

Metacognitive asymmetries in visual perception

Matan Mazor¹, Rani Moran^{1,2}, & Stephen M. Fleming^{1,2,3}

¹ Wellcome Centre for Human Neuroimaging, UCL

² Max Planck UCL Centre for Computational Psychiatry and Ageing Research

³ Department of Experimental Psychology, UCL

Author note

Correspondence concerning this article should be addressed to Matan Mazor, 12 Queen Square, London WC1N 3BG. E-mail: mtnmzor@gmail.com

Abstract

People have better metacognitive sensitivity for decisions about the presence compared to the absence of objects. However, it is not only objects themselves that can be present or absent, but also parts of objects and other visual features. Asymmetries in visual search indicate that a disadvantage for representing absence may operate at these levels as well. Furthermore, a processing advantage for surprising signals suggests that a presence/absence asymmetry may be explained by absence being passively represented as a default state, and presence as a default-violating surprise. It is unknown whether metacognitive asymmetry for judgements about presence and absence extend to these different levels of representation (object, feature, and default-violation). To address this question and test for a link between the representation of absence and default reasoning more generally, here we measured metacognitive sensitivity for discrimination judgments between stimuli that are identical except for the presence or absence of a distinguishing feature, and for stimuli that differ in their compliance with an expected default state. We find that the presence of local and global stimulus features gives rise to faster, more confident responses, but contrary to our hypothesis, has no effect on metacognitive sensitivity. In contrast, an additional post-hoc experiment confirmed robust asymmetries in metacognitive sensitivity for a classical visual detection task (detecting a grating in noise). Our results weigh against our proposal of a link between the detection metacognitive asymmetry and default reasoning, and are instead consistent with a low-level visual origin of this phenomenon.

Keywords: absence, presence, metacognition

Word count: 9797

Metacognitive asymmetries in visual perception

1 Introduction

At any given moment, there are many more things that are not there than things that are there. As a result, and in order to efficiently represent the environment, perceptual and cognitive systems have evolved to represent presences, and absence is implicitly represented as a default state (Oaksford, 2002; Oaksford & Chater, 2001). One corollary of this is that presence can be inferred from bottom-up sensory signals, but absence is never explicitly represented in sensory channels and must instead be inferred based on top-down expectations about the likelihood of detecting a hypothetical signal, had it been present. Experiments on human subjects accordingly suggest that representing absence is more cognitively demanding than representing presence, even in simple perceptual tasks, as is evident in slower reactions to stimulus absence than stimulus presence in near-threshold visual detection (Mazor, Friston, & Fleming, 2020), in a general difficulty to form associations with absence (Newman, Wolff, & Hearst, 1980), and in the late acquisition of explicit representations of absence in development (e.g., Sainsbury, 1971; Coldren & Haaf, 2000; for a review on the representation of nothing see Hearst, 1991).

An overarching difficulty in representing absence may reflect the metacognitive nature of absence representations; to represent something as absent, one must assume that they would have detected it had it been present. In philosophical writings, this form of higher-order, metacognitive inference-about-absence is known as *argument from epistemic closure*, or *argument from self-knowledge* (*If it was true, I would have known it*; Walton, 1992; De Cornulier, 1988). Strikingly, quantitative measures of metacognitive insight are consistently found to be lower for decisions about absence than for decisions about presence. When asked to rate their subjective confidence

following near-threshold detection decisions, subjective confidence ratings following ‘target absent’ judgments are commonly lower, and less aligned with objective accuracy, than following ‘target present’ judgments (Fig. 1; Kanai, Walsh, & Tseng, 2010; Kellij, Fahrenfort, Lau, Peters, & Odegaard, 2018; Mazor et al., 2020; Meuwese, Loon, Lamme, & Fahrenfort, 2014).

Metacognitive asymmetries have not only been observed for judgments about the presence or absence of whole physical objects and stimuli, but also for the presence or absence of cognitive variables such as memory traces. For instance, in recognition memory, subjects typically show poor metacognitive sensitivity for judgments about the absence of memories (such as when judging that they haven’t seen a study item before; Higham, Perfect, & Bruno, 2009). Unlike the absence of a visual stimulus, the absence of a memory is not localized in space and does not correspond with a specific representation of ‘nothing’.

One way of conceptualizing these findings is that absence asymmetries emerge as a function of default reasoning – absences are considered the “default”, and information about perceptual or mnemonic presence is accumulated and tested against this default. For instance, an asymmetry may emerge in recognition memory because the presence of memories is actively represented, and the absence of memories is assumed as the default unless evidence is available for the contrary. In the same way, other visual features that are not typically treated as presences or absences may still be coded relative to a default – assuming one state unless evidence is available for the contrary (e.g., assuming that a cookie is sweet rather than salty). However, whether a metacognitive asymmetry in processing presence and absence generalizes to these more abstract violations of default expectations remains unknown. Here we set out to map out the structure of absence representations by testing for metacognitive asymmetries in the presence and

absence of attributes at different levels of representation – from concrete objects, to visual features, to violations of default expectations.

Our choice of stimuli draws inspiration from visual search – a field where asymmetries are observed for a variety of stimulus types and features. In visual search, participants typically take longer to search for a target that is marked by the absence of a distinguishing feature, as compared to searching for a target that is marked by the presence of a feature relative to distractors (Treisman & Gormican, 1988; Treisman & Souther, 1985). Interestingly, *search asymmetries* have been demonstrated not only for the absence or presence of concrete physical features, but also for the presence or absence of deviations from a more abstract default state, which can be based on experience, culture, and contextual expectations (see methods; Von Grünau & Dubé, 1994; Frith, 1974; Gandolfo & Downing, 2020; Wang, Cavanagh, & Green, 1994). Of special interest for our study are these latter symmetries due to expectation violations, and their relation with asymmetries induced by the presence or absence of local and global features. Observing a metacognitive asymmetry for expectation violations as well as for the presence and absence of objects features would support a strong link between the representation of absence and default reasoning, where differences in metacognitive sensitivity reflect differences in the processing of information that agrees or contrasts with the expected default state.

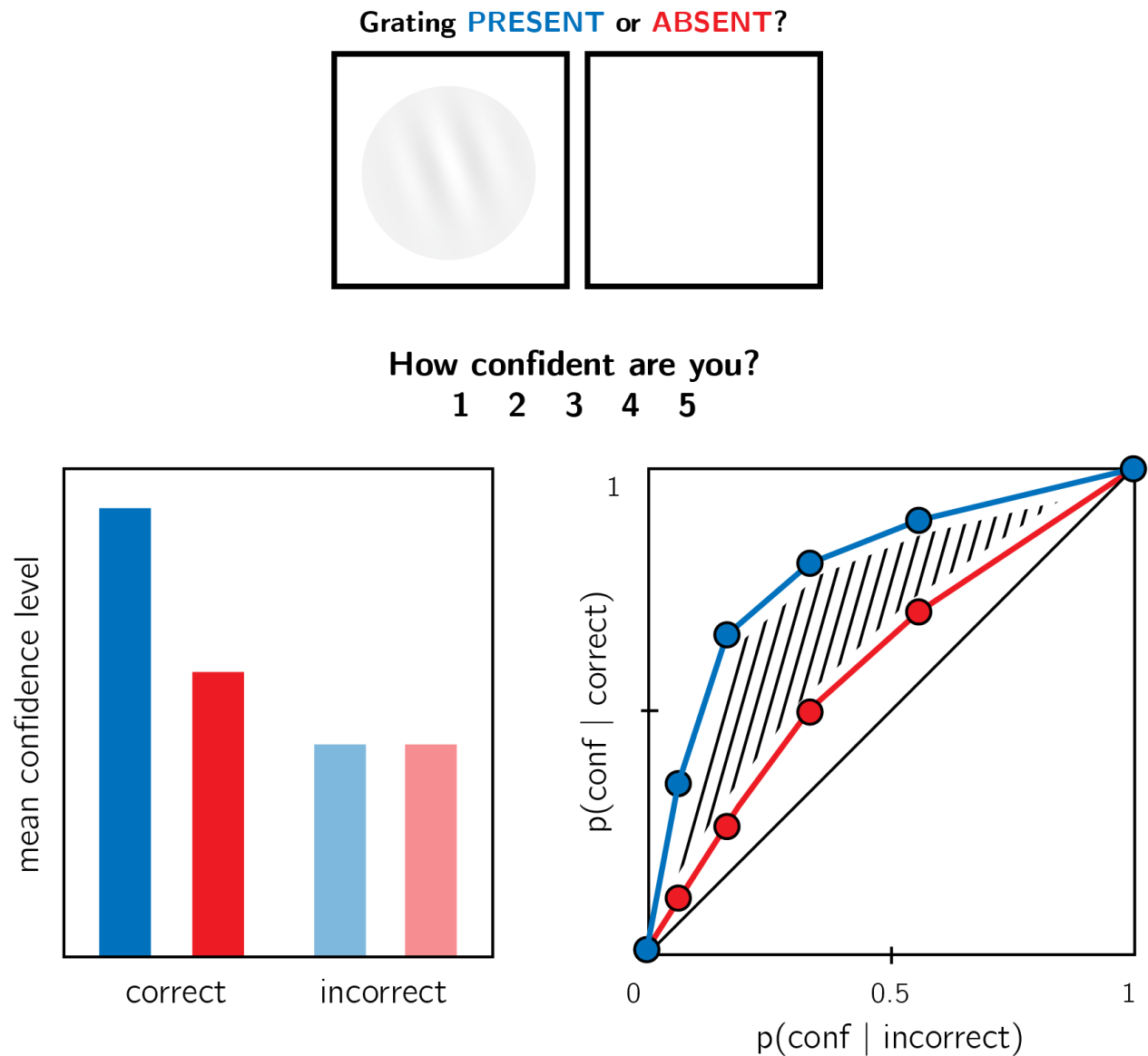


Figure 1. In visual detection, subjective confidence ratings following judgments about target absence are typically lower, and less correlated with objective accuracy than following judgments about target presence. Top panel: a typical detection experiment. The participant reports whether a visual grating was present or absent, and then rates their subjective decision confidence. Bottom left: typically, mean confidence in ‘yes’ responses (blue) is higher than in ‘no’ responses (red). This effect is much more pronounced in correct trials. Bottom right: the interaction between accuracy and response type on confidence (metacognitive asymmetry) manifests as a lower area

under the response-conditional ROC curve for ‘no’ responses compared with ‘yes’ responses. Plots do not directly correspond to a specific dataset, but portray typical results in visual detection.

