NEUROSCIENCE of Consciousness

Neuroscience of Consciousness, 2022, **8(1)**, 1–8

DOI: https://doi.org/10.1093/nc/niac010

Research Article

The positive evidence bias in perceptual confidence is unlikely post-decisional

Jason Samaha[1,*] and Rachel Denison[2]

[1]Department of Psychology, University of California, 1156 High St, Santa Cruz, CA 95064, USA; [2]Department of Psychological and Brain Sciences, Boston University, 64 Cummington Mall, Boston, MA 02215, USA

*Correspondence address. Psychology, University of California, 1156 High St, Santa Cruz, CA 95064, USA. Tel: +831-459-4630; E-mail: jsamaha@ucsc.edu

Abstract

Confidence in a perceptual decision is a subjective estimate of the accuracy of one's choice. As such, confidence is thought to be an important computation for a variety of cognitive and perceptual processes, and it features heavily in theorizing about conscious access to perceptual states. Recent experiments have revealed a "positive evidence bias" (PEB) in the computations underlying confidence reports. A PEB occurs when confidence, unlike objective choice, overweights the evidence for the correct (or chosen) option, relative to evidence against the correct (or chosen) option. Accordingly, in a perceptual task, appropriate stimulus conditions can be arranged that produce selective changes in confidence reports but no changes in accuracy. Although the PEB is generally assumed to reflect the observer's perceptual and/or decision processes, post-decisional accounts have not been ruled out. We therefore asked whether the PEB persisted under novel conditions that addressed two possible post-decisional accounts: (i) post-decision evidence accumulation that contributes to a confidence report solicited after the perceptual choice and (ii) a memory bias that emerges in the delay between the stimulus offset and the confidence report. We found that even when the stimulus remained on the screen until observers responded, and when observers reported their choice and confidence simultaneously, the PEB still emerged. Signal detection-based modeling showed that the PEB was not associated with changes to metacognitive efficiency, but rather to confidence criteria. The data show that memory biases cannot explain the PEB and provide evidence against a post-decision evidence accumulation account, bolstering the idea that the PEB is perceptual or decisional in nature.

Keywords: confidence; metacognition; visual perception; bias; signal detection theory

Introduction

Human perceptual decision-making not only results in a choice about what is perceived but can also produce a sense of confidence in the accuracy of that choice. Several lines of inquiry suggest that the same source of evidence informs both one's choice and one's confidence in that choice. For instance, subjective reports of confidence are well correlated with choice accuracy across a variety of perceptual and mnemonic tasks (Song *et al.* 2011; Kiani *et al.* 2014; Grimaldi *et al.* 2015; Ais *et al.* 2016; Samaha and Postle 2017). Moreover, certain neurons in monkey parietal cortex encode both the choice and confidence level of the animal (Kiani and Shadlen 2009). The close relation between choice accuracy and confidence has been formalized in computational approaches that define confidence as the probability of being correct, thereby casting confidence as an optimal readout of choice uncertainty (Kepecs *et al.* 2008; Kiani *et al.* 2014; Meyniel *et al.* 2015; Sanders *et al.* 2016).

In contrast, a number of recent experiments have demonstrated a bias in confidence reports that renders confidence dissociable from choice accuracy. The so-called "positive evidence bias" (PEB) refers to the finding that confidence seems to overweight the evidence in favor of the correct, or in some formulations, the chosen option (the "positive evidence"), whereas the difficulty of the choice itself is governed by the balance of evidence between choice alternatives (Zylberberg *et al.* 2012; Aitchison *et al.* 2015; Koizumi *et al.* 2015; Samaha *et al.* 2016, 2017, 2019; Peters *et al.* 2017; Rausch *et al.* 2017; Miyoshi *et al.* 2018; Odegaard *et al.* 2018; Miyoshi and Lau 2020; Pereira *et al.* 2020). For example, Koizumi *et al.* (2015) demonstrated, in a motion direction discrimination task, that increasing the number of dots moving in the direction of the correct choice while simultaneously increasing the number of dots moving randomly does not change accuracy but increases confidence.

We have recently demonstrated the PEB in orientation discrimination and estimation tasks. In our protocol, a single luminance-modulated grating is embedded in white noise, and we examine the effect of increasing both the noise and the grating contrast. We observed that proportional increases in grating and noise contrast selectively boost confidence ratings but not discrimination or estimation accuracy (Samaha *et al.* 2016, 2019). Given the importance of accurate confidence for guiding appropriate behaviors (Folke *et al.* 2017; Desender *et al.* 2018), and the role of confidence

Received: 14 January 2022; **Revised:** 13 May 2022; **Accepted:** 27 June 2022

effects in theoretical accounts of conscious perception (Dehaene and Changeux 2011; Lau and Rosenthal 2011; Brown *et al.* 2019), an understanding of why confidence diverges from accuracy in the PEB is warranted.

Theoretically, the PEB in orientation tasks could emerge at perceptual stages of processing if the orientation strength appears stronger or more visible to the observer when grating and noise contrast are increased, thereby causing the change in confidence. The PEB could occur at decision stages, e.g. if stimulus information other than orientation is used to inform confidence. A third option is that the bias emerges after the initial perceptual decision is made. Here, we consider two such post-decisional accounts of how confidence and accuracy can dissociate and test whether they explain the PEB.

The first account is termed post-decision evidence accumulation (Navajas *et al.* 2016). It is based on the idea that choice evidence continues to accumulate even after the initial decision is made. If, as is the case with many prior experiments on the PEB, confidence is solicited *after* the stimulus choice, then confidence can be based on different levels of evidence than the choice, leading to a dissociation. This account has been leveraged to explain changes of mind, among other confidence-related results (Navajas *et al.* 2016). Note that to explain the PEB, post- (but not pre-) decision evidence would need to be selectively biased by the correct (or chosen) option.

