Stage 1 Registered Report: Metacognitive asymmetries in visual perception

Matan Mazor1, Rani Moran2, & Stephen M. Fleming1,2,3

1 Wellcome Centre for Human Neuroimaging, UCL

2 Max Planck UCL Centre for Computational Psychiatry and Ageing Research

3 Department of Experimental Psychology, UCL

Accepted in principle for publication in *Neuroscience of Consciousness*. Correspondence concerning this article should be addressed to Matan Mazor, 12 Queen Square, London WC1N 3BG. E-mail: [mtnmzor@gmail.com](mailto: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 measure 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.

*Keywords:* absence, presence, metacognition

*Word count:* 4131

Metacognitive asymmetries in visual perception

# Introduction

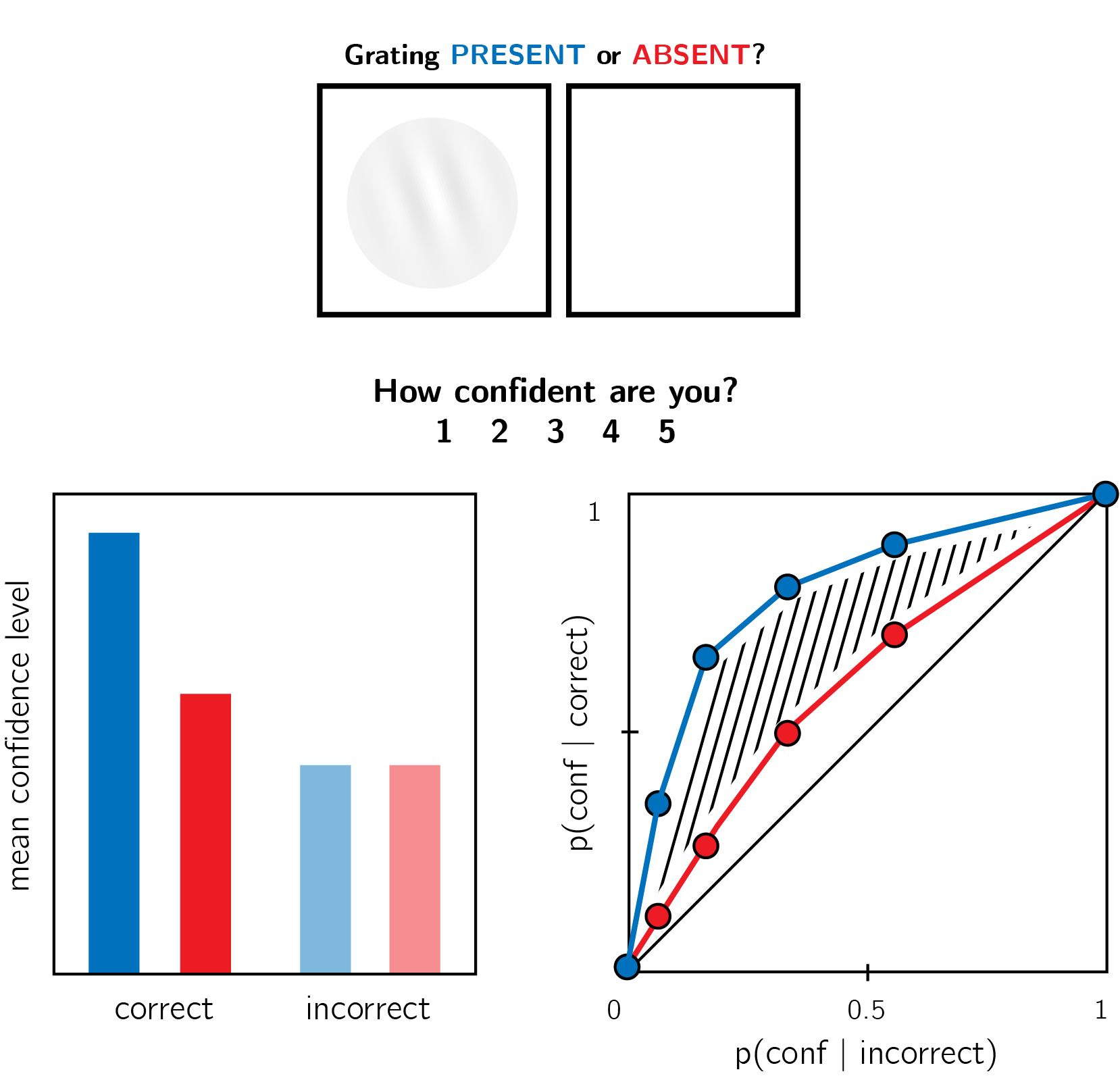
At any given moment, there are many more things that are not there than things that are there. As a result, and in order to efficiently represent the environment, perceptual and cognitive systems have evolved to represent presences, and absence is implicitly represented as a default state (Oaksford, 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 by 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 presence of a distinguishing feature, as compared to searching for a target that is marked by the absence 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 (Dosher, 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.