While traditional accounts interpreted visual search asymmetries as reflecting a qualitative advantage for the cognitive representation of presence (affording a parallel search in the case of feature-present search only; Treisman & Gormican, 1988), other models attribute the asymmetry to differences in the distributions of perceptual signals already at the sensory level (Doshier, Han, & Lu, 2004; Vincent, 2011). Similarly, in the case of metacognitive asymmetries, the idea that decisions about absence are qualitatively different from decisions about presence has been challenged by an excellent fit of simple models that assume unequal variance for the signal-present and signal-absent sensory distributions, a model that does not assume any qualitative difference between the two decisions (Kellij et al., 2018). Deciding between these model families is beyond the scope of this project. However, identifying metacognitive asymmetries for abstract cognitive variables such as familiarity could help refine these models, for instance by revealing that representing deviations from a default state is an overarching principle of cognitive organization, one that goes beyond specific features of visual perception.

2 Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

We ran six experiments, that were identical except for the identity of the two stimuli S_1 and S_2 (and of the stimulus used for backward masking; see section 2.6 for details). Our choice of stimuli for this study was based on the visual search literature. For some stimulus pairs S_1 and S_2 ,

searching for one S_1 among multiple S_2 s is more efficient than searching for one S_2 among multiple S_1 s. Such *search asymmetries* have been reported for stimulus pairs that are identical except for the presence and absence of a distinguishing feature. Importantly, distinguishing features vary in their level of abstraction, from concrete *local features* (finding a Q among Os is easier than the inverse search; Treisman & Souther, 1985), through *global features* (finding a curved line among straight lines is easier than the inverse search; Treisman & Gormican, 1988), and up to the presence or absence of abstract *expectation violations* (searching for an upward-tilted cube among downward-tilted cubes is easier than the inverse search, in line with a general expectation to see objects on the ground rather than floating in space; Von Grünau & Dubé, 1994). We treat these three types of asymmetries as reflecting a default-reasoning mode of representation, where the absence of features and/or the adherence of objects to prior expectations is tentatively accepted as a default by the visual system, unless evidence is available for the contrary (Treisman & Gormican, 1988; Treisman & Souther, 1985). In this study, we test for metacognitive asymmetries for two stimulus features in each category, in six separate experiments with different participants (Fig. 2). For each of the following stimulus pairs, searching for S_1 among multiple instances of S_2 has been found to be more efficient than the inverse search:

- **Local feature: Addition of a stimulus part.** Q and O were used as S_1 and S_2 , respectively (Treisman & Souther, 1985).
- **Local feature: Open ends.** C and O were used as S_1 and S_2 , respectively (Takeda & Yagi, 2000; Treisman & Gormican, 1988; Treisman & Souther, 1985).
- **Global feature: Orientation.** Tilted and vertical lines were used S_1 and S_2 , respectively (Treisman & Gormican, 1988).
- **Global feature: Curvature.** Curved and straight lines were used as S_1 and S_2 , respectively (Treisman & Gormican, 1988).
- **Expectation violation: Viewing angle.** Upward and Downward tilted cubes were used as S_1 and S_2 , respectively (Von Grünau & Dubé, 1994).

- **Expectation violation: Letter inversion.** Flipped and normal N were used as S_1 and S_2 , respectively (Frith, 1974; Wang et al., 1994).

The experiments quantified participants' metacognitive sensitivity for discrimination judgments between S_1 and S_2 .

2.1 Participants

The research complied with all relevant ethical regulations, and was approved by the Research Ethics Committee of University College London (study ID number 1260/003). Participants were recruited via Prolific, and gave informed consent prior to their participation. They were selected based on their acceptance rate (>95%) and for being native English speakers. For each of the six experiments, we aimed to collect data until we reached 106 included participants (after applying our pre-registered exclusion criteria). The entire experiment took 10-15 minutes to complete. Participants were paid between £1.25 to £2 for their participation, maintaining a median hourly wage of £7 or higher.

2.2 Procedure

Experiments were programmed using the jsPsych and P5 JavaScript packages (De Leeuw, 2015; McCarthy, 2015), and were hosted on a JATOS server (???)

After instructions, a practice phase, and a multiple-choice comprehension check, the main part of the experiment started. It comprised 96 trials separated into 6 blocks. Only the last 5 blocks were analyzed.

On each trial, participants made discrimination judgments on masked stimuli, and rated their subjective decision confidence on a continuous scale. After a fixation cross (500 ms), the target stimulus (S_1 or S_2) was presented in the center of the screen for 50 ms, followed by a mask

(100 ms). Stimulus onset asynchrony was calibrated online in a 1-up-2-down procedure (Levitt, 1971), with a multiplicative step factor of 0.9, and starting at 30 milliseconds. Participants then used their keyboard to make a discrimination judgment. Stimulus-key mapping was counterbalanced between participants. Following response, subjective confidence ratings were given on an analog scale by controlling the size of a colored circle with the computer mouse. High confidence was mapped to a big, blue circle, and low confidence to a small, red circle. We chose a continuous (rather than a more typical discrete) confidence scale in order to ensure sufficient variation in confidence ratings within the dynamic range of individual participants. This variation is useful for the extraction of response conditional ROC curves. The confidence rating phase terminated once participants clicked their mouse, but not before 2000 ms. No trial-specific feedback was delivered about accuracy. In order to keep participants motivated and engaged, block-wise feedback was delivered between experimental blocks about overall accuracy, mean confidence in correct responses, and mean confidence in incorrect responses.

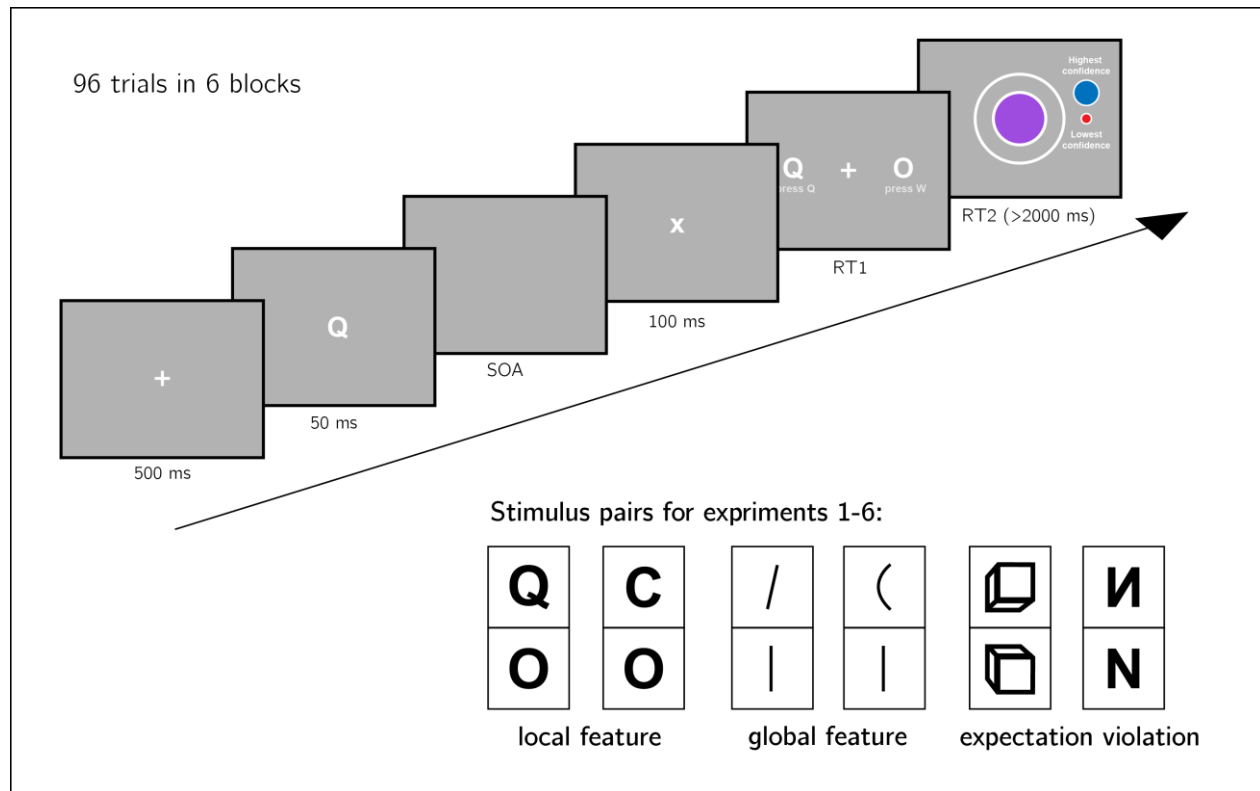


Figure 2. Experiment design. Metacognitive asymmetry effects were tested for six stimulus features in six separate experiments, encompassing three levels of abstraction: local features, global features, and expectation violations. The presented trial corresponds to the first stimulus pair, where *Q* and *O* are used as stimuli.

2.2.1 Randomization.

The order and timing of experimental events was determined pseudo-randomly by the Mersenne Twister pseudorandom number generator, initialized in a way that ensures registration time-locking (Mazor, Mazor, & Mukamel, 2019).

2.3 Data analysis

We used R (Version 3.6.0; R Core Team, 2019) and the R-packages *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2018), *broom* (Version 0.5.6; Robinson & Hayes, 2020), *cowplot* (Version 1.0.0; Wilke, 2019), *dplyr* (Version 1.0.4; Wickham et al., 2020), *ggplot2* (Version

3.3.1; Wickham, 2016), *lmerTest* (Version 3.1.2; Kuznetsova, Brockhoff, & Christensen, 2017), *lsr* (Version 0.5; Navarro, 2015), *MESS* (Version 0.5.6; Ekstrøm, 2019), *papaja* (Version 0.1.0.9997; Aust & Barth, 2020), *pracma* (Version 2.2.9; Borchers, 2019), *pwr* (Version 1.3.0; Champely, 2020), and *tidyr* (Version 1.1.0; Wickham & Henry, 2020) for all our analyses.

For each of the six stimulus pairs $[S_1, S_2]$, we tested the following hypotheses:

- **Hypothesis 1:** Subjective confidence is higher for S_1 responses than for S_2 responses.

For each of the six stimulus pairs, we tested the null hypothesis that subjective confidence for S_1 responses is equal to or lower than subjective confidence for the feature-absent stimulus ($H_o: conf_{S_1} \leq Conf_{S_2}$).

2. **Hypothesis 2:** Metacognitive sensitivity, measured as the area under the response conditional ROC curves, is higher for S_1 responses than for S_2 responses.

For each of the six stimulus pairs, we tested the null hypothesis that metacognitive sensitivity for S_1 responses is equal to or lower than metacognitive sensitivity for the S_2 responses ($H_o: auROC_{S_1} \leq auROC_{S_2}$).

3. **Hypothesis 3:** Metacognitive sensitivity, measured as the area under the response conditional ROC curves, is higher for S_1 responses than for S_2 responses, to a greater extent than expected from an equivalent equal-variance SDT model.