The second account is that the PEB does not arise in perceptual decision-making, *per se*, but is a bias that arises in short-term memory. This is plausible because several experiments have observed a PEB in memory-based confidence judgments (Zawadzka *et al.* 2017; Miyoshi *et al.* 2018), and to date, experiments demonstrating the PEB in perception tasks have used relatively short, fixed-duration stimulus presentations and/or separate choice and confidence responses. Thus, the confidence judgment is typically rendered a second or more after the stimulus was perceived. Although a relevant experiment found that confidence ratings were biased by the magnitude (rather than balance) of evidence even when using a simultaneous choice and confidence report, the stimuli used were short (86 ms), allowing time between perception and response input (Aitchison *et al.* 2015). And although some work using longer stimuli (Kiani and Shadlen 2009) has produced behavioral patterns consistent with a PEB (Maniscalco *et al.* 2016; Zawadzka *et al.* 2017; Miyoshi *et al.* 2018), positive evidence was not explicitly manipulated in those studies, and such patterns may be compatible with other models (Rausch and Zehetleitner 2019). Thus, experiments directly manipulating positive evidence have not yet combined simultaneous choice and confidence responses with stimuli that remain visible until a response is given.

Here, we test both of these accounts in a new experiment in which observers issue their choice and confidence reports at the same moment in time with a single key press (minimizing post-decisional evidence accumulation) and where confidence is reported while the stimulus is still visible (eliminating any memory-based bias). Using a signal detection theory (SDT) analysis, we find that the PEB persists under these conditions and is not associated with changes in metacognitive efficiency. These results suggest that the PEB likely arises during perception or decision-making, rather than during a post-decisional stage.

Materials and methods

Participants

Twenty-six participants (age range: 18–35 years; 17 female) from the University of Wisconsin-Madison community completed the experiment. Twenty-three of the participants provided data deemed suitable for hypothesis testing (see the 'Staircase Procedure' section). All participants reported normal or corrected visual acuity, provided written informed consent, and were compensated monetarily. Sample size was based on prior experiments we have done looking at the PEB across two or more conditions (Samaha *et al.* 2016, 2019). This experiment was conducted in accordance with the University of Wisconsin Institutional Review Board and the Declaration of Helsinki.

Stimuli

Visual stimuli were composed of a sinusoidal luminance grating (1.1 cycles per degree, zero phase) embedded in white noise and presented centrally within a circular aperture [2.5 degrees of visual angle (DVA)]. The orientation of the grating was randomly chosen on each trial to be 45° or −45° tilted from vertical. The noise component of the stimulus was created anew on each trial by randomly sampling each pixel's luminance from a uniform distribution. A fixation point (light gray, 0.19 DVA) was centered on the screen throughout the trial. Stimuli were presented on a gray background on an iMac computer screen (52 cm wide by 32.5 cm tall; 1920 by 1200 pixel resolution; 60-Hz refresh rate) using PsychToolbox 3 (Pelli 1997; Kleiner *et al.* 2007) running in MATLAB 2015b (MathWorks, Natick, MA) viewed from a chin rest at a distance of 62 cm.

Staircase procedure

The PEB can be demonstrated by embedding gratings in noise under two different conditions (Fig. 1a): one high contrast grating averaged with high contrast noise (which we will refer to as 'high positive evidence' or high PE) and one low contrast grating averaged with low contrast noise ['low positive evidence' or low PE (Samaha *et al.* 2016, 2019)]. If there is a PEB, then the high PE stimulus will produce higher confidence ratings but not higher accuracy. We therefore started each main task with a staircase procedure designed to find two levels of grating contrast that produce equal levels of discrimination accuracy when embedded in low and high contrast noise. Specifically, we started with a 50% and a 100% Michelson contrast noise patch and used the Quest procedure as implemented in PsychToolbox 3 to find a grating contrast that produced ∼75% accuracy when averaged with the 50% noise patch and another contrast level that produced the same accuracy when averaged with the 100% noise patch. In previous work, we had used a single staircase to find a grating contrast threshold that when averaged with 100% noise should produce ∼75% accuracy and then simply halved the contrast of the grating and the noise to make the low PE stimulus. Although this procedure previously worked to match accuracy at the group level (as predicted by Weber's law), here we matched accuracy empirically at the individual participant level by separately staircasing high and low PE stimuli. To this end, we ran 20 practice trials followed by 200 trials of the staircase before each of the two main task conditions (fixed duration vs. response-dependent; see the 'Main Task Procedure' section). High PE and low PE staircases were interleaved, and the final grating contrast for the low and high PE stimuli was computed as the mean of the Quest posterior distribution. The thresholds from some participants, however, deviated substantially from the predicted value of 50% lower grating contrast in the low PE condition. To ensure that variability in the efficacy of the staircase did not overly influence the results, we excluded three subjects whose low PE thresholds were less than 20% or greater than 80% of their high PE thresholds. The staircase followed the same task structure as the main tasks (described next), with the