# Methods

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

We will run six experiments, that will be identical except for the identity of the two stimuli and . Our choice of stimuli for this study is based on the visual search literature. For some stimulus pairs and , searching for one among multiple s is more efficient than searching for one among multiple 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 will 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 among multiple instances of has been found to be more efficient than the inverse search:

1. **Local feature: Addition of a stimulus part**. *Q* and *O* will be used as and , respectively (Treisman & Souther, 1985).
2. **Local feature: Open ends**. *C* and *O* will be used as and , respectively (Takeda & Yagi, 2000; Treisman & Gormican, 1988; Treisman & Souther, 1985).
3. **Global feature: Curvature**. Curved and straight lines will be used as and , respectively (Treisman & Gormican, 1988).
4. **Global feature: Orientation**. Tilted and vertical lines will be used and , respectively (Treisman & Gormican, 1988).
5. **Expectation violation: Letter inversion**. Reversed and normal *N* will be used as and , respectively (Frith, 1974; Wang et al., 1994).
6. **Expectation violation: Viewing angle**. Upward and Downward tilted cubes will be used as and , respectively (Von Grünau & Dubé, 1994).

The experiments will quantify participants’ metacognitive sensitivity for discrimination judgments between and .

## Participants

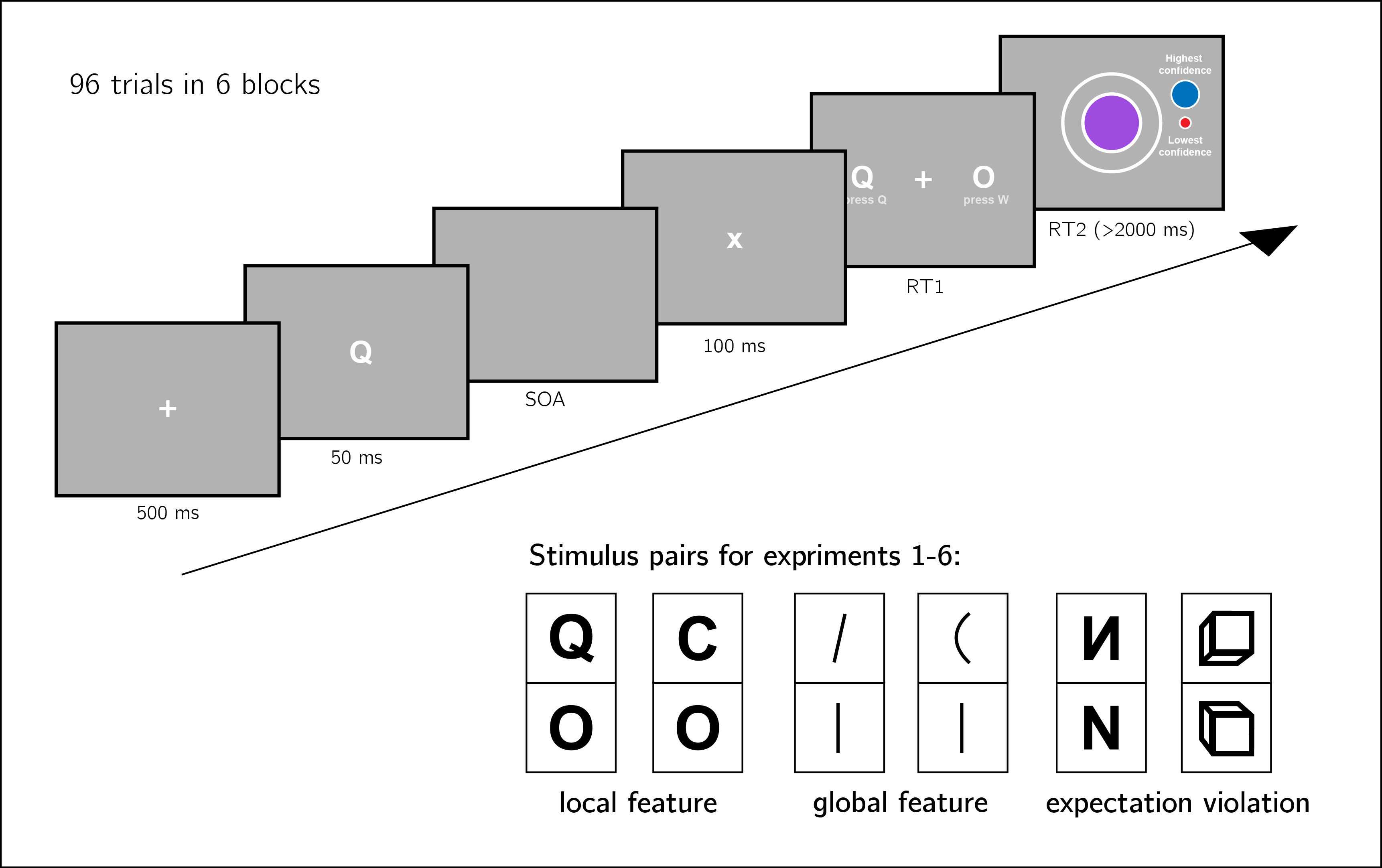
The research complies with all relevant ethical regulations, and was approved by the Research Ethics Committee of University College London (study ID number 1260/003). Participants will be recruited via Prolific, and will give informed consent prior to their participation. They will be selected based on their acceptance rate (>95%) and for being native English speakers. For each of the six experiments, we will collect data until we reach 106 included participants (after applying our pre-registered exclusion criteria). The entire experiment will take 15 minutes to complete. Participants will be paid £2 for their participation, equivalent to an hourly wage of £8.