For each of the six stimulus pairs, we tested the null hypothesis that difference between metacognitive sensitivities for S_1 and S_2 responses is lower than the difference expected from an equivalent equal-variance SDT model ($H_o: (auROC_{S_1} - auROC_{S_2}) \leq (\widehat{auROC}_{S_1} - \widehat{auROC}_{S_2})$) where \widehat{auROC} is the expected auROC under an equal variance SDT model with equal sensitivity, criterion, and distribution of confidence ratings in incorrect responses).

4. **Hypothesis 4:** S_1 responses are faster on average than S_2 responses.

For each of the six stimulus pairs, we tested the null hypothesis that log-transformed response times for S_1 responses are equal to or higher than log-transformed response times for S_2 responses ($H_o: \log(RT_{S_1}) \geq \log(RT_{S_2})$).

Hypotheses 1 and 2 correspond to the effects of stimulus type on metacognitive bias and metacognitive sensitivity, respectively. Although these two measures are theoretically independent, both bias and sensitivity are found to vary between detection ‘yes’ and ‘no’ responses.

Based on pilot data and previous experiments examining near-threshold perceptual detection and discrimination, we did not expect a response bias (such that the probability of responding S_1 is significantly different from 0.5 across participants). However, such a response bias, if found, may bias metacognitive asymmetry estimates as measured with response-conditional ROC curves. Hypothesis 3 was designed to confirm that metacognitive asymmetry is higher than that expected from an equivalent equal-variance SDT model with the same response bias, sensitivity, and distribution of confidence ratings in incorrect responses as in the actual data. We interpreted conflicting results for Hypotheses 2 and 3 as evidence for a metacognitive asymmetry that is driven or masked by a response bias.

Hypothesis 4 is motivated by two observations from previous studies. First, detection ‘yes’ responses are faster than detection ‘no’ responses (Mazor et al., 2020). And second, when participants are not under strict time pressure, reaction time inversely scales with confidence (Calder-Travis, Charles, Bogacz, & Yeung, 2020; Henmon, 1911; Moran, Teodorescu, & Usher, 2015; Pleskac & Busemeyer, 2010). Based on these findings, if S_1 and S_2 responses are similar to detection ‘yes’ and ‘no’ responses not only in explicit confidence judgments, but also in response times, we should also expect a response time difference for these stimulus pairs.

2.3.1 Dependent variables and analysis plan. Response conditional ROC

curves were extracted by plotting the empirical cumulative distribution of confidence ratings for correct responses against the same cumulative distribution for incorrect responses. This was done separately for the two responses S_1 and S_2 , resulting in two curves. The area under the response-conditional ROC curve is a measure of metacognitive sensitivity (Fleming & Lau, 2014). The difference between the areas for the two responses is a measure of metacognitive asymmetry (Meuwese et al., 2014). This difference was used to test Hypothesis 2.

In order to test hypothesis 3, SDT-derived response-conditional ROC curves were plotted in the following way. For each response, we plotted the empirical cumulative distribution for incorrect responses on the x axis against the cumulative distribution for correct responses that would be expected in an equal-variance SDT model with matching sensitivity and response bias on the y axis. The difference between the areas of these theoretically derived response-conditional ROC curves was compared against the difference between the true response-conditional ROC curves.

For visualization purposes only, confidence ratings were divided into 20 bins, tailored for each participant to cover their dynamic range of confidence ratings.

For each of the six experiments, Hypotheses 1-4 were tested using a one tailed t-test at the group level with $\alpha = 0.05$. The summary statistic at the single subject level was difference in mean confidence between S_1 and S_2 responses for Hypothesis 1, difference in area under the response-conditional ROC curve between S_1 and S_2 responses (ΔAUC) for Hypothesis 2, difference in ΔAUC between true confidence distributions and SDT-derived confidence distributions for hypothesis 3, and difference in mean log response time between S_1 and S_2 responses for Hypothesis 4.

In addition, a Bayes factor was computed using the BayesFactor R package (Morey, Rouder, Jamil, & Morey, 2015) and using a Jeffrey-Zellner-Siow (Cauchy) Prior with an rscale parameter of 0.65, representative of the similar standardized effect sizes we observe for Hypotheses 1-4 in our pilot data.

We based our inference on the resulting Bayes Factors.

2.3.2 Statistical power. Statistical power calculations were performed using the R-pwr packages pwr (Champely, 2020) and PowerTOST (Labes, Schütz, Lang, & Labes, 2020).

- Hypothesis 1 (MEAN CONFIDENCE): With 106 participants, we had statistical power of 95% to detect effects of size 0.32, which is less than the standardized effect size we observed for confidence in our pilot sample ($d = 0.66$).
- Hypothesis 2 (METACOGNITIVE ASYMMETRY): With 106 participants, we had statistical power of 95% to detect effects of size 0.32, which is less than the standardized effect size we observed for metacognitive sensitivity in our pilot sample ($d = 0.73$).
- Hypothesis 3 (METACOGNITIVE ASYMMETRY: CONTROL): With 106 participants, we had statistical power of 95% to detect effects of size 0.32, which is less than the standardized effect size we observed for metacognitive sensitivity, controlling for response bias, in our pilot sample ($d = 0.81$).
- Hypothesis 4 (RESPONSE TIME): With 106 participants, we had statistical power of 95% to detect effects of size 0.32, which is less than the standardized effect size we observed for response time in our pilot sample ($d = 0.61$).

Finally, in case that the true effect size equals 0, a Bayes Factor with our chosen prior for the alternative hypothesis will support the null in 95 out of 100 repetitions, and will support the null with a BF_{01} higher than 3 in 79 out of 100 repetitions. In a case where the true effect size is sampled from a Cauchy distribution with a scale factor of 0.65, a Bayes Factor with our chosen prior for the alternative hypothesis will support the alternative hypothesis in 76 out of 100 repetitions, support the alternative hypothesis with a BF_{10} higher than 3 in 70 out of 100 repetitions, and support the null hypothesis with a BF_{01} higher than 3 in 15 out of 100 hypotheses (based on an adaptation of simulation code from Lakens, 2016).

2.3.3 Rejection criteria. Participants were excluded for performing below 60% accuracy, for having extremely fast or slow reaction times (below 250 milliseconds or above 5 seconds in more than 25% of the trials), and for failing the comprehension check. Finally, for type-2 ROC curves to be generated, some responses must be incorrect. and for them to be informing some variability in confidence ratings is necessary. Thus, participants who committed less than two of each error type (for example, mistaking a *Q* of *O* and mistaking an *O* for *Q*), or that reported less than two different confidence levels for each of the two responses were excluded from all analyses.

Trials with response time below 250 milliseconds or above 5 seconds were excluded.

2.4 Data availability

All raw data is fully available on OSF and on the study's GitHub repository:

<https://github.com/matanmazor/asymmetry>.

2.5 Code availability

All analysis code is openly shared on the study's GitHub repository:

<https://github.com/matanmazor/asymmetry>. For complete reproducibility, the RMarkdown file used to generate the final version of the manuscript, including the generation of all figures and extraction of all test statistics, is also available on our GitHub repository.

2.6 Deviations from pre-registration

We aimed to follow our pre-specified methods as close as possible (the approved Stage 1 protocol is available in the following link: doi.org/10.31234/osf.io/av9ns). The following is an exhaustive list of deviations from our pre-registered plan.

- *Stimulus used for backward masking:* We planned to use the same stimulus (the letter Z) for backward masking in all six experiments. This mask was effective in Experiments 1 and 2, but in Experiment 3 overly high accuracy levels indicated that for these stimuli the mask was not salient enough. For a subset of participants in Exp. 3, an overlay of all 10 stimuli from all 6 experiments (*Q*, *O*, *C*, vertical, tilted, and curved lines, upward-tilted and downward-tilted cubes, and normal and flipped Ns) was used. For the remaining participants and experiments, we used four dollar signs as our mask.
- *Rejection criteria:* In our pre-registration we explain that informative response-conditional ROC curves can only be generated if participants make errors. When analyzing the data we came to realize that an additional prerequisite for response-conditional ROC curves to be informative is that the variance in confidence ratings is higher than zero, otherwise the curve is diagonal. We therefore required that participants report at least two different confidence

levels for each response. Participants that did not meet this additional criterion were excluded from all analyses.

- *Monetary compensation:* For some of the experiments, we noticed that participants completed the experiment more quickly than what we had originally estimated. We therefore reduced our offered payment for some of the experiments, while maintaining a median hourly wage of £7 or higher.

3 Results

A summary of the results from all six experiments is available in section 3.7 and in Figures 3, 4 and 5.

3.1 Experiment 1: *Q* vs. *O*

In Experiment 1, we examined discrimination judgments between the two letters *Q* and *O*. Based on a search asymmetry for these letters (*Q*s are found faster than *O*s than vice-versa; Treisman & Souther, 1985), we hypothesized that a similar asymmetry would emerge in subjective confidence judgments, such that metacognitive sensitivity for *Q* responses will be higher than for *O* responses. We used the letter *Z* as our backward mask.

205 participants were recruited from Prolific for Experiment 1.

Median completion time was 13.12 minutes. Mean accuracy was 0.74. Participants reported seeing an *O* on 0.47 of the trials. Overall we excluded 71 participants based on our exclusion criteria, leaving 134 participants for the main analysis. Due to a technical error in data collection, this figure is higher than that specified in our preregistration document (N=106). Going forward, only data from included participants is analyzed.

Mean accuracy among the included participants was $M = 0.74$, 95% CI [0.73, 0.75]. Mean SOA in the last trial was $M = 47.50$, 95% CI [39.39, 55.61]. Participants showed no consistent bias in their responses (quantified as the probability of a ‘Q’ response minus 0.5; $M = 0.02$, 95% CI [0.00, 0.04]). On a scale of 0 to 1, mean confidence level was $M = 0.49$, 95% CI [0.45, 0.53]. Confidence was higher for correct than for incorrect responses ($M_d = 0.15$, 95% CI [0.13, 0.17], $t(133) = 14.85$, $p < .001$).

Hypothesis 1: In line with our hypothesis, confidence was generally higher for Q (feature present) responses than for O (feature absent) responses ($t(133) = 7.52$, $p < .001$; Cohen’s $d = 0.65$; $BF_{10} = 1.07 \times 10^9$; see Fig. 3, panel 1).

Hypothesis 2: In order to measure metacognitive asymmetry, we extracted the response-conditional type-2 ROC (rc-ROC) curves for the two responses (Q and O) in the discrimination task. This was done by plotting the cumulative distribution of confidence ratings (high to low) for correct responses against the same distribution for incorrect responses. The area under the rc-ROC curve (auROC) was then taken as a measure of metacognitive sensitivity (Kanai et al., 2010; Meuwese et al., 2014). In line with our hypothesis, auROC for Q responses ($M = 0.72$, 95% CI [0.70, 0.74]) was higher than for O responses ($M = 0.68$, 95% CI [0.66, 0.70]; $t(133) = 2.96$, $p = .002$; Cohen’s $d = 0.26$; $BF_{10} = 6.56$; see Figure 4, panel 1), similar to the documented metacognitive asymmetry for detection judgments.