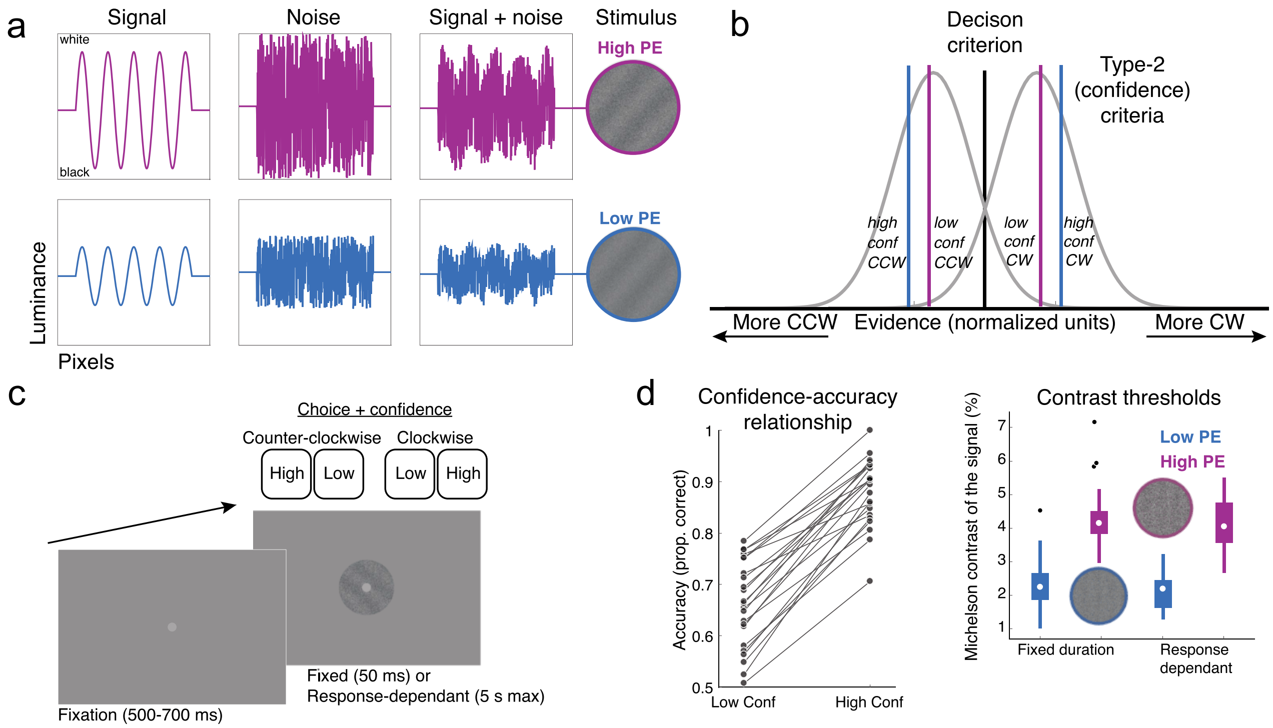

Figure 1. (a) Schematic of stimuli used in the positive evidence (PE) manipulation. High PE stimuli were composed of 100% contrast white noise averaged with a higher contrast grating (contrast thresholds in panel D, right), whereas low PE stimuli contained 50% contrast noise averaged with a lower contrast grating. (b) SDT model of confidence and performance. Gaussian distributions represent normalized internal evidence for clockwise (CW) and counterclockwise (CCW) stimuli across trials. Type-2 (confidence) criteria (colored vertical lines) are exceeded when a certain level of evidence (distance from decision criterion) is passed. The PEB could arise if, on high PE trials, the confidence criteria shift closer to the decision boundary, leading to more frequent reports of "high confidence" without changes in accuracy. Note that such a shift would occur in the normalized evidence space if, e.g. the means and variances of the absolute evidence distributions scaled for high PE stimuli but the absolute confidence criteria did not change (not shown). Also note that the PEB can be modeled in two-dimensional SDT space (Maniscalco *et al.* 2016; Samaha *et al.* 2017, 2020; Miyoshi *et al.* 2018; Miyoshi and Lau 2020). (c) Task schematic. On each trial, a high or low PE stimulus tilted either 45° or −45° from vertical was presented. In different blocks, the stimulus was either presented for a fixed duration of 50 ms, or until a response was made (up to 5 seconds). Choice (CW or CCW) and confidence level (high or low) were given with a single button press. (d) Left, accuracy is higher for decisions endorsed with high confidence. Right, boxplots of contrast thresholds for each PE level and task, determined from a pre-task adaptive procedure. Note that these are the contrast levels of the gratings prior to averaging with 100% (high PE) or 50% (low PE) contrast white noise. Inset shows example (vertical) stimuli from an observer with a low PE threshold of 2.5% and high PE threshold of 5%

only exception that the contrast of the grating component of the stimulus was adapted using Quest.

Main task procedure

To test whether the PEB persisted when the target stimulus remained onscreen until the participant's response, we tested both a fixed-duration condition in which the stimulus duration was 50 ms and a response-dependent condition in which the stimulus was displayed until a response was made. The only difference between conditions was the duration of the target stimulus.

Each trial began with an intertrial interval (ITI) randomly drawn from a uniform distribution of durations between 500 and 700 ms. During the ITI, only the central fixation point was displayed. A target stimulus followed, which was randomly selected on each trial to be 45° or −45° tilted from vertical and have high or low PE. Choice (clockwise/counterclockwise) and confidence (two levels: high or low) were indicated with a single keypress. With their left hand, participants used the "F" and "D" key to indicate counterclockwise tilt with low and high confidence, respectively. A right-hand response using the "J" or "K" key indicated clockwise tilt with low and high confidence, respectively. In this way, the confidence report was made with the same amount of time allotted for evidence accumulation as the orientation choice (Fig. 1c). Responses were required within 5 seconds of stimulus onset, and trials with late responses were repeated at the end of the block. Participants completed 3 blocks of 80 trials each of the fixed-duration condition and then 3 blocks of 80 trials each of the response-deponent condition, or vice versa (counterbalanced across participants). Before starting either duration condition for the first time, the staircase was run to find high and low PE thresholds for that condition. Those threshold contrast values were then used for the main task.

Model-free analysis

For each combination of PE and stimulus duration, we computed mean confidence ratings and accuracy (proportion correct) across trials. We submitted these measures to separate 2-by-2 repeated-measures ANOVA with PE (high or low) and stimulus duration (fixed or response-dependent) as factors. If the PEB emerges in memory, then we should see a confidence bias in the fixed-duration stimulus condition, but not in the response-dependent condition, corresponding to an interaction between PE and duration when predicting confidence. If the PEB is present in both conditions, we expect a main effect of PE on confidence. Additionally, we expect no main effect of PE on the proportion of correct discrimination responses if we adequately controlled performance. Lastly, we analyzed the mean and median response time (RT) in all conditions to determine whether changes in confidence (without changes in accuracy) were associated with decision duration, as predicted by some drift-diffusion models of confidence (Kiani and Shadlen 2009; Kiani *et al.* 2014; Zylberberg *et al.* 2016).

Model-based analysis

In addition to a model-free analysis (which ensures that results are not based on any particular model assumptions), we also used an SDT modeling framework (Fig. 1b) to assess the impact of PE (high or low) and stimulus duration (fixed or response-dependent) on behavior. The model, called meta-d′ (Maniscalco and Lau 2012; Fleming and Lau 2014), estimates (i) type-1 sensitivity (d′); (ii) type-2 criteria (confidence criteria), which reflect how much (normalized) stimulus evidence is needed to commit to a high or low confidence response; and (iii) metacognitive efficiency (meta-d′ – d′), which reflects how well confidence ratings distinguish between correct and incorrect responses relative to an ideal observer with no information loss between the type-1 (choice) and type-2 (confidence) decisions. Meta-d – d takes on negative values when metacognitive sensitivity is worse than what would be expected by an ideal (according to SDT) metacognitive observer, whereas values close to zero indicate no deviation from optimality. Note, however, that optimality here rests on the assumption that stimulus evidence distributions are Gaussians of equal variance, an assumption that has recently been challenged in model simulations showing that heuristics, like the PEB, can actually improve metacognitive accuracy when distributions are not equal variance (Miyoshi and Lau 2020).