## Procedure

Experiments were programmed using the jsPsych and P5 JavaScript packages (De Leeuw, 2015; McCarthy, 2015), and will be hosted on a JATOS server (Lange, Kühn, & Filevich, 2015).

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

On each trial, participants will make discrimination judgments on masked stimuli, and rate their subjective decision confidence on a continuous scale. After a fixation cross (500 ms), the target stimulus ( or ) will be presented in the center of the screen for 50 ms, followed by a mask (100 ms). Stimulus onset asynchrony will be 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 will then use their keyboard to make a discrimination judgment. Stimulus-key mapping will be counterbalanced between participants. Following response, subjective confidence ratings will be given on an analog scale by controlling the size of a colored circle with the computer mouse. High confidence will be 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 will terminate once participants click their mouse, but not before 2000 ms. No trial-specific feedback will be delivered about accuracy. In order to keep participants motivated and engaged, block-wise feedback will be 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 will be tested for six stimulus features in six separate experiments, ecompassing three levels of abstraction: local features, global features, and expectation violations. The presented trial corresponds to the first stimulus pair, were *Q* and *O* are used as stimuli.

### Randomization.

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

## Data analysis

We will use 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 1.0.0; Wickham et al., 2020), *ggplot2* (Version 3.3.1; Wickham, 2016), *lsr* (Version 0.5; Navarro, 2015), *MESS* (Version 0.5.6; Ekstrøm, 2019), *papaja* (Version 0.1.0.9942; 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 [, ], will test the following hypotheses:

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

For each of the six stimulus pairs, we will test the null hypothesis that subjective confidence for responses is equal to or lower than subjective confidence for the feature-absent stimulus ().

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

For each of the six stimulus pairs, we will test the null hypothesis that metacognitive sensitivity for responses is equal to or lower than metacognitive sensitivity for the responses ().

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

For each of the six stimulus pairs, we will test the null hypothesis that difference between metacognitive sensitivities for and responses is lower than the difference expected from an equivalent equal-variance SDT model ().

1. **Hypothesis 4**: responses are faster on average than responses.

For each of the six stimulus pairs, we will test the null hypothesis that log-transformed response times for responses are equal to or higher than log-transformed response times for responses ().

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 do not expect a response bias (such that the probability of responding is significantly different from 0.5 across participants). However, such a response bias, if found, may bias our metacognitive asymmetry estimate as measured with response-conditional ROC curves. Hypothesis 3 is 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 will interpret 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; Pleskac & Busemeyer, 2010). Based on these findings, if and 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.

### Dependent variables and analysis plan.

Response conditional ROC curves will be extracted by plotting the empirical cumulative distribution of confidence ratings for correct responses against the same cumulative distribution for incorrect responses. This will be done separately for the two responses and , 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 will be used to test Hypothesis 2.