Hypothesis 3: Metacognitive asymmetry was not significantly higher than what is expected based on an equal-variance SDT model with the same response bias and sensitivity as the subjects ($t(133) = 0.97$, $p = .167$; Cohen’s $d = 0.08$). A Bayes Factor indicated that our results are more likely under a model that assumes no additional metacognitive asymmetry ($BF_{01} = 6.07$).

Hypothesis 4: In line with our hypothesis, *Q* responses were faster on average than *O* responses by 37 ms. ($t(133) = -2.99, p = .002$; Cohen's $d = 0.26$; $BF_{10} = 7.05$; see Fig. 3, panel 1).

In summary, in Experiment 1 we found that *Q* responses were faster and accompanied by higher subjective confidence, in line with a processing advantage for feature-presence.

Metacognitive asymmetry however did not go beyond what is expected from an equal-variance SDT model for these stimuli, taking into account response biases.

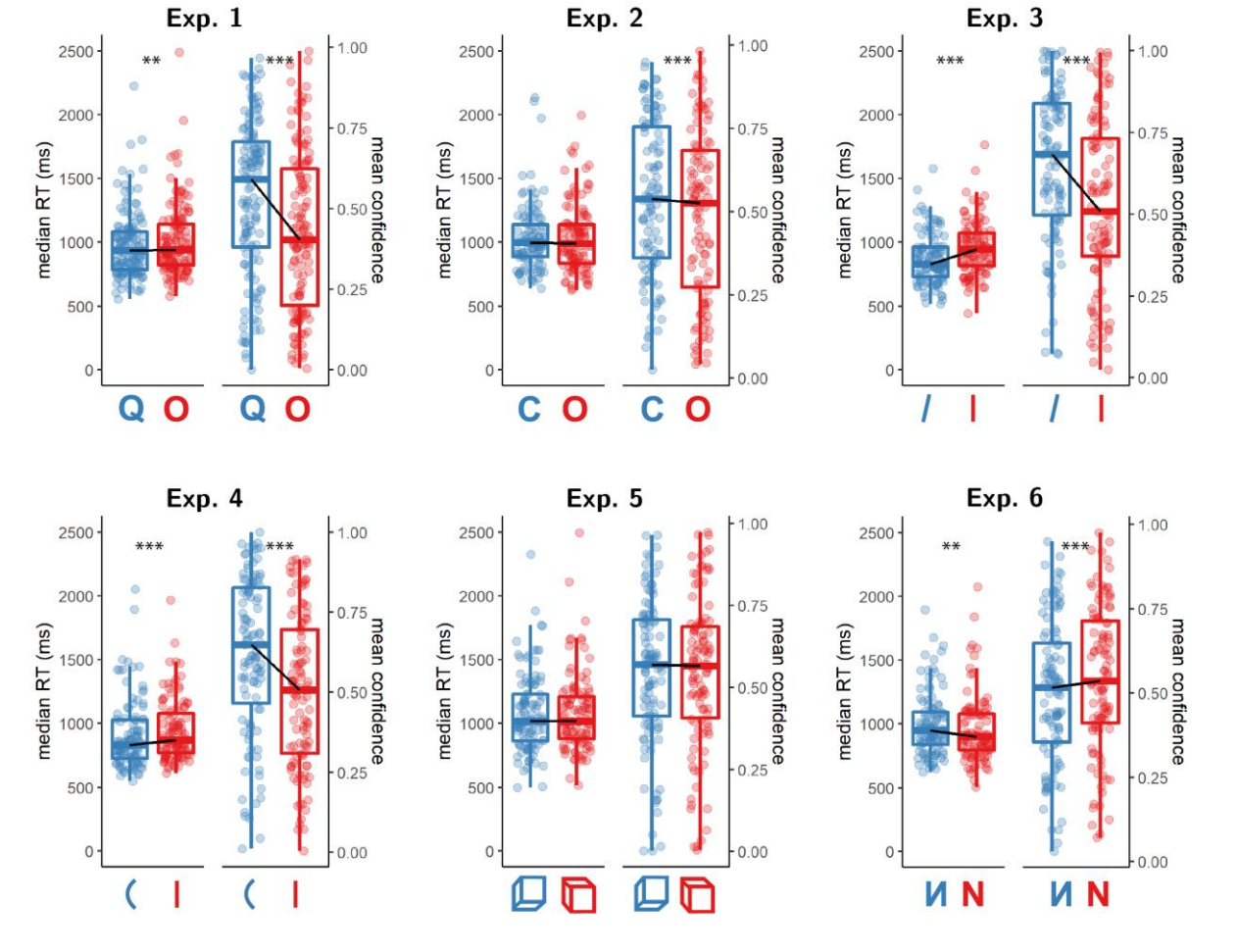


Figure 3. Reaction time and confidence distributions for Experiments 1-6. Box edges and central lines represent the 25, 50 and 75 quantiles. Whiskers cover data points within four inter-

quartile ranges around the median. Black lines connect the median values for the two responses.

Stars represent significance in a two-sided t-test: \bullet : $p < 0.01$, $*$: $p < 0.001$

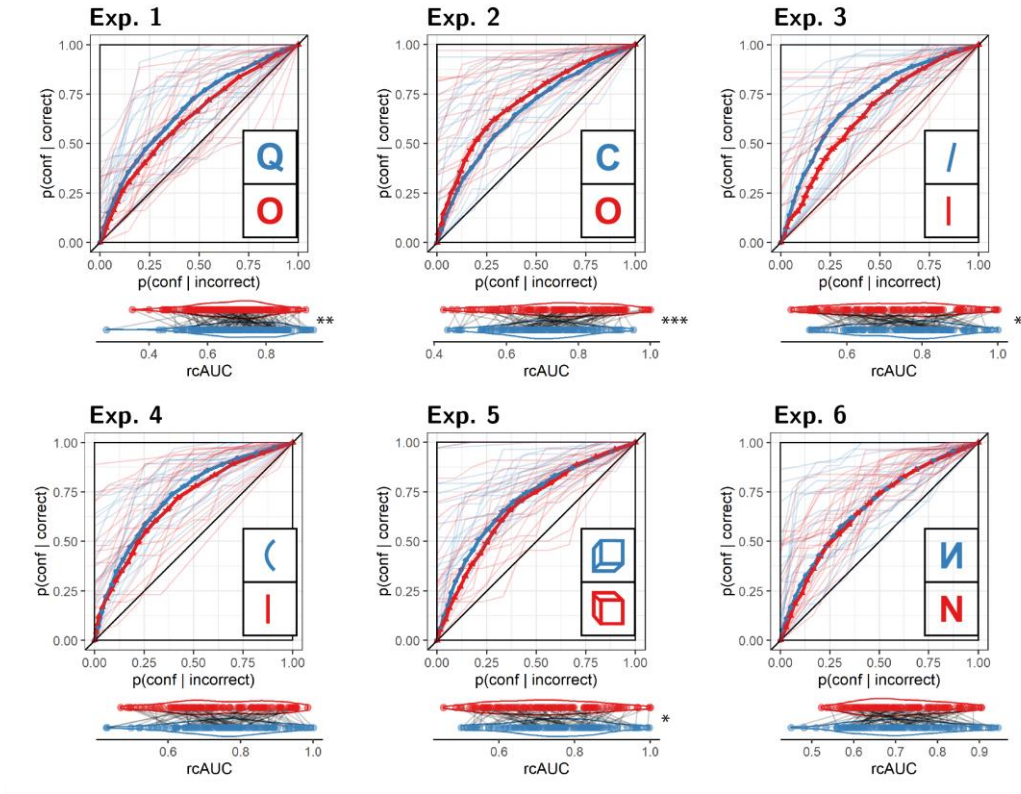


Figure 4. Response conditional ROC curves for Experiments 1-6. The area under the curve is a measure of metacognitive sensitivity. Error bars stand for the standard error of the mean. For illustration, the response-conditional ROC curves of the first 20 participants of each Experiment are plotted in low opacity. Below each ROC: distributions of the area under the curve for the two responses, across participants. Each line corresponds to a participant. Stars represent significance in a two-sided t-test: \bullet : $p < 0.05$, \bullet : $p < 0.01$, \bullet : $p < 0.001$

3.2 Experiment 2: C vs. O

In Experiment 2, we examined discrimination judgments between the two letters *C* and *O*. Based on a search asymmetry for these letters (*C*s are found faster among *O*s than vice versa;

Treisman & Souther, 1985; Takeda & Yagi, 2000; Treisman & Gormican, 1988), we hypothesized that a similar asymmetry would emerge in subjective confidence judgments, such that metacognitive sensitivity for perceiving a *C* will be higher than for perceiving an *O*. We used the letter *Z* as our backward mask.

143 participants were recruited from Prolific for Experiment 2.

Median completion time was 12.80 minutes. Mean accuracy was 0.75, and participants reported seeing an *O* on 0.43 of the trials. Overall we excluded 37 participants, leaving 106 participants for the main analysis. Going forward, only data from included participants is analyzed.

Mean accuracy among included participants was $M = 0.74$, 95% CI [0.73, 0.75]. The mean SOA of the last trial was $M = 40.18$, 95% CI [34.37, 46.00]. Participants showed a consistent bias toward reporting a *C* rather than an *O* ($M = 0.07$, 95% CI [0.05, 0.08]). On a scale of 0 to 1, mean confidence level was $M = 0.52$, 95% CI [0.48, 0.56]. Confidence was higher for correct than for incorrect responses ($M_d = 0.17$, 95% CI [0.15, 0.19], $t(105) = 15.05$, $p < .001$).

Hypothesis 1: In line with our hypothesis, confidence was generally higher for *C* (feature present) responses than for *O* (feature absent) responses ($M_d = 0.05$, 95% CI [0.03, ∞], $t(105) = 3.59$, $p < .001$; Cohen's $d = 0.35$; $BF_{10} = 42.62$; see Figure 4, panel 2).

Hypothesis 2: Opposite to our prediction, auROC for *C* responses ($M = 0.70$, 95% CI [0.68, 0.72]) was *lower* than for *O* responses ($M = 0.75$, 95% CI [0.73, 0.78]; $t(105) = -3.53$, $p > .999$; Cohen's $d = 0.34$; see Figure 4, panel 2.). Bayes Factor strongly supported the

alternative ($BF_{10} = 35.19$). Note that our prior on effect sizes was symmetric around zero, such that support for the alternative is obtained for negative, as well as positive effects.