We derived a single estimate of confidence criteria for each observer by averaging over the absolute values of the criteria associated with clockwise and counterclockwise stimuli. The resulting metric indicates normalized (z) evidence criteria needed to commit to a "high confidence" decision; lower values therefore indicate more liberal type-2 criteria (less evidence is needed). Note that because evidence is normalized in SDT (see Fig. 1b), type-2 criterion changes can occur because either the absolute amount of evidence needed to respond with high confidence changes or the distributions of evidence themselves change while the absolute positions of the criteria remain the same. We do not commit to either mechanistic interpretation of criterion changes here but simply use this measure as a means of replicating any shift in confidence with PE in a widely used model-based framework. The model was estimated from choice and confidence ratings for each participant using a Bayesian model called HMeta-d (Fleming 2017), wherein the calculation of d′ follows the model of Lee (2008). This open-source software is available at https://github.com/metacoglab/HMeta-d. We used the MATLAB implementation (function *fit_meta_d_mcmc.m*), which implements a Markov-Chain Monte Carlo sampling procedure to estimate posterior distributions over parameters. We ran the function with 3 chains, 1000 burn-in samples, and 10 000 recorded samples in each chain.

We then entered each of the three parameters (confidence criteria, meta-d′ – d′, and d′) into separate 2-by-2 repeated-measures ANOVAs with PE (high or low) and stimulus duration (fixed or response-dependent) as factors. If increasing PE boosts confidence ratings in both stimulus duration conditions, we expect a main effect of PE on confidence criteria and possibly on meta-d′ – d′ (if the PEB also alters metacognitive efficiency). Specifically, higher confidence ratings would correspond to lower confidence criteria, i.e. less normalized evidence required to make a "high confidence" response. If, on the other hand, the PEB is caused by a memory bias, then confidence measures should interact with stimulus duration, as PEB would only occur in the fixed-duration condition where judgments are made on the basis of a memory representation. If our staircase procedure is adequately controlled for accuracy, we expect no effects of PE on d′.

Results

Staircase thresholds

Contrast thresholds from the staircase procedure conformed well to Weber's law: in the fixed-duration staircase, mean [±standard error of the mean (SEM)] Michelson contrast for the low PE (2.33 ± 0.15%) and high PE (4.35 ± 0.20%) produced a ratio of 53 ± 2.5%. The response-dependent staircase produced threshold estimates for low PE (2.12 ± 0.12%) and high PE (4.12 ± 0.16%) with a ratio of 51 ± 2.5%. Threshold contrasts were numerically, although not significantly ($P = 0.18$), lower for the response-dependent task, indicating that slightly lower thresholds are achieved when the observer has unlimited time to view the stimulus (Fig. 1d).

General confidence behavior

Before addressing our main hypotheses, we confirmed that each participant used their confidence responses appropriately in that higher confidence was associated with higher accuracy. Collapsing across PE levels and task, every participant had higher mean (±SEM) accuracy on trials endorsed with high, as compared to low, confidence [Fig. 1d; $P(\text{correct})_{\text{LowConf}} = 0.66(0.017)$, $P(\text{correct})_{\text{HighConf}} = 0.88(0.013)$, $t(22) = 16.56$, $P = 6.5^{-14}$]. Moreover, mean (±SEM) confidence was higher on correct compared to error trials for both fixed-duration stimuli [$\text{Conf}_{\text{error}} = 1.40(0.050)$, $\text{Conf}_{\text{correct}} = 1.65(0.051)$, $t(22) = -10.48$, $P = 5.0^{-10}$] and response-dependent viewing conditions [$\text{Conf}_{\text{error}} = 1.26(0.043)$, $\text{Conf}_{\text{correct}} = 1.55(0.049)$, $t(22) = -9.11$, $P = 6.3^{-9}$]. Median reaction times were also faster for correct compared to error trials for both fixed-duration stimuli [$\text{RT}_{\text{error}} = 782(36)$, $\text{RT}_{\text{correct}} = 700(23)$, $t(22) = 3.96$, $P = 6.6^{-4}$] and response-dependent stimuli [$\text{RT}_{\text{error}} = 1044(75)$, $\text{RT}_{\text{correct}} = 872(47)$, $t(22) = 4.42$, $P = 2.7^{-4}$].

Stimulus viewing time

Since viewing time was controlled by the observer in the response-dependent condition, average stimulus duration varied from person to person between a range of 0.7 (minimum) and 1.8 (maximum) seconds, with a mean (SD) across subjects of 1.05 (0.30) seconds. In the fixed-duration condition, stimuli were always presented for 50 ms.

Model-free results

These data are shown in Fig. 2. We found no main effect of PE on discrimination mean (±SEM) accuracy [$p(\text{correct})_{\text{LowPE}} = 0.79$ (0.015), $P(\text{correct})_{\text{HighPE}} = 0.78$ (0.013), $F(1,22) = 0.10$, $P = 0.756$] nor an effect of stimulus duration [$P(\text{correct})_{\text{FixedDuration}} = 0.80(0.015)$, $P(\text{correct})_{\text{ResponseDependant}} = 0.77(0.016)$, $F(1,22) = 2.01$, $P = 0.170$], indicating that the staircasing procedure effectively equated discriminability across PE levels and tasks. Additionally, there was no interaction between PE and duration when predicting accuracy [$F(1,22) = 0.05$, $P = 0.833$]. Consistent with the PEB, however, we observed a significant main effect of PE on mean (±SEM) confidence ratings [$\text{Conf}_{\text{LowPE}} = 1.53(0.047)$, $\text{Conf}_{\text{HighPE}} = 1.56(0.046)$, $F(1,22) = 5.08$, $P = 0.034$], showing that high PE was associated with higher confidence (consistent with a more liberal threshold for reporting high confidence). This main effect argues against post-decision evidence accumulation as the source of the PEB. Importantly, PE and duration did not interact to predict confidence [$F(1,22) = 0.4$, $P = 0.532$], suggesting that the PEB persists

