Self-Modeling in Inference about Absence

A Thesis
Presented to
The Division of Wellcome Centre for Human Neuroimaging; Institute of Neurology
University College London

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Matan Mazor

April 2021

Approved for the Division
(Brain Sciences)

Stephen M. Fleming

Karl J. Friston

Acknowledgements

Rani, Roni, Halely, Dina, Karl, Josh, Alisa, Lesley, B7, SIPS & Neurohackademy (Ariel and Tal), Steve, Noam, scihub, Roy, Chudi, Dan, Peter Zeidman, Rani, Dominic and the Hari Krishna volunteers, my participants, imaging team, Kamlyn, Maddi and Shiv, the developers of JATOS (Elisa and Kristian), labjs (Felix), thesisdown (Chester Ismay) and the numerous other pacakges that I've used, stackoverflow community,

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

Table of Contents

List of Tables

List of Figures

Abstract

Representing the absence of things is qualitatively different from representing their presence. Specifically, to represent something as absent one must know that they would have known if it was present. This form of counterfactual reasoning critically relies on having a mental self-model which specifies expected perceptual and cognitive states under different world states. This thesis addresses open questions regarding inference about absence in perceptual decision making: its reliance on prior metacognitive knowledge, relative encapsulation from metacognitive monitoring, neural underpinning, and its relation with default-reasoning and predictive-coding. First, the timing of decisions about the absence of an item has been shown to be sensitive to search time and accuracy in previous trials, but it remains unknown how decisions about the absence of an item are made in the very first trials of the experiment, before previous trials are available. In a set of behavioural experiments I provide evidence for that implicit metacognitive knowledge about spatial attention supports inference about the absence of items already in these first trials, and that this implicit knowledge is dissociable from explicit metacognitive knowledge about search difficulty. Second, subjective confidence in perceptual decisions is mostly sensitive to perceptual evidence supporting the decision, but decisions about stimulus absence are unique in that they are based on the absence of evidence, rendering positive evidence unavailable. Using reverse-correlation I identify positive stimulus features that contribute to decision confidence in decisions about absence, and discuss these findings in the context of sensory noise estimation. Third, neuroimaging studies of metacognitive monitoring have identified a network of frontal and parietal regions that are sensitive to decision confidence. Using functional MRI, I find that these regions are mostly invariant to whether subjective confidence is rated with respect to decisions about presence or absence. In interpreting these results, I formulate computational models that monitor fluctuations in external stimulus strength and in internal attentional states. Finally, in a series of six behavioural experiments I show that different levels of the cognitive hierarchy are sensitive to different notions of absence. I conclude with a discussion of specific ways in which inference about absence can be used by cognitive scientists for probing implicit metacognitive beliefs and studying the mental self-model.

Dedication

You can have a dedication here if you wish.

Chapter 1

This will automatically install the {remotes} package and {thesisdown}

Placeholder

1.1 Inference about absence

1.2 Probabilistic reasoning, criterion setting, and self knowledge

Symmetrical definition:

Dissymmetrical definition:

1.2.1 Second-order cognition

1.2.2 Computational models of detection

The High-Threshold model

Signal Detection Theory

1.3 Detection: "I would have noticed it"

1.4 Visual search: "I would have found it"

1.5 Memory: "I would have remembered it"

1.6 The development of a self-model

1.7 This thesis

Chapter 2

Zero-shot search termination reveals a dissociation between implicit and explicit metacognitive knowledge

Placeholder

Matan Mazor, Stephen M. Fleming

2.1 Introduction

2.2 Experiment 1

2.2.1 Participants

2.2.2 Procedure

Randomization

2.2.3 Data analysis

Rejection criteria

Data preprocessing

Hypotheses and analysis plan

2.2.4 Results

Additional analyses

2.3 Experiment 2

2.3.1 Participants

2.3.2 Procedure

2.3.3 Results

Additional Analyses

Exploratory analysis: task experience

Exploratory analysis: search time estimates

2.4 Discussion

2.4.1 Is implicit metacognitive knowledge metacognitive?

2.4.2 Inference about absence as a tool for studying implicit self knowledge

2.4.3 Conclusion

Chapter 3

Prospective search time estimates for unseen displays reveal a rich intuitive theory of visual search

Placeholder

Matan Mazor, Max Siegel & Joshua B. Tenenbaum

3.1 Introduction

3.2 Experiments 1 and 2: shape, orientation, and color

3.2.1 Participants

3.2.2 Procedure

Familiarization

Estimation

Visual Search

3.2.3 Results

Search times

Estimation accuracy

3.3 Experiments 3 and 4: complex, unfamiliar stimuli

3.3.1 Participants

3.3.2 Procedure

3.3.3 Results

Estimation accuracy

Cross-participant correlations

Estimation time

Visual search asymmetry

Chapter 4

Distinct neural contributions to metacognition for detecting (but not discriminating) visual stimuli

Placeholder

4.1 Introduction

4.2 Methods and Materials

4.2.1 Participants

4.2.2 Design and procedure

4.2.3 Scanning parameters

4.2.4 Analysis

4.2.5 Exclusion criteria

4.2.6 Response conditional type-II ROC curves

4.2.7 Imaging analysis

fMRI data preprocessing

Regions of Interest

Univariate analysis

Main Design Matrix (DM-1)

Global Confidence Design Matrix (GC-DM)

Quadratic-Confidence Design Matrix (post-hoc analysis; QC-DM)

Categorical-Confidence Design Matrices (post-hoc analysis; CC-DM)

Multivariate analysis

4.2.8 Statistical inference

4.3 Results

4.4 Behavioural results

4.4.1 Imaging results

Parametric effect of confidence

Interaction of linear confidence effects with task and response

Interaction of nonlinear confidence effects with task and response

4.4.2 Computational models

4.5 Discussion

Chapter 5

Paradoxical evidence weightings in confidence judgments for detection and discrimination

Matan Mazor, Lucie Charles, Karl J. Friston & Stephen M. Fleming

In two experiments we asked what sensory evidence is incorporated into decisions and confidence judgments in perceptual decisions about stimulus presence or absence (detection) and stimulus category (discrimination). We successfully replicated the positive evidence bias in discrimination confidence ratings: subjective confidence was boosted more by supporting evidence than it was undermined by conflicting evidence, in line with a detection disposition to the discrimination task. We further find that detection judgments show the same positive evidence bias as discrimination confidence ratings. Paradoxically, confidence ratings in detection present a discrimination-like evidence weighting, with equal weighting of positive and negative evidence. First-order perceptual decision making models fail to account for the entire set of findings.