Hypothesis 3: Metacognitive sensitivity for *C* responses was still significantly higher than for *O* responses after controlling for bias (Cohen's $d=0.49$; $BF_{10} = 6.46 \times 10^3$).

Hypothesis 4: Contrary to our hypothesis, response times for *C* and for *O* responses were highly similar, with a median difference of 6 ms. ($t(105) = 0.01$, $p = .504$; Cohen's $d = 0.00$; $BF_{01} = 8.57$; see Figure 4, panel 2).

In summary, in Experiment 2 we found a dissociation between our two confidence-related measures. As we hypothesized, participants were generally more confident in their *C* (feature present) responses, but their metacognitive sensitivity was higher following *O* (feature absent) responses. We found no reliable difference in response times between these two responses.

3.3 Experiment 3: tilted vs. vertical lines

In Experiment 3, we examined discrimination judgments between tilted and vertical lines. Based on a search asymmetry for these stimuli (tilted lines are found faster among vertical lines than vice versa; Treisman & Gormican, 1988), we hypothesized that a similar asymmetry would emerge in subjective confidence judgments, such that metacognitive sensitivity for perceiving a tilted line will be higher than for perceiving a vertical line. As described in section 2.6, overly high accuracy in the first few participants led us to change our masking stimulus, first to an overlay of all stimuli and then to four dollar signs. We present here the combined results from these last two cohorts of participants. The results were qualitatively similar in the two cohorts.

304 participants were recruited from Prolific for Experiment 3.

Median completion time was 12.43 minutes. Mean accuracy was 0.86, and participants reported seeing a vertical line on 0.44 of the trials. Overall we excluded 198 participants, leaving 106 participants for the main analysis. Going forward, only data from included participants is analyzed.

Mean accuracy among included participants was $M = 0.79$, 95% CI [0.78, 0.81]. The mean SOA of the last trial was $M = 30.83$, 95% CI [25.99, 35.68]. Participants showed a consistent bias toward reporting a tilted rather than a vertical line ($M = 0.06$, 95% CI [0.04, 0.08]). On a scale of 0 to 1, mean confidence level was $M = 0.61$, 95% CI [0.56, 0.65]. Confidence was higher for correct than for incorrect responses ($M_d = 0.18$, 95% CI [0.15, 0.20], $t(105) = 13.42$, $p < .001$).

Hypothesis 1: In line with our hypothesis, confidence was generally higher for tilted lines (feature present) responses than for vertical lines (feature absent) responses ($M_d = 0.12$, 95% CI [0.09, ∞], $t(105) = 7.18$, $p < .001$; Cohen's $d = 0.70$; $BF_{10} = 8.89 \times 10^7$; see Figure 4, panel 3).

Hypothesis 2: Contrary to our prediction, auROC for reports of seeing a tilted line ($M = 0.76$, 95% CI [0.74, 0.78]) was similar to auROC for reports of seeing a vertical line according to Bayes Factor analysis ($M = 0.73$, 95% CI [0.70, 0.75]; Cohen's $d = 0.18$; $BF_{01} = 1.59$; see Figure 4, panel 3.). This difference in metacognitive sensitivity was however significant in a standard t-test ($t(105) = 1.88$, $p = .031$).

Hypothesis 3: (The lack of) metacognitive asymmetry was not different from what would be expected based on an equal-variance SDT model with the same response bias and sensitivity ($t(105) = -0.70$, $p = .759$; Cohen's $d = 0.07$; $BF_{01} = 6.74$).

Hypothesis 4: In line with our hypothesis, response times for ‘tilted’ responses were significantly faster than response times for ‘vertical’ responses, with a median difference of 68 ms. ($t(105) = -5.82, p < .001$; Cohen’s $d = 0.56$; $BF_{10} = 1.83 \times 10^5$; see Figure 4, panel 3).

In summary, in Experiment 3 we found that ‘tilted’ (feature present) responses were faster and accompanied by higher subjective confidence than ‘vertical’ (feature absent) responses, in the absence of a difference in metacognitive sensitivity between the two responses.

3.4 Experiment 4: curved vs. straight lines

In Experiment 4, we examined discrimination judgments between curved and vertical lines. Based on a search asymmetry for these stimuli (curved lines are found faster among vertical lines than vice versa; Treisman & Gormican, 1988), we hypothesized that a similar asymmetry would emerge in subjective confidence judgments, such that metacognitive sensitivity for perceiving a tilted line will be higher than for perceiving a vertical line. We used four dollar signs (\$\$\$\$) as our mask.

211 participants were recruited from Prolific for Experiment 4.

Median completion time was 12.08 minutes. Mean accuracy was 0.84, and participants reported seeing a straight line on 0.44 of the trials. Overall we excluded 104 participants, leaving 107 participants for the main analysis. Going forward, only data from included participants is analyzed.

Mean accuracy among included participants was $M = 0.79$, 95% CI [0.77, 0.80]. The mean SOA of the last trial was $M = 28.01$, 95% CI [24.22, 31.79]. Participants showed a consistent bias toward reporting a curved rather than a vertical line ($M = 0.06$, 95% CI [0.04,

0.07]). On a scale of 0 to 1, mean confidence level was $M = 0.57$, 95% CI [0.53, 0.61].

Confidence was higher for correct than for incorrect responses ($M_d = 0.21$, 95% CI [0.18, 0.24], $t(106) = 14.96$, $p < .001$).

Hypothesis 1: In line with our hypothesis, confidence was generally higher for curved lines (feature present) responses than for straight lines (feature absent) responses ($M_d = 0.12$, 95% CI [0.09, ∞], $t(106) = 8.25$, $p < .001$; Cohen's $d = 0.80$; $BF_{10} = 1.61 \times 10^{10}$; see Figure 4, panel 4).

Hypothesis 2: Contrary to our prediction, auROC for reports of seeing a curved line ($M = 0.76$, 95% CI [0.73, 0.78]) was similar to auROC for reports of seeing a straight line ($M = 0.75$, 95% CI [0.73, 0.78]; $t(106) = 0.30$, $p = .382$; Cohen's $d = 0.03$; $BF_{01} = 8.23$; see Figure 4, panel 4.).

Hypothesis 3: (The lack of) metacognitive asymmetry was not different from what would be expected based on an equal-variance SDT model with the same response bias and sensitivity ($t(106) = -1.93$, $p = .972$; Cohen's $d = 0.19$; $BF_{01} = 1.45$).

Hypothesis 4: In line with our hypothesis, response times for 'curved' responses were significantly faster than response times for 'straight' responses, with a median difference of 51 ms ($t(106) = -4.36$, $p < .001$; Cohen's $d = 0.42$; $BF_{10} = 558.55$; see Figure 4, panel 4).

In summary, similar to Experiment 3, 'curved' (feature-present) responses were faster and accompanied by higher subjective confidence than 'straight' (feature absent) responses. However, similar to the results of Experiment 3, here also We did not find a metacognitive asymmetry for these stimuli.

3.5 Experiment 5: upward-tilted vs. downward-tilted cubes

In Experiment 5, we examined discrimination judgments between upward-tilted and downward-tilted cubes. Based on a search asymmetry for these stimuli (upward-tilted cubes are found faster among downward-tilted cubes than vice versa, in line with an expectation to see objects on the ground and not floating in space; Von Grünau & Dubé, 1994), we hypothesized that a similar asymmetry would emerge in subjective confidence judgments, such that metacognitive sensitivity for perceiving an upward-tilted cube will be higher than for perceiving a downward-tilted cube. We used four dollar signs (\$\$\$\$) as our mask.

162 participants were recruited from Prolific for Experiment 5.

Median completion time was 13.30 minutes. Mean accuracy was 0.79, and participants reported seeing a downward-tilted cube on 0.51 of the trials. Overall we excluded 56 participants, leaving 106 participants for the main analysis. Going forward, only data from included participants is analyzed.

Mean accuracy among included participants was $M = 0.77$, 95% CI [0.76, 0.78]. The mean SOA of the last trial was $M = 29.51$, 95% CI [23.20, 35.81]. Participants showed no consistent response bias ($M = -0.01$, 95% CI [-0.03, 0.00]). On a scale of 0 to 1, mean confidence level was $M = 0.55$, 95% CI [0.51, 0.59]. Confidence was higher for correct than for incorrect responses ($M_d = 0.23$, 95% CI [0.20, 0.26], $t(105) = 13.89$, $p < .001$).

Hypothesis 1: Contrary to our hypothesis, confidence was similar for upward-tilted (feature present) responses and downward-tilted (feature absent) responses ($M_d = 0.00$, 95% CI [-0.02, ∞], $t(105) = 0.12$, $p = .452$; Cohen's $d = 0.01$; $BF_{01} = 8.51$; see Figure 4, panel 5).

Hypothesis 2: Contrary to our hypothesis, auROC for reports of seeing an upward-tilted cube ($M = 0.75$, 95% CI [0.73, 0.77]) was similar to auROC for reports of seeing a downward-tilted cube ($M = 0.72$, 95% CI [0.70, 0.75]; Cohen's $d = 0.22$; $BF_{10} = 1.38$; see Figure 4, panel 5.). In contrast with the Bayes Factor analysis, a t-test revealed a significant metacognitive asymmetry, with higher metacognitive sensitivity for perceiving an upward-tilted (default-violating) cube ($t(105) = 2.29$, $p = .012$).

Hypothesis 3: (The lack of) metacognitive asymmetry was not different from what would be expected based on an equal-variance SDT model with the same response bias and sensitivity (Cohen's $d=0.22$; $BF_{10} = 1.28$). Here also, frequentist and Bayesian analyses conflicted, with a t-test revealing a significant metacognitive advantage for upward-tilted (default violating) responses when controlling for bias ($t(105) = 2.25$, $p = .013$).

Hypothesis 4: Contrary to our hypothesis, response times for 'upward-tilted' responses were similar to response times for 'downward-tilted' responses with a median difference of 9 ($t(105) = -0.82$, $p = .207$; Cohen's $d = 0.08$; $BF_{01} = 6.19$; see Figure 4, panel 5).

In summary, in Experiment 5 we found no sign of processing asymmetry between upward and downward-tilted cubes in response-times and confidence. A significant metacognitive asymmetry was observed when using null-hypothesis significance testing, but was not supported by our Bayes Factor analysis. In accordance with our pre-registered plan to commit to the Bayes Factor analysis in interpreting the results, in what follows we interpret these findings as providing no support for a metacognitive asymmetry for upward and downward tilted cubes.

3.6 Experiment 6: flipped vs. normal letters

In Experiment 6, we examined discrimination judgments between flipped and normal *N* stimuli. Based on a search asymmetry for these stimuli (flipped *N*s are found faster among normal *N*s than vice versa; Frith, 1974; Wang et al., 1994), we hypothesized that a similar asymmetry would emerge in subjective confidence judgments, such that metacognitive sensitivity for perceiving a flipped *N* will be higher than for perceiving a normal *N*. We used four dollar signs (\$\$\$\$) as our mask.