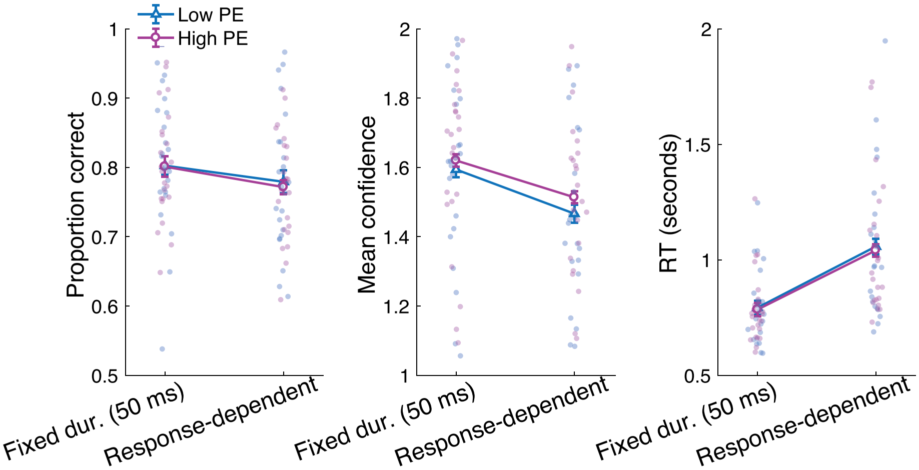

Figure 2. Model-free results depicting orientation discrimination accuracy (left), average confidence rating (middle), and RTs (right) as a function of PE (high or low) and stimulus duration (fixed or response-dependent). A significant main effect of PE on confidence without an effect on accuracy confirms the PEB. A lack of interaction between PE and duration for confidence ratings suggests the bias is not task-dependent. No PE effect was observed on mean RT, although a duration effect is evident. Error bars show ±1 within-subject SEM (Morey 2008); dots are individual observers

even when no memory is required. We also observed a main effect of duration on confidence [$\text{Conf}_{\text{FixedDuration}} = 1.61(0.051)$, $\text{Conf}_{\text{ResponseDependant}} = 1.49(0.047)$, $F(1,22) = 10.95$, $P = 0.003$], indicating lower confidence when the stimulus duration was response-dependent.

To compliment the accuracy and confidence results, we also assessed mean and median RTs as a function of PE and stimulus duration (Fig. 2). Using mean (±SEM) RT, there was no main effect of PE [$\text{RT}_{\text{HighPE}} = 922(41)$, $\text{RT}_{\text{LowPE}} = 915(41)$, $F(1,22) = 1.72$, $P = 0.203$], indicating that average RT changes did not accompany PE-related changes in confidence. There was also no PE-by-duration interaction for RT [$F(1,22) = 0.16$, $P = 0.696$]. However there was a clear main effect of duration [$\text{RT}_{\text{FixedDuration}} = 788(31)$, $\text{RT}_{\text{ResponseDependant}} = 1049(63)$, $F(1,22) = 19.84$, $P = 0.0002$], reflecting the fact that participants took longer to respond when the stimulus viewing duration was under their control. These results held when using median RT as well. There was no main effect of PE [$\text{RT}_{\text{LowPE}} = 785(28)$, $\text{RT}_{\text{HighPE}} = 778$ (28), $F(1,22) = 2.69$, $P = 0.115$], no interaction with duration [$F(1,22) = 0.97$, $P = 0.53$], and a significant main effect of duration [$\text{RT}_{\text{FixedDuration}} = 712(23)$, $\text{RT}_{\text{ResponseDependant}} = 899(49)$, $F(1,22) = 16.62$, $P = 0.0005$].

Model-based results

Using the SDT model of metacognition, meta-d′, we estimated type-1 sensitivity (d′), type-2 criteria, which determine how much (normalized) evidence is needed to endorse a response with high confidence, and type-2 sensitivity adjusted for type-1 sensitivity (meta-d′ – d′). These data are summarized in Fig. 3. The staircase procedure effectively matched perceptual sensitivity across PE conditions. We found no main effect of PE on mean (±SEM) d′ [$\text{d}'_{\text{HighPE}} = 1.64(0.09)$, $\text{d}'_{\text{LowPE}} = 1.70(0.11)$, $F(1,22) = 0.38$, $P = 0.541$]. There was also no main effect of stimulus duration on d′ [$\text{d}'_{\text{FixedDuration}} = 1.75(0.11)$, $\text{d}'_{\text{ResponseDependant}} = 1.59(0.11)$, $F(1,22) = 1.42$, $p = 0.245$], indicating that the staircase matched performance between the fixed and response-dependent duration conditions. Duration and PE did not interact to predict d′ [$\text{F}(1,22) = 0.07$, $P = 0.794$].

In contrast to the lack of effects on type-1 performance, the PE manipulation produced a significant main effect on type-2 criteria [$\text{t2crit}_{\text{HighPE}} = 0.79(0.09)$, $\text{t2crit}_{\text{LowPE}} = 0.87(0.08)$, $F(1,22) = 10.38$, $P = 0.0039$], which recapitulates the finding that confidence was overall higher with high PE. High PE caused more liberal type-2 criteria, indicating that less evidence, relative to the normalized distribution of evidence, was required to respond with high confidence. This main effect, coupled with a lack of significant interaction with stimulus duration [$F(1,22) = 1.31$, $P = 0.265$], suggests that the PEB does not depend on the stimulus being presented for a short duration. If anything, the type-2 bias was stronger in the response-dependent duration condition (Fig. 3). These findings indicate that the PEB does not arise in memory. In addition, we observed a significant main effect of stimulus duration on type-2 criteria, such that participants used more conservative type-2 criteria in the response-dependent duration task compared to the fixed-duration task [$\text{t2crit}_{\text{FixedDuration}} = 0.79(0.09)$, $\text{t2crit}_{\text{ResponseDependant}} = 0.93(0.08)$, $F(1,22) = 16.52$, $P = 0.0005$], recapitulating the finding that confidence was lower in the response-dependent task. To our knowledge, this is the first report of changes in type-2 criteria and confidence with stimulus duration without concurrent changes in accuracy (sensitivity).