In order to test hypothesis 3, SDT-derived response-conditional ROC curves will be plotted in the following way. For each response, we will plot 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 will be compared against the difference between the true response-conditional ROC curves.

For visualization purposes only, confidence ratings will be 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 will be tested using a one tailed t-test at the group level with . The summary statistic at the single subject level will be difference in mean confidence between and responses for Hypothesis 1, difference in area under the response-conditional ROC curve between and responses () for Hypothesis 2, difference in between true confidence distributions and SDT-derived confidence distributions for hypothesis 3, and difference in mean log response time between and responses for Hypothesis 4.

In addition, a Bayes factor will be 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 will base our inference on the resulting Bayes Factors.

### Statistical power.

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

1. Hypothesis 1 (MEAN CONFIDENCE): With 106 participants, we will have a 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 ().
2. Hypothesis 2 (METACOGNITIVE ASYMMETRY): With 106 participants, we will have a 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 ().
3. Hypothesis 3 (METACOGNITIVE ASYMMETRY: CONTROL): With 106 participants, we will have a 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 ().
4. Hypothesis 4 (RESPONSE TIME): With 106 participants, we will have a 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 ().

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 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 higher than 3 in 70 out of 100 repetitions, and support the null hypothesis with a higher than 3 in 15 out of 100 hypotheses (based on an adaptation of simulation code from Lakens, 2016).

### Rejection criteria.

Participants will be 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. Thus, only participants who committed at least two errors of each error type (for example, mistaking a *Q* of *O* and mistaking an *O* for *Q*), will be included.

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

## Data availability

All raw data will be made fully available on OSF and on the study’s GitHub respository: <https://github.com/matanmazor/asymmetry>. Pilot data is available at: <https://github.com/matanmazor/asymmetry/blob/master/Experiments/Q_in_O/results/pilot/jatos_results_batch3.csv>

## Code availability

All analysis code will be 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, will be available on our GitHub repository.

## Competing interests

The authors declare no competing interests.

# References

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). Acurious 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.

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>

Lange, K., Kühn, S., & Filevich, E. (2015). " Just another tool for online studies”(JATOS): An easy solution for setup and management of web servers supporting online studies. *PloS One*, *10*(6), e0130834.

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

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*, *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.

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”. *URLh Http://Cran/R-Projectorg/Web/Packages/BayesFactor/BayesFactor Pdf I (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>

Sainsbury, R. (1971). The “feature positive effect” and simultaneous discrimination learning. *Journal of Experimental Child Psychology*, *11*(3), 347–356.