127 participants were recruited from Prolific for Experiment 6.

Median completion time was 12.76 minutes. Mean accuracy was 0.74, and participants reported seeing a normal *N* on 0.50 of the trials. Overall we excluded 21 participants, leaving 106 participants for the main analysis. Going forward, only data from included participants is analyzed.

Mean accuracy among included participants was $M = 0.73$, 95% CI [0.72, 0.74]. The mean SOA in the last trial was $M = 37.26$, 95% CI [33.07, 41.46]. Participants showed no consistent response bias ($M = 0.00$, 95% CI [-0.02, 0.02]). On a scale of 0 to 1, mean confidence level was $M = 0.53$, 95% CI [0.49, 0.57]. Confidence was higher for correct than for incorrect responses ($M_d = 0.17$, 95% CI [0.15, 0.20], $t(105) = 16.45$, $p < .001$).

Hypothesis 1: Contrary to our hypothesis, confidence was *lower* for flipped (feature present) responses than for normal (feature absent) responses. This result was in the opposite direction to what we had expected, so was not significant in a one-tailed t-test ($M_d = -0.04$, 95% CI [-0.06, ∞], $t(105) = -3.32$, $p = .999$; Cohen's $d = 0.32$). However, a Bayes Factor favoured the alternative over the null ($BF_{10} = 18.92$; see Figure 4, panel 6).

Hypothesis 2: Contrary to our hypothesis, auROC for reports of seeing a flipped N ($M = 0.71$, 95% CI [0.69, 0.73]) was similar to auROC for reports of seeing a normal N ($M = 0.71$, 95% CI [0.69, 0.73]; $t(105) = 0.08$, $p = .468$; Cohen's $d = 0.01$; $BF_{01} = 8.54$; see Figure 4, panel 6.).

Hypothesis 3: (The lack of) metacognitive asymmetry was not different from what would be expected based on an equal-variance SDT model with the same response bias and sensitivity ($t(105) = 0.26$, $p = .396$; Cohen's $d = 0.03$; $BF_{01} = 8.28$).

Hypothesis 4: Contrary to our hypothesis, response times for 'flipped' responses were significantly *slower* than response times for 'normal' responses, with a median difference of 30 ms. ($t(105) = 2.81$, $p = .997$; Cohen's $d = 0.27$; $BF_{10} = 4.66$; see Figure 4, panel 6).

In summary, in Experiment 6 we found a significant difference in response speed and subjective confidence in the opposite direction to what we expected, with a processing advantage for the default-complying stimulus (N) compared to the default-violating stimulus (flipped N). We found no metacognitive asymmetry for these stimuli.

3.7 Experiments 1-6: summary

Overall, the pattern of results from Experiments 1-6 only partly matched our hypotheses in some cases, and stood in direct contrast to them in other cases (see fig. 5). A reliable metacognitive asymmetry was observed only in Experiment 2, and this asymmetry was in the opposite direction to what we had predicted, with a metacognitive advantage for O (feature absent) over C (feature present) responses. A metacognitive advantage for reporting Q over O s (Exp. 1) was not reliably above what is expected based on an equal-variance signal detection model.

For both local and global visual features (Experiments 1-4) we observed differences in mean confidence and response times that were consistent with our hypothesis of a processing advantage for the representation of the presence compared to the absence of visual features. In Experiments 5 and 6, we tested more abstract expectation violations. In Experiment 5, discrimination between upward-tilted and downward-tilted cubes showed no asymmetry in response time, confidence, or metacognitive sensitivity. In Experiment 6, participants were less confident and slower in their reports of seeing a flipped N, contrary to our prediction that default-violating signals should be easier to perceive. We found no evidence for a metacognitive asymmetry for these stimuli.

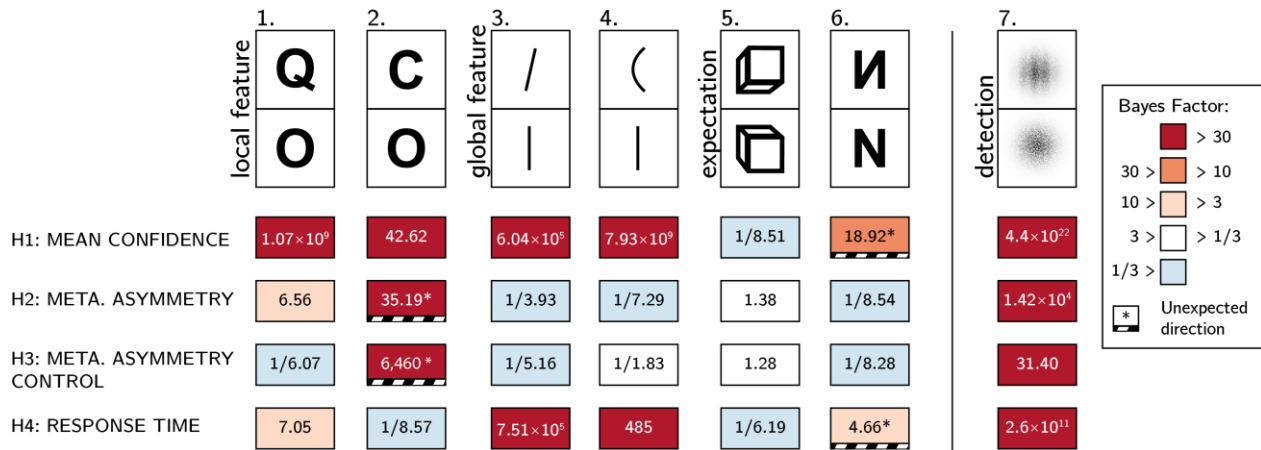


Figure 5. Summary of results from Experiments 1-6, and exploratory Experiment 7. Rows correspond to our four pre-registered hypotheses: difference in confidence, a difference in metacognitive sensitivity, a difference in metacognitive sensitivity when controlling for response and confidence bias, and a difference in response times.

3.8 Experiment 7 (exploratory): grating vs. noise

Results from Experiments 1-6 revealed that search asymmetry is not always accompanied by an asymmetry in metacognitive sensitivity. Given that we did not observe a true metacognitive asymmetry in the expected direction for any of our stimulus pairs, we were concerned that our experimental design may have been unsuitable for detecting classical metacognitive asymmetries in detection, for example due to an insufficient number of trials, the masking procedure, or the confidence report scheme. As a positive control, we collected data for an additional experiment that more closely resembled typical detection experiments. In this experiment, participants discriminated between two stimuli: random noise and a noisy grating (presented to participants as a ‘zebra’ stimulus; see Fig. 6). In a previous lab-based study, similar stimuli produced a robust metacognitive asymmetry between target absent (noise) and target present (noisy grating) responses (Mazor et al., 2020). We used black and white concentric circles as a mask. Apart from the choice of stimuli and mask, the procedure was identical to that of our pre-registered experiments.

127 participants were recruited from Prolific for exploratory Experiment 7. For this positive control, all four hypotheses were fulfilled.

Median completion time was 10.70 minutes. Mean accuracy was 0.73, and participants reported seeing a grating on 0.48 of the trials. Overall we excluded 36 participants, leaving 105 participants for the main analysis. Going forward, only data from included participants is analyzed.

Mean accuracy among included participants was $M = 0.76$, 95% CI [0.74, 0.77]. The mean SOA of the last trial was $M = 53.87$, 95% CI [38.85, 68.89]. Participants showed no

consistent response bias ($M = 0.01$, 95% CI [0.00, 0.03]). On a scale of 0 to 1, mean confidence level was $M = 0.55$, 95% CI [0.51, 0.59]. Confidence was higher for correct than for incorrect responses ($M_d = 0.15$, 95% CI [0.13, 0.17], $t(104) = 12.58$, $p < .001$).

Hypothesis 1: In line with our hypothesis, confidence was higher for reports of target presence than for reports of target absence ($M_d = 0.20$, 95% CI [0.17, ∞], $t(104) = 14.07$, $p < .001$; Cohen's $d = 1.37$; $BF_{10} = 4.39 \times 10^{22}$; see Figure 6, right panel).

Hypothesis 2: In line with our hypothesis, auROC for reports of target presence ($M = 0.75$, 95% CI [0.73, 0.77]) was significantly higher than for reports of target absence ($M = 0.68$, 95% CI [0.66, 0.70]; $t(104) = 5.20$, $p < .001$; Cohen's $d = 0.51$; $BF_{10} = 1.42 \times 10^4$; see Figure 6, left panel).

Hypothesis 3: In line with our hypothesis, this metacognitive asymmetry was significantly stronger than what is expected based on an equal-variance SDT model with the same response bias and sensitivity ($t(104) = 3.49$, $p < .001$; Cohen's $d = 0.34$; $BF_{10} = 31.40$).

Hypothesis 4: In line with our hypothesis, reports of target presence were significantly faster than reports of target absence, with a median difference of 124 ms. ($t(104) = -8.84$, $p < .001$; Cohen's $d = 0.86$; $BF_{10} = 2.63 \times 10^{11}$ see Figure 6, right panel).

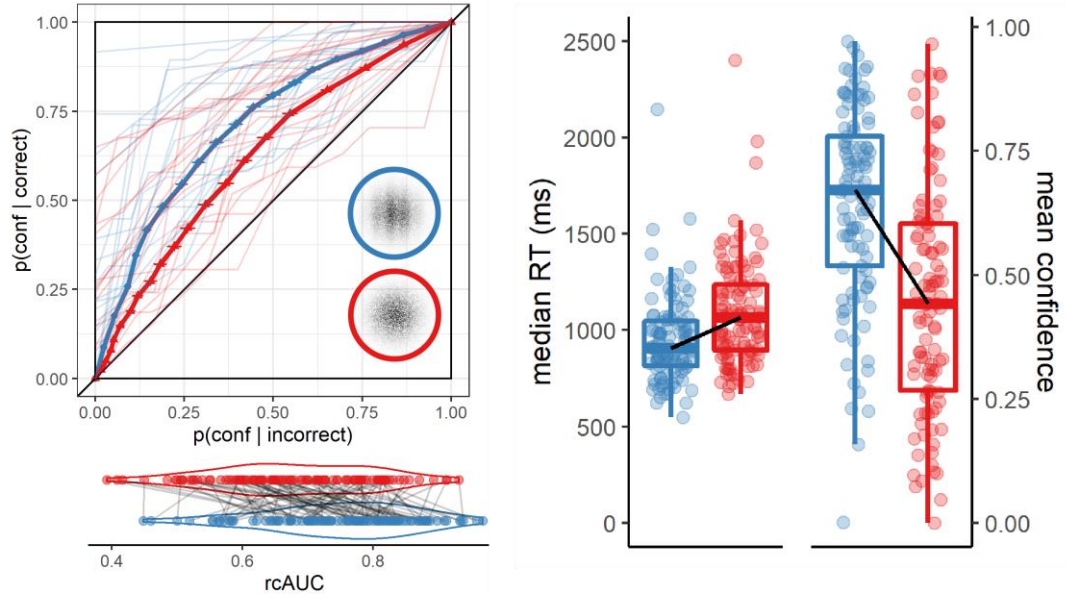


Figure 6. Response conditional ROC curves (left panel) and confidence and reaction time distributions (right panel) for Exp. 7 (detection positive control). The structure of this figure is similar to Figs 3-4.