In the above analysis, the confidence criteria associated with each choice (i.e. on each "side" of the decision criterion) were collapsed (see the Methods section) to provide a single measure of type-2 criterion. To check that the side of the criteria was not an explanatory factor, we re-ran the above analysis but with the (absolute value) of the un-collapsed criteria and with "side" as an additional factor in the ANOVA. There remained a main effect of PE [$F(1,22) = 10.38$, $P = 0.004$] and task [$F(1,22) = 16.52$, $P = 0.0005$] in predicting criterion, but there was no main effect of "side" [$F(1,22) = 0.83$, $P = 0.37$], indicating the criteria were approximately symmetrical about zero. Furthermore, "side" did not significantly interact with task [$F(1,22) = 3.21$, $P = 0.076$] nor with PE [$F(1,22) = 3.11$, $P = 0.081$], suggesting that it is reasonable to average criteria across the factor "side" and that the main effect of PE is still present when including side as a predictor.

The effects of PE on confidence reports were specific to type-2 criteria. For metacognitive efficiency (meta-d′ – d′), which describes how well confidence tracks performance while accounting for task difficulty (under SDT assumptions that evidence is Gaussian and equal variance), we observed no main effect of PE [$\text{meta-d}' - \text{d}'_{\text{HighPE}} = -0.23(0.08)$, $\text{meta-d}' - \text{d}'_{\text{LowPE}} = -0.30(0.12)$, $F(1,22) = 0.32$, $P = 0.578$] nor of task [$\text{meta-d}' - \text{d}'_{\text{FixedDuration}} =$

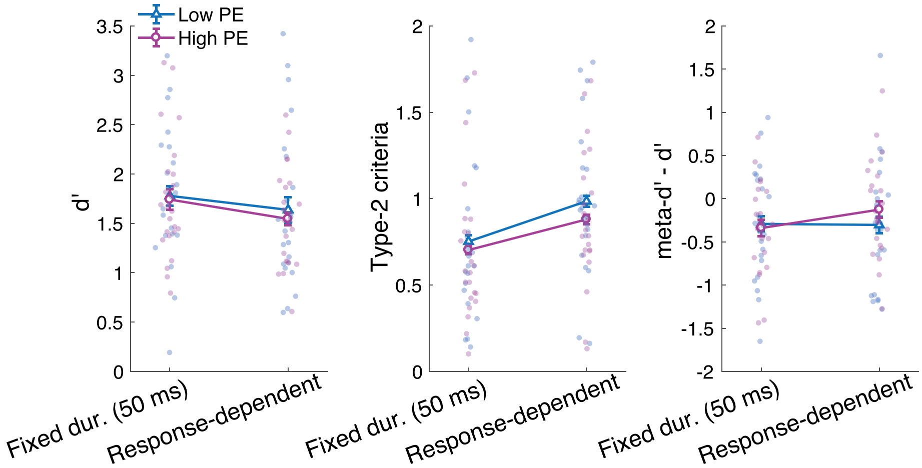

Figure 3. Model-based results showing d′ (left), type-2 criteria (middle), and metacognitive efficiency (meta-d′ – d′; right) as a function PE (high or low) and stimulus duration (fixed or response-dependent). A main effect of PE on type-2 criteria indicates more liberal criteria with high PE, and no effect on d′ confirms the PEB. A lack of interaction on type-2 criteria suggests that the PEB is not task-dependent, and a lack of any effects on metacognitive sensitivity suggests that the PEB is driven by a bias in overall confidence level rather than a change in the relation between confidence and accuracy. Error bars show ±1 within-subject SEM (Morey 2008); dots are individual observers

−0.31(0.09), meta-d′ – $\text{d}'_{\text{ResponseDependant}}$ = −0.21(0.11), $F(1,22) = 0.78$, $P = 0.385$], and there was no interaction [$F(1,22) = 1.41$, $P = 0.248$]. Moreover, no simple effects (tested via paired-sample *t*-tests) were significant (all *P* values > 0.179). This suggests that the PEB is a bias in overall confidence and does not alter the introspective accuracy of confidence judgments.

Discussion

We investigated a previously documented bias in confidence ratings that is thought to occur because of an overreliance of confidence computations on evidence for the chosen alternative (Zylberberg *et al.* 2012, 2014; Koizumi *et al.* 2015; Maniscalco *et al.* 2016; Samaha *et al.* 2016, 2017; Peters *et al.* 2017; Odegaard *et al.* 2018). We tested the PEB under novel circumstances where memory requirements were eliminated and post-decision evidence accumulation was discouraged. Our main finding, evident in both a model-free analysis and an SDT-based model of confidence behavior, is that the PEB persisted under these new circumstances, thereby making memory and post-decision evidence accounts of the PEB less likely. Furthermore, we found that the PEB is not associated with a reduction of metacognitive efficiency, replicating Maniscalco *et al.* (2016), but was instead accounted for by a shift in type-2 criteria, such that less (normalized) evidence was needed to endorse a decision with high confidence when both signal and noise were increased. (Note that this type-2 criteria shift is measured in a variance-normalized evidence space and may or may not correspond to a change in the absolute amount of evidence needed to commit to a "high confidence" response).

Subjective reports of perception, such as confidence or visibility ratings, often figure as evidence in debates about the nature of conscious perception, the role of consciousness in behavior, and the neural correlates of consciousness (Lamme 2003; de Lafuente and Romo 2005; Lau and Passingham 2006; Block 2011; Dehaene and Changeux 2011; Hesselmann *et al.* 2011; King and Dehaene 2014; Vandenbroucke *et al.* 2014; Samaha 2015; Brown *et al.* 2019). The phenomenon of blindsight has been particularly informative in the study of conscious perception precisely because it dissociates subjective and objective aspects of perception (Weiskrantz 1986; Cowey and Stoerig 1991; Azzopardi and Cowey 1997; Overgaard 2011). Whenever similar, albeit far less dramatic, dissociations occur in typical individuals, as is the case in the PEB, it raises the possibility of insight into the mechanisms involved in subjective conscious perception independent of objective task performance. However, such insight is only meaningful for perception if the change in the subjective report actually reflects something about an individual's conscious experience, rather than a change in post-perceptual processes. By reducing the possibility that memory effects and post-decision evidence accumulation are sources of the PEB, our results indicate that a perceptual account of the PEB remains viable—along with the possibility that the bias is caused at some decisional, rather than the perceptual stage of processing.

