis of Optimal Performance and an Empirical Test. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00132

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*(2), 113–120.

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America, 109*(9), 3593–3598. doi:10.1073/pnas.1120118109

Zak, I., Katkov, M., Gorea, A., & Sagi, D. (2012). Decision criteria in dual discrimination tasks estimated using external-noise methods. *Attention, Perception, & Psychophysics, 74*(5), 1042–1055. doi:10.3758/s13414-012-0269-0

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature, 370*(6485), 140–143. doi:10.1038/370140a0