3.9 Exploratory analysis

3.9.1 zROC analysis.

In a signal-detection framework, metacognitive asymmetry appears when the signal distribution has both a higher mean and higher variance than that of the noise distribution. This unequal variance setting produces higher metacognitive sensitivity for judgments of signal presence, compared to judgments of signal absence. A direct measure for the ratio between the variances of the two distributions is the slope of the *type-1 zROC curve*. A zROC curve is constructed by applying the inverse of the normal cumulative density function to false alarm and hit rates for different confidence thresholds. The slope of the zROC curve equals 1 exactly when the variance of the signal and noise distributions are equal. In detection experiments, the slope is often shallower than 1, indicating a wider signal distribution. Indeed, in our positive control experiment (Exp. 7), the median zROC slope was 0.86 and significantly

shallower than 1 ($t(103) = -5.08, p < .001$; $BF_{10} = 8.66 \times 10^3$ for a t-test on the log-slope against zero). Measuring the slope of the zROC curve in our six pre-registered experiments, we asked whether our ‘feature-present’ distributions had higher variance than our ‘feature-absent’ distributions.

zROC slopes were numerically shallower than one in Experiments 1 (Q vs. O; median slope = 0.95), 3 (line tilt; median slope = 0.94), 4 (line curvature; 0.97) and 5 (cube orientation; 0.95). This was significant only in Experiment 5 ($t(101) = -2.09, p = .039$). In agreement with the results of our rcROC analysis, the zROC slope in Exp. 2 (C vs. O) was significantly higher than one, suggesting that the representation of the letter ‘O’ was more variable than that of the letter ‘C’ (median slope = 1.09; $t(104) = 2.29, p = .024$).

Previous studies reported similar variance structures for these stimuli when presented in visual search arrays. For example, confidence in a vertical/tilted visual search task revealed higher variance in the representation of tilted (feature positive) compared to vertical (feature negative) stimuli (Vincent, 2011). Similarly, reverse correlation analysis revealed higher variance in the representation of *Q* (feature positive) compared to *O* (feature negative) stimuli (Saiki, 2008). Finally, and in agreement with our results, variance in the representation of *O* (feature negative) was found to be higher than in the representation of *C* (feature positive) (Doshier et al., 2004). Note that for the case of line tilt and *Q* vs. *O*, finding a high-variance target among low-variance distractors is easier than finding a low-variance target among high-variance distractors. However, the opposite is true for *C* vs. *O*, where a low-variance target (*C*) renders the search easier. This last observation challenges the suggestion that variance structure is the determining factor for visual search asymmetries (Doshier et al., 2004; Saiki, 2008; Treisman & Gormican, 1988; Vincent, 2011).

3.9.2 Inter-subject correlations. Across experiments, asymmetry in mean confidence (Hypothesis 1) and in response time (Hypothesis 4) were mostly aligned. This is consistent with previous reports of a negative correlation between response times and confidence across trials within participants (Calder-Travis et al., 2020; Henmon, 1911; Moran et al., 2015; Pleskac & Busemeyer, 2010). To test if this was the case across participants too, and not only across experiments, we fitted a mixed-effects regression model to data from all seven experiments with experiment as a random effect ($\Delta RT \sim \Delta conf + (1 + \Delta conf|exp)$). The association between confidence and RT effects was significant in this model ($p < 0.001$; see Fig. 7; upper panel). In contrast, metacognitive asymmetry (difference between the area under the response conditional ROC curves, controlling for response bias) was not significantly associated with asymmetry in either confidence ratings ($p = 0.41$; see Fig. 7; lower panel) or reaction time ($p = 0.54$).

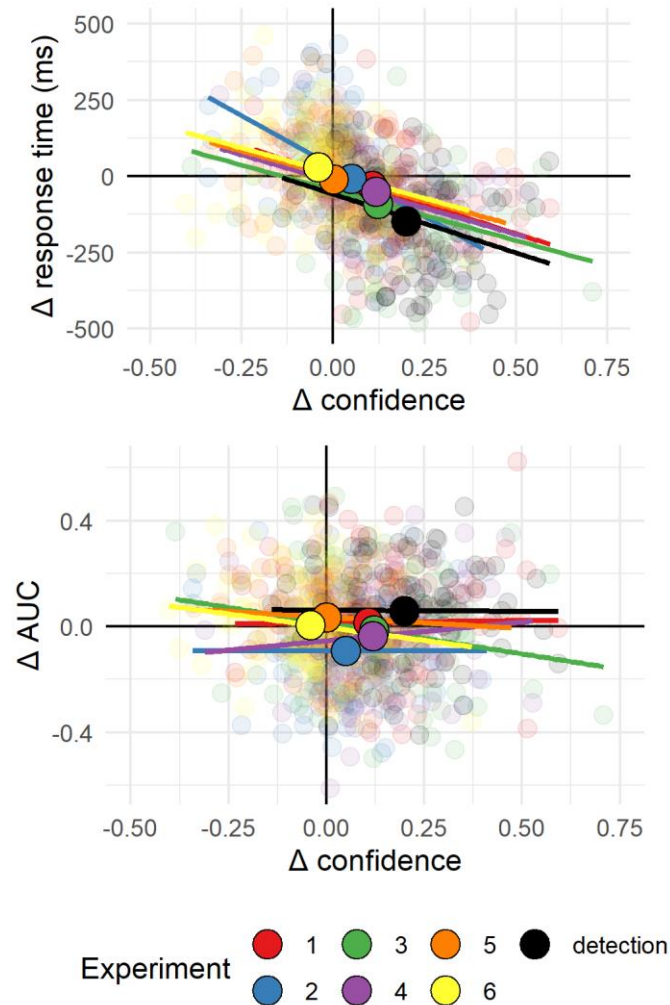


Figure 7. Upper panel: Difference in mean confidence between S1 and S2 responses plotted against difference in mean response time between S1 and S2 responses across the seven experiments. Lower panel: Difference in mean confidence between S1 and S2 responses plotted against difference in metacognitive sensitivity, controlling for response bias, across the seven experiments. Semi-transparent circles represent individual subjects. Opaque circles are the means for each of the seven experiments, across participants. Lines indicate the best-fitting linear regression line for experiments 1-7.

4 Discussion

In perceptual detection, judgments about the presence or absence of a target stimulus differ in several ways: first, participants are more confident in stimulus presence than absence (e.g., Meuwese et al., 2014; Kellij et al., 2018), second, confidence ratings in judgments of stimulus presence have greater metacognitive sensitivity, being more aligned with objective accuracy (Kellij et al., 2018; Mazor et al., 2020; Meuwese et al., 2014), and finally, participants are faster to report stimulus presence (Mazor et al., 2020). In our positive control detection experiment (Experiment 7) we replicated these effects. We found a mean difference of 20% confidence between decisions about the presence or absence of a grating, a metacognitive asymmetry of 0.07 in AUC units (ranging from 0 to 1), and a median difference of 124 milliseconds in response time between reports of target presence and absence. In our six pre-registered experiments, we focused on these three behavioural signatures of decisions about the presence and absence of a stimulus, and asked whether they extend to discrimination tasks where stimuli are distinct in the presence or absence of sub-stimulus features such as the presence of an additional line in a letter, the curvature of a line, or more abstractly, the presence of a surprising default-reasoning signal.

Starting from the end, Experiments 5 and 6 provide evidence against the proposal that asymmetries in confidence, reaction time and metacognitive sensitivity emerge for default-violating signals at all levels of representation. Stimulus pairs in Exp. 5 (cube orientation) and 6 (letter inversion) produced response-conditional ROC curves that were more consistent with the absence of metacognitive asymmetry than with our prior distribution on effect sizes (see section 2.3.1 for the specifics of our Bayesian hypothesis testing, including our prior on effect sizes). Given that these stimuli have been shown to produce reliable asymmetries in visual search (Frith,

1974; Malinowski & Hübner, 2001; Shen & Reingold, 2001; Von Grünau & Dubé, 1994; Wang et al., 1994), we can safely conclude that not all default violations that produce an asymmetry in visual search also produce an asymmetry in metacognitive sensitivity.

Moreover, in Exp. 6, default-complying *N* responses were faster, and accompanied by higher levels of subjective confidence, than default-violating flipped-*N* responses. This is in contrast to our prediction of a processing advantage for default-violating signals, and in line with previous reports of a processing advantage for familiar over unfamiliar stimuli in the context of face perception and reading. For example, in a breaking continuous flash suppression (bCFS) paradigm, inverted faces took longer to break into awareness than upright faces (Stein & Peelen, 2021). A similar processing advantage for familiar stimuli has been documented for the perception of words (Albonico, Furubacke, Barton, & Oruc, 2018) and Chinese letters (Xue, Chen, Jin, & Dong, 2006). One possibility is that the perception of highly familiar stimuli such as letters and faces is supported by specific expert brain systems, affording a processing advantage beyond the general superior processing of default-violating signals (Xue et al., 2006; Yovel & Kanwisher, 2005). Indeed, Exp. 6 was the only experiment in which we observed a processing advantage for familiar over unfamiliar stimuli.

Next, in Experiments 3 and 4 we looked at two features that have a global effect on stimulus appearance: tilt and curvature. Based on visual search asymmetries, Treisman and Gormican (1988) proposed that tilt and curvature are represented as positive features in the visual system. This takes us one step closer to typical detection experiments: participants now detect the presence or absence of a basic visual feature. In agreement with our Hypotheses 1 and 4, participants were more confident in identifying tilted and curved lines (mean differences of 0.12 and 0.12 on a 0-1 confidence scale), and were faster in giving these responses (mean differences

of 67.67 and 50.57 ms). However, we did not observe a metacognitive asymmetry for these global visual features, further suggesting that the presence of a search asymmetry does not immediately entail the presence of a metacognitive asymmetry.

Our strongest candidate for a stimulus pair for which we expected to find a presence-absence asymmetry was *Q* vs. *O* (Exp. 1). The difference between these two letters is the presence of an additional line stroke: a concrete stimulus part that is localized in space and is independent of the rest of the stimulus. Theoretically, participants could approach this task as a detection task: ignore the common denominator (*O*) and focus on the presence or absence of the distinctive feature (';'). As we hypothesized, participants were more confident in their *Q* responses (mean difference of 0.11 on a 0-1 confidence scale). Participants were also faster in their *Q* responses (median difference of 37 ms). However, unlike stimulus-level detection, a small difference of 0.04 units in the area under the response conditional ROC curves was not significantly different than what is expected based on a null SDT model.