If the bias is genuinely perceptual, which future work will continue to address, it would suggest that different computations underlie subjective visual appearance and objective decision-making. The PEB is thought to occur because subjective reports are overreliant on the magnitude of evidence for a choice, whereas objective performance is driven by the balance of evidence between choice alternatives. If the contents of consciousness reflect such biased computations, future work may capitalize on this dissociation to identify brain processes that down-weight unchosen alternatives. As an example of this strategy, a recent experiment recorded population activity in the superior colliculus (SC) of macaques while inducing the PEB with motion stimuli (Odegaard *et al.* 2018). It was found that choice-predictive activity in SC neurons was sensitive to confidence and accuracy when the two measures were correlated but were not sensitive to confidence when accuracy was held constant using the PEB. Continued use of the PEB paradigm could serve as a sensitive tool to reveal neural dynamics associated specifically with subjective reports of sensory processing.

A limitation of the present work is that, even though we included a condition where observers responded with their choice and confidence simultaneously while still viewing the stimulus, internally, observers may first commit to their choice and then accumulate more evidence before deciding on their confidence. Under this hypothesis, a post-decision evidence accumulation account could still explain the PEB in our data if, after internally committing to a choice, observers accumulate additional evidence for the confidence report. In the "fixed-duration" condition, this additional evidence would have to come from persistent neural signals; in the "until response" condition, the additional evidence could come from the stimulus itself, which remains on the screen until the report. Such an internal sequential decision account may be difficult to exclude on the basis of behavioral data alone. However, the current data place considerable constraints on the way these internal decisions would have to play out.

To explain the current data, the post-choice evidence accumulation for confidence (i) would need to be biased in favor of the correct or chosen option and (ii) could not cause observers to revise their choice, even though they have every opportunity to do so before the unspeeded, simultaneous report of choice and confidence. Therefore, dissociations between accuracy and confidence would have to arise from changes in confidence in the absence of changes of mind. Importantly, these two criteria are in regards to the 'comparison' between high and low PE conditions. To explain the PEB, then, these two criteria would have to apply differentially to the high and low PE trials. That is, post-decision evidence would (i) have to be more biased for high PE compared to low PE trials and (ii) not lead to more choice revisions for one PE condition. Only if these two criteria are met could post-decision evidence accumulation result in higher confidence for high vs. low PE (due to criterion 1) with no difference in accuracy (due to criterion 2)

Whether such constraints are likely to be met or not is up for debate, but, results from prior experiments suggest that the post-decision account of the PEB may be unlikely. Specifically, prior work using short-duration stimuli probed the PEB with various delays, up to 12 seconds, between stimulus presentation and choice/confidence responses (Samaha *et al.* 2019). The experiment found the PEB to be independent of delay: even after 12 seconds of delay without stimulus presentation, confidence responses still showed a PEB. It seems unlikely that additional evidence would have been accumulated yet not used to alter the observer's decisions, given the 12 seconds of delay participants had to revise their choice. Several experiments conducted by Yu *et al.* (2015) fit a series of evidence accumulation models to data with variable inter-judgment intervals (i.e. the time between the choice and confidence response) and found that post-decision evidence did not selectively accumulate for the correct or chosen option (i.e. no "confirmation bias" was found in the post-decision dynamics). Future work could systematically vary stimulus presentation time, type-1 decision time, and type-2 decision time to gain more insight into whether post-decision evidence accumulation contributes to the PEB. Additionally, using models that explicitly contain post-decision evidence accumulation (Pleskac and Busemeyer 2010) in tasks that manipulate positive evidence could glean further insight into the origins of this bias.

Dissociating confidence and performance has further promise in aiding research into the function of subjective perceptual states. Using the PEB in an orientation estimation task, we recently demonstrated that enhancing confidence independent of accuracy was associated with stronger serial dependencies in orientation estimation between trials (Samaha *et al.* 2019). This finding suggested that a candidate function of subjective confidence is enhancing temporal continuity between perceptual inputs. Using a related confidence–accuracy dissociation paradigm, Desender *et al.* (2018) found that reducing confidence without changing performance led participants to seek out additional perceptual information before committing to a choice. This would suggest that an additional function of subjective confidence is to inform future decision-making policies. Future work will be needed to determine other functions of confidence computations as well as the ultimate perceptual or decisional locus of this bias in a subjective visual report.

Supplementary data

Supplementary data is available at *NCONSC* online.

Data availability

In accordance with the practices of open science and reproducibility, all task scripts, raw data, and code used in the present analyses are freely available through the Open Science Framework (https://osf.io/5m9wd/).

Funding

This work was supported by MH095984 and John Templeton Foundation grant number 48365 to J.S. and R.D.

Conflict of interest statement

None declared.

References

Ais J, Zylberberg A, Barttfeld P *et al.* Individual consistency in the accuracy and distribution of confidence judgments. *Cognition* 2016;**146**:377–86.

Aitchison L, Bang D, Bahrami B *et al.* Doubly Bayesian analysis of confidence in perceptual decision-making. *PLOS Comput Biol* 2015;**11**:e1004519.

Azzopardi P, Cowey A. Is blindsight like normal, near-threshold vision? *Proc Natl Acad Sci* 1997;**94**:14190–4.

Block N. Perceptual consciousness overflows cognitive access. *Trends Cogn Sci* 2011;**15**:567–75.

Brown R, Lau H, LeDoux JE. Understanding the higher-order approach to consciousness. *Trends Cogn Sci* 2019;**23**:754–68.

Cowey A, Stoerig P. The neurobiology of blindsight. *Trends Neurosci* 1991;**14**:140–5.

de Lafuente V, Romo R. Neuronal correlates of subjective sensory experience. *Nat Neurosci* 2005;**8**:1698–703.

Dehaene S, Changeux J-P. Experimental and theoretical approaches to conscious processing. *Neuron* 2011;**70**:200–27.

Desender K, Boldt A, Yeung N. Subjective confidence predicts information seeking in decision making. *Psychol Sci* 2018;**29**:0956797617744771.

Fleming SM. HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neurosci Conscious* 2017;**3**:nix007.