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>

# Supplementary information

# Pilot data and analysis

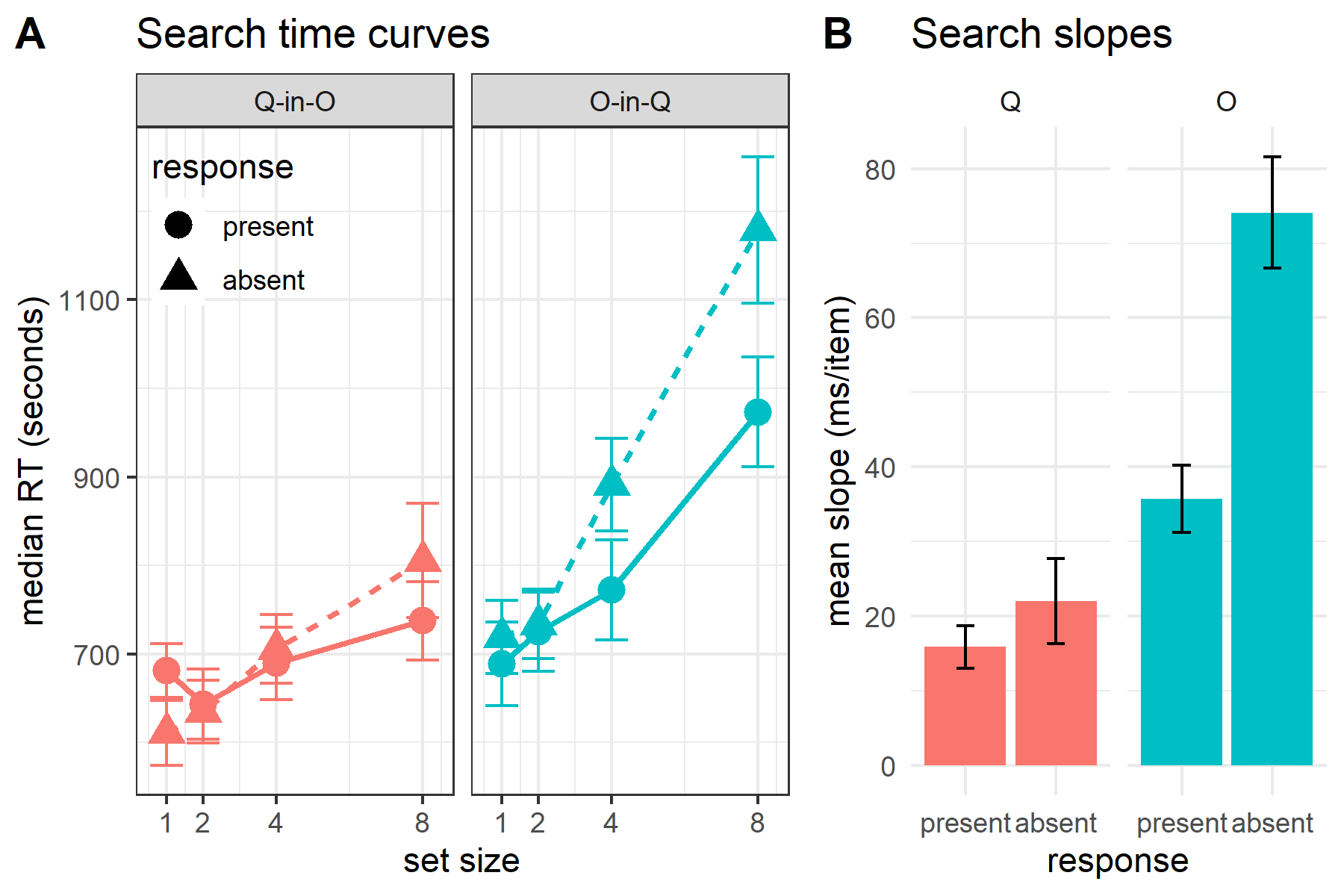
## Pilot Experiment

30 participants were recruited from Prolific for our pilot experiment. We followed a similar procedure to the one described in the Methods section, using *Q* and *O* as our stimuli. In this pilot study we also replicated the visual-search asymmetry for these stimuli. To keep the experiment short and participants engaged, participants only completed 32 discrimination trials. The experiment took about 13.5 minutes to complete. Subjects were paid £1.25 for their participation.

Mean accuracy was 0.80 in the discrimination task, 0.96 in the Q-in-O search task, and 0.95 in the O-in-Q search task. We excluded participants for performing below 70% accuracy in one or two of the search tasks, for performing below 60% accuracy in the discrimination task, for having extremely fast or slow reaction times in one or more of the tasks (below 250 milliseconds or above 5 seconds in more than 25% of the trials), and for failing the comprehension check. Overall we excluded 0 participants, leaving 30 participants for the main analysis.

## Visual search task: search asymmetry replication

Search time analysis was performed on correct trials with reaction time between 250 and 5000 milliseconds. Search slopes for the response time/set size function were extracted for each participant, task and response, and then subjected to a two-way analysis of variance. As expected, we observed significant effects for target identity (mean search slope for *Q-in-O* search: 18.93 ms/item; mean search slope for *O-in-Q* search: 54.89 ms/item; , , ) and response (mean search slope for target absent responses: 48.04; mean search slope for target present responses: 25.78; , , ; see Figure ). An interaction effect was also significant (, , ). These results match previous reports of steeper search slopes for searching a circle among circles crossed by a line than for the inverse search (e.g., Treisman & Souther, 1985).



*Figure* *3:* A: Median search time by distractor set size for the two search tasks and two responses. Correct responses only. Error bars represent the standard error of the median. B: mean search slope per target (Q or O) and response (present or absent). Error bars represent the standard error of the mean.