Finally, In Experiment 2 we looked at discrimination between *C* and *O*s based on evidence from visual search that open edges are represented as a positive feature in the visual system (Takeda & Yagi, 2000; Treisman & Gormican, 1988; Treisman & Souther, 1985). As we hypothesized, *C* responses were accompanied by higher levels of subjective confidence (mean difference of 0.05 on a 0-1 confidence scale), and were delivered faster than *O* responses (with a modest but significant difference of 6 ms between the two responses). However, in striking contrast to our original hypothesis, metacognitive sensitivity was *lower* for *C* responses (mean difference of 0.05 AUC units), even when controlling for response bias. This result strongly supports different underlying mechanisms behind search and metacognitive asymmetries. Furthermore, the results of Experiment 2 suggest distinct factors mediate the processing

advantage for presence over absence (as reflected in shorter response times and higher confidence for *C* responses), and the metacognitive asymmetry between presence and absence (as reflected in improved metacognitive sensitivity for *O* responses).

C and *O* are unique in that the difference between them corresponds to two contrasting notions of presence and absence. On the one hand, *C* is marked by the presence of one additional feature - open edges (Treisman & Gormican, 1988; Treisman & Souther, 1985). On the other hand, it is marked by the absence of a piece: there is simply less of it relative to *O*. These two notions of presence and absence are typically coupled in detection. For example, the presence of a grating on a screen corresponds to the presence of additional features (such as orientation, contrast, and phase) as well as of more ‘visual stuff’, relative to the blank background. A compelling interpretation of the results of Exp. 2 is that it is the presence or absence of visual features such as open edges that is driving the difference in confidence and response time, whereas a more quantitative notion of presence or absence (the amount of ‘visual stuff’ presented) is driving the metacognitive asymmetry between these two responses. We note however that based on this interpretation, we would expect a metacognitive sensitivity to operate also in Experiment 1, where *O* is missing a piece relative to *Q*. As described above, Experiment 1 provided no evidence for such a metacognitive asymmetry beyond what is expected from an equal-variance signal-detection model.

When interpreting our findings in a broader context, it is useful to note that in all six experiments we used backward masking for controlling the visibility level of our stimuli. Different visibility manipulations have been shown to affect detection metacognitive sensitivity in different ways. For example, whereas metacognitive sensitivity in detection ‘no’ responses is at chance when backward masking is used, it is significantly higher than chance when the

attentional blink is used to control stimulus visibility (Kanai et al., 2010). Similarly, phase scrambling but not attentional blink produces a metacognitive advantage for ‘yes’ responses (Kellij et al., 2018). While our positive control (Exp. 7) produced a reliable metacognitive asymmetry between judgments of target presence and absence, it was also the only experiment where stimulus visibility was controlled with low contrast, in addition to backward masking (for the purpose of compatibility with previous experiments; see Fig. 6). Based on our findings alone, we cannot rule out the possibility that using other visibility manipulations may reveal metacognitive asymmetries for the presence or absence of abstract default violations.

Together, our findings weigh against our original proposal that metacognitive asymmetries in perceptual detection are a signature of higher-order default reasoning. Unlike search asymmetries that extend to abstract levels of representations such as familiarity (Wang et al., 1994; Wolfe, 2001) and even social features such as ethnicity and gender (Gandolfo & Downing, 2020; Levin & Angelone, 2001), metacognitive asymmetries in visual discrimination are grounded in concrete visual processing. Furthermore, we provide evidence for a dissociation between asymmetries in metacognition and in response time and confidence, where the latter is linked to activation of basic feature-detectors, for example of orientation, open ends, or curvature.

5 Conclusion

In a set of six experiments, we sought to test the proposal that a metacognitive asymmetry between the representation of stimulus presence and absence is one instance of a more general asymmetry between the representation of default states and default-violating surprises. Our findings provide evidence against this idea. A metacognitive asymmetry was not observed in near-threshold discrimination between stimulus pairs that differ in their alignment with default

expectations. This was the case even in pairs that showed substantial asymmetries in response time and overall confidence levels. Results from our pre-registered set of analyses are most consistent with a narrow interpretation of the presence/absence metacognitive asymmetry in visual detection, that is limited to concrete, spatially localized presences. Furthermore, a metacognitive asymmetry between *Cs* and *Os* in the opposite direction to what is observed in visual search indicates that different cognitive and perceptual mechanisms underlie these two apparently similar phenomena.

6 References

- Albonico, A., Furubacke, A., Barton, J. J., & Oruc, I. (2018). Perceptual efficiency and the inversion effect for faces, words and houses. *Vision Research*, 153, 91–97.
- Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Borchers, H. W. (2019). *Pracma: Practical numerical math functions*. Retrieved from <https://CRAN.R-project.org/package=pracma>
- Calder-Travis, J., Charles, L., Bogacz, R., & Yeung, N. (2020). Bayesian confidence in optimal decisions.
- Champely, S. (2020). *Pwr: Basic functions for power analysis*. Retrieved from <https://CRAN.R-project.org/package=pwr>
- Coldren, J. T., & Haaf, R. A. (2000). Asymmetries in infants' attention to the presence or absence of features. *The Journal of Genetic Psychology*, 161(4), 420–434.
- De Cornulier, B. (1988). Knowing whether, knowing who, and epistemic closure. *Questions and Questioning*, 182–192.
- De Leeuw, J. R. (2015). JsPsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1–12.
- Dosher, B. A., Han, S., & Lu, Z.-L. (2004). Parallel processing in visual search asymmetry. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1), 3.

- Ekstrøm, C. T. (2019). *MESS: Miscellaneous esoteric statistical scripts*. Retrieved from <https://CRAN.R-project.org/package=MESS>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8, 443.
- Frith, U. (1974). A curious effect with reversed letters explained by a theory of schema. *Perception & Psychophysics*, 16(1), 113–116.
- Gandolfo, M., & Downing, P. E. (2020). Asymmetric visual representation of sex from human body shape. *Cognition*, 205, 104436.
- Hearst, E. (1991). Psychology and nothing. *American Scientist*, 79(5), 432–443.
- Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, 18(3), 186.
- Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 57.
- Kanai, R., Walsh, V., & Tseng, C.-h. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, 19(4), 1045–1057.
- Kellij, S., Fahrenfort, J., Lau, H., Peters, M. A., & Odegaard, B. (2018). The foundations of introspective access: How the relative precision of target encoding influences metacognitive performance.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.

<https://doi.org/10.18637/jss.v082.i13>

Labes, D., Schütz, H., Lang, B., & Labes, M. D. (2020). Package “powertost”. *Power*, 2, 49.

Lakens, D. (2016). *Power analysis for default Bayesian t-tests*. Retrieved from

<http://daniellakens.blogspot.com/2016/01/power-analysis-for-default-bayesian-t.html>

Levin, D. T., & Angelone, B. L. (2001). Visual search for a socially defined feature: What causes the search asymmetry favoring cross-race faces? *Perception & Psychophysics*, 63(3), 423–435.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477.

Malinowski, P., & Hübner, R. (2001). The effect of familiarity on visual-search performance: Evidence for learned basic features. *Perception & Psychophysics*, 63(3), 458–463.

Mazor, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *Elife*, 9, e53900.

Mazor, M., Mazor, N., & Mukamel, R. (2019). A novel tool for time-locking study plans to results. *European Journal of Neuroscience*, 49(9), 1149–1156.

McCarthy, L. (2015). P5. Js. URL: [Https://P5js.Org](https://P5js.Org), 3.

- Meuwese, J. D., Loon, A. M. van, Lamme, V. A., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: Comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics*, 76(4), 1057–1068.
- Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive Psychology*, 78, 99–147.
- Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from <https://CRAN.R-project.org/package=BayesFactor>
- Morey, R. D., Rouder, J. N., Jamil, T., & Morey, M. R. D. (2015). Package “bayesfactor”. URL [http://Cran/R-Projectorg/Web/Packages/BayesFactor/BayesFactor Pdf I](http://Cran/R-Projectorg/Web/Packages/BayesFactor/BayesFactor%20Pdf%20I) (Accessed 1006 15).
- Navarro, D. (2015). *Learning statistics with r: A tutorial for psychology students and other beginners*. (Version 0.5). Adelaide, Australia: University of Adelaide. Retrieved from <http://ua.edu.au/ccs/teaching/lsr>
- Newman, J. P., Wolff, W. T., & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology: Human Learning and Memory*, 6(5), 630.
- Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking & Reasoning*, 8(2), 135–151.
- Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, 5(8), 349–357.

- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Robinson, D., & Hayes, A. (2020). *Broom: Convert statistical analysis objects into tidy tibbles*. Retrieved from <https://CRAN.R-project.org/package=broom>
- Saiki, J. (2008). Stimulus-driven mechanisms underlying visual search asymmetry revealed by classification image analyses. *Journal of Vision*, 8(4), 30–30.
- Sainsbury, R. (1971). The “feature positive effect” and simultaneous discrimination learning. *Journal of Experimental Child Psychology*, 11(3), 347–356.
- Shen, J., & Reingold, E. M. (2001). Visual search asymmetry: The influence of stimulus familiarity and low-level features. *Perception & Psychophysics*, 63(3), 464–475.
- Stein, T., & Peelen, M. V. (2021). Dissociating conscious and unconscious influences on visual detection effects. *Nature Human Behaviour*, 1–13.
- Takeda, Y., & Yagi, A. (2000). Inhibitory tagging in visual search can be found if search stimuli remain visible. *Perception & Psychophysics*, 62(5), 927–934.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1), 15.
- Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3), 285.

- Vincent, B. T. (2011). Search asymmetries: Parallel processing of uncertain sensory information. *Vision Research*, 51(15), 1741–1750.
- Von Grünau, M., & Dubé, S. (1994). Visual search asymmetry for viewing direction. *Perception & Psychophysics*, 56(2), 211–220.
- Walton, D. (1992). Nonfallacious arguments from ignorance. *American Philosophical Quarterly*, 29(4), 381–387.
- Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics*, 56(5), 495–500.
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Henry, L. (2020). *Tidyr: Tidy messy data*. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Wilke, C. O. (2019). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*.
Retrieved from <https://CRAN.R-project.org/package=cowplot>
- Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics*, 63(3), 381–389.

- Xue, G., Chen, C., Jin, Z., & Dong, Q. (2006). Language experience shapes fusiform activation when processing a logographic artificial language: An fMRI training study. *Neuroimage*, 31(3), 1315–1326.
- Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, 15(24), 2256–2262.