Fleming SM, Lau HC. How to measure metacognition. *Front Hum Neurosci* 2014;**8**. 10.3389/fnhum.2014.00443.

Folke T, Jacobsen C, Fleming SM *et al.* Explicit representation of confidence informs future value-based decisions. *Nat Hum Behav* 2017;**1**:0002.

Grimaldi P, Lau H, Basso MA. There are things that we know that we know, and there are things that we do not know we do not know: confidence in decision-making. *Neurosci Biobehav Rev* 2015;**55**:88–97.

Hesselmann G, Hebart M, Malach R. Differential BOLD activity associated with subjective and objective reports during "blindsight" in normal observers. *J Neurosci* 2011;**31**:12936–44.

Kepecs A, Uchida N, Zariwala HA *et al.* Neural correlates, computation and behavioural impact of decision confidence. *Nature* 2008;**455**:227–31.

Kiani R, Corthell L, Shadlen MN. Choice certainty is informed by both evidence and decision time. *Neuron* 2014;**84**:1329–42.

Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science (New York, N Y)* 2009;**324**:759–64.

King J-R, Dehaene S. A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philos Trans R Soc Lond B Biol Sci* 2014;**369**:20130204.

Kleiner M, Brainard D, Pelli D *et al.* What's new in psychtoolbox-3. *Perception* 2007;**36**:1–16.

Koizumi A, Maniscalco B, Lau H. Does perceptual confidence facilitate cognitive control? *Atten Percept Psychophys* 2015;**77**:1295–306.

Lamme VAF. Why visual attention and awareness are different. *Trends Cogn Sci* 2003;**7**:12–8.

Lau H, Passingham RE. Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc Natl Acad Sci* 2006;**103**:18763–8.

Lau H, Rosenthal D. Empirical support for higher-order theories of conscious awareness. *Trends Cogn Sci* 2011;**15**:365–73.

Lee MD. BayesSDT: software for Bayesian inference with signal detection theory. *Behav Res Methods* 2008;**40**:450–6.

Maniscalco B, Lau H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious Cogn* 2012;**21**:422–30.

Maniscalco B, Peters MAK, Lau H. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Atten Percept Psychophys* 2016;**78**:1–15.

Meyniel F, Sigman M, Mainen ZF. Confidence as Bayesian probability: from neural origins to behavior. *Neuron* 2015;**88**:78–92.

Miyoshi K, Kuwahara A, Kawaguchi J. Comparing the confidence calculation rules for forced-choice recognition memory: a winner-takes-all rule wins. *J Mem Lang* 2018;**102**:142–54.

Miyoshi K, Lau H. A decision-congruent heuristic gives superior metacognitive sensitivity under realistic variance assumptions. *Psychol Rev* 2020;**127**:655–71.

Morey RD. Confidence intervals from normalized data: a correction to cousineau (2005). *Tutor Quant Methods Psychol* 2008;**4**:61–4.

Navajas J, Bahrami B, Latham PE. Post-decisional accounts of biases in confidence. *Curr Opin Behav Sci* 2016;**11**:55–60.

Odegaard B, Grimaldi P, Cho SH *et al.* Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proc Natl Acad Sci* 2018;**115**:201711628.

Overgaard M. Visual experience and blindsight: a methodological review. *Exp Brain Res* 2011;**209**:473–9.

Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 1997;**10**:437–42.

Pereira M, Faivre N, Iturrate I *et al.* Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proc Natl Acad Sci* 2020;**117**:8382–90.

Peters MAK, Thesen T, Ko YD *et al.* Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat Hum Behav* 2017;**1**. 10.1038/s41562-017-0139.

Pleskac TJ, Busemeyer JR. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol Rev* 2010;**117**:864–901.

Rausch M, Hellmann S, Zehetleitner M. Confidence in masked orientation judgments is informed by both evidence and visibility. *Atten Percept Psychophys* 2017;**80**:1–21.

Rausch M, Zehetleitner M. The folded X-pattern is not necessarily a statistical signature of decision confidence. *PLoS Comput Biol* 2019;**15**:e1007456.

Samaha J. How best to study the function of consciousness? *Front Psychol* 2015;**6**. 10.3389/fpsyg.2015.00604.

Samaha J, Barrett JJ, Sheldon AD *et al.* Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Front Psychol* 2016;**7**. 10.3389/fpsyg.2016.00851.

Samaha J, Iemi L, Haegens S *et al.* Spontaneous brain oscillations and perceptual decision-making. *Trends Cogn Sci* 2020;**24**:639–53.

Samaha J, Iemi L, Postle BR. Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. *Conscious Cogn* 2017;**54**:47–55.

Samaha J, Postle BR. Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proc Royal Soc B* 2017;**284**:20172035.

Samaha J, Switzky M, Postle BR. Confidence boosts serial dependence in orientation estimation. *J Vis* 2019;**19**:25.

Sanders JI, Hangya B, Kepecs A. Signatures of a statistical computation in the human sense of confidence. *Neuron* 2016;**90**:499–506.

Song C, Kanai R, Fleming SM *et al.* Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Conscious Cogn* 2011;**20**:1787–92.

Vandenbroucke ARE, Sligte IG, Barrett AB *et al.* Accurate metacognition for visual sensory memory representations. *Psychol Sci* 2014;**25**:861–73.

Weiskrantz L. *Blindsight: A Case Study and Implications*. Oxford, UK: Oxford University Press, 1986.

Yu S, Pleskac TJ, Zeigenfuse MD. Dynamics of postdecisional processing of confidence. *J Exp Psychol Gen* 2015;**144**:489.

Zawadzka K, Higham PA, Hanczakowski M. Confidence in forced-choice recognition: what underlies the ratings? *J Exp Psychol Learn Mem Cogn* 2017;**43**:552–64.

Zylberberg A, Barttfeld P, Sigman M. The construction of confidence in a perceptual decision. *Front Integr Neurosci* 2012;**6**. 10.3389/fnint.2012.00079.

Zylberberg A, Fetsch CR, Shadlen MN. The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *ELife* 2016;**5**:e17688.

Zylberberg A, Roelfsema PR, Sigman M. Variance misperception explains illusions of confidence in simple perceptual decisions. *Conscious Cogn* 2014;**27**:246–53.