## Discrimination task: metacognitive asymmetry

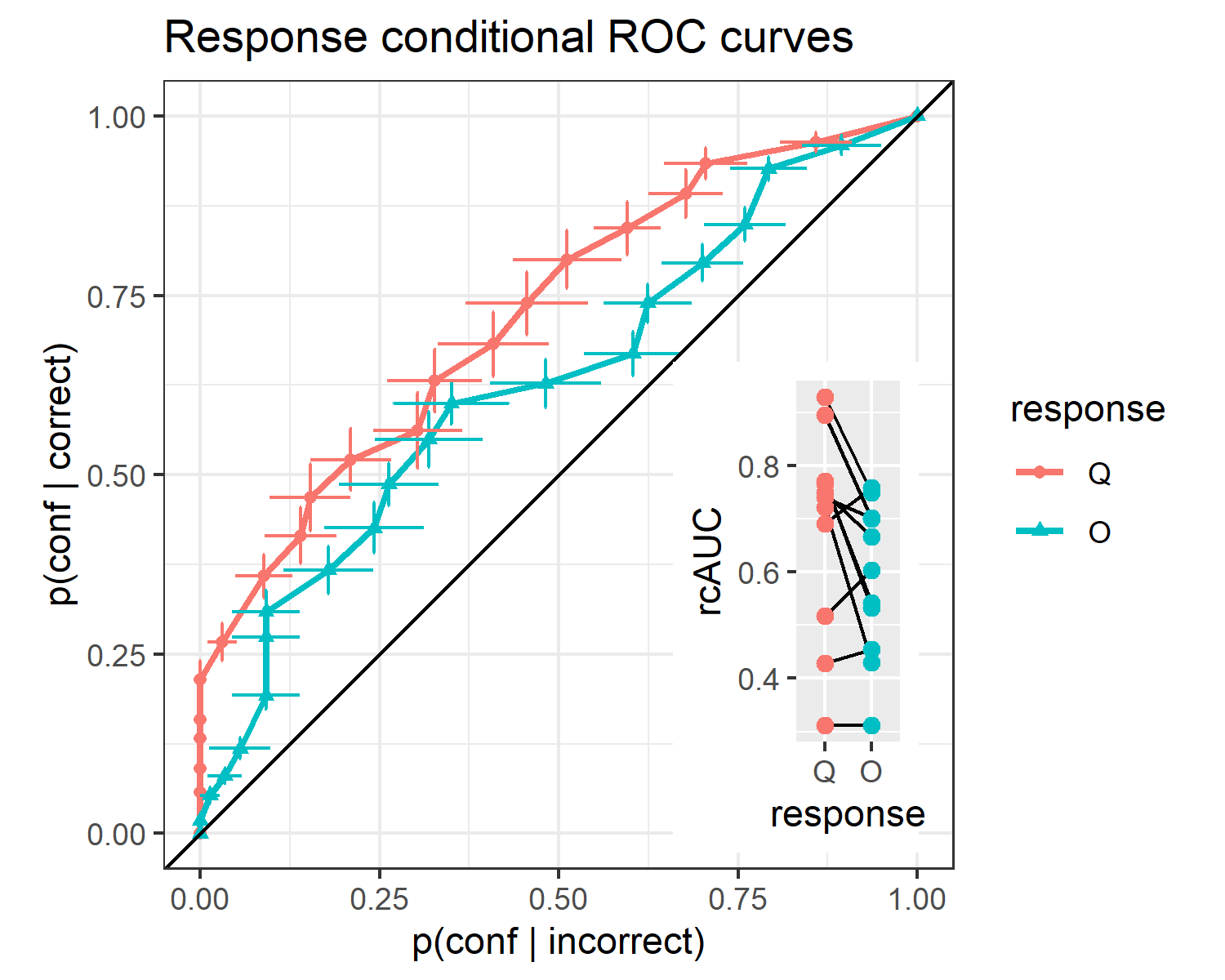
Mean accuracy in the discrimination task was 0.76 (, 95% CI , ). mean SOA in the 32th trial was , 95% CI , . Participants showed no consistent bias in their responses (, 95% CI , ). On a scale of 0 to 1, mean confidence level was 0.53 (, 95% CI , ). Confidence was higher for correct than for incorrect responses (, 95% CI , , , ).

*Hypothesis 1*: In line with our hypothesis, confidence was generally higher for *Q* (feature present) responses than for *O* (feature absent) responses (, 95% CI , , , ; Cohen’s d = 0.66).

*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 (, 95% CI , ) was higher than for *O* responses (, 95% CI , ; , ; Cohen’s d = 0.73; see Figure 4), mirroring the metacognitive asymmetry for detection judgments.

*Hypothesis 3*: Metacognitive asymmetry was significantly higher than what would be expected based on an equal-variance SDT model with the same response bias and sensitivity (, , Cohen’s d=0.81).

*Hypothesis 4*: In line with our hypothesis, *Q* responses were faster on average than *O* responses (, ; Cohen’s d = 0.67).



*Figure* *4:* Response conditional ROC curves for the two discrimination responses. The area under the curve is a measure of metacognitive sensitivity. Bottom right inset: distributions of the area under the curve for the two responses, across participants. Overall, participants had lower metacognitive insight into the accuracy of their ‘O’ responses.