5.1 Introduction

When considering two alternative hypotheses, the probability of a chosen hypothesis to be correct is not only a function of the likelihood of the observations under the chosen hypothesis, but also of the likelihood of the observations under the unchosen one. For example, when deciding that a random dot display was drifting to the right and not to the left, confidence should not only positively weigh motion energy to the right (*positive evidence*), but also negatively weigh motion energy to the left (*negative evidence*). However, in their subjective confidence ratings subjects put unproportional weight on positive evidence, giving rise to a *positive evidence bias* (Koizumi, Maniscalco, & Lau, 2015; Sepulveda et al., 2020; Zylberberg, Barttfeld, & Sigman, 2012). Put differently, confidence ratings in discrimination are sensitive not only to the *relative evidence* of the chosen hypothesis compared with the unchosen one, but also to the *sum evidence* for the two hypotheses (also termed *visibility*; Rausch, Hellmann, & Zehetleitner, 2018).

Focusing on sum rather than relative evidence is rational if subjects are rating their confidence not in the identity of the stimulus, but in the presence or absence of a signal. For example, when judging the direction of motion in a random dot kinematogram, if motion energy is high both to the left and to the right, confidence in the direction of motion should be low (low relative evidence), but confidence in the presence of coherent motion, regardless of its direction, should be high (high sum evidence). A positive evidence bias in discrimination judgments may indicate that participants are rating their confidence not in the accuracy of their choice, but in the presence of a signal.

This implied link between metacognitive evaluation and detection (judgments about the presence or absence of a signal) has led us to examine the contribution of perceptual evidence to decision and confidence in perceptual detection tasks. We were interested in three questions: first, when faced with a detection task where targets are drawn from two stimulus classes, would detection decision be sensitive to sum evidence (like in discrimination confidence), or to the relative evidence for presence for one category over the other? Second: would confidence in the presence of a target stimulus be susceptible to the same positive evidence bias as confidence in stimulus type? And finally, when making decisions about the absence of a signal, would confidence ratings be sensitive to some form of positive evidence for absence, or be entirely independent of sensory evidence?

In two experiments participants performed discrimination and detection decisions on noisy stimuli, and rated their confidence in their decisions. Using reverse correlation analysis we measured the influence of random fluctuations in stimulus energy on their responses and confidence ratings, as well as markers of a processing asymmetry between detection 'yes' and 'no' responses (response time, general confidence, and metacognitive sensitivity). To anticipate our results, we fully replicated previous findings of a positive evidence bias in discrimination responses (Zylberberg et al., 2012). Paradoxically, although detection decisions were sensitive to sum evidence as expected, we found no positive evidence bias in confidence judgments following detection 'yes' responses. In Experiment 2, where reverse correlation revealed an accumulation of positive evidence for stimulus absence, we find no metacognitive sensitivity between the two detection responses. We discuss our findings as drawing a link between discrimination confidence ratings and detection responses, but not detection confidence ratings.

5.2 Experiment 1

5.2.1 Methods

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). 10 participants were recruited via the UCL subject recruiting system, and gave their informed consent prior to their participation. Each participant performed four sessions of 600 trials each, in blocks of 100 trials. Sessions took place on different days and consisted of 3 discrimination blocks interleaved with 3 detection blocks.

Experimental procedure

The experimental procedure for Experiment 1 largely followed the procedure described in Zylberberg et al. (2012), Experiment 1. Participants observed a random-dot kinematogram for a fixed duration of 700 ms. In discrimination trials, the direction of motion was one of two opposite directions with equal probability, and participants reported the observed direction by pressing one of two arrow keys on a standard keyboard. In detection blocks participants reported whether there was coherent motion by pressing one of two arrow keys on a standard keyboard. In half of the detection trials dots moved coherently to one of two opposite directions, and in the other half they moved randomly.

In both detection and discrimination blocks, following a decision participants indicated their confidence in their decision. Confidence was reported on a continuous scale ranging from chance to complete certainty. To avoid response bias in confidence reports, the orientation (vertical or horizontal) and polarity (e.g., right or left) of the scale was set to agree with the type 1 response. For example, following a down arrow press, a vertical confidence bar was presented where 'guess' is at the center of the screen and 'certain' appeared at the lower end of the scale (see Fig. 5.1).

To control for response requirements, for 5 subjects the dots moved to the right or to the left, and for the 5 other subjects they moved upward or downward. The first group made discrimination judgments with the right and left keys and detection judgments with the up and down keys, and this mapping was reversed for the second group. The number of coherently moving dots ("motion coherence") was adjusted to maintain performance at around 70% accuracy for detection and discrimination tasks independently. This was achieved by measuring mean accuracy once in every 20 trials, and adjusting coherence by a step of 3% if accuracy fell below 60% or went above 80%.

Stimuli for discrimination blocks were generated using the exact same procedure reported in Zylberberg et al. (2012)[1]. Trials started with a presentation of a fixation cross for one second, immediately followed by stimulus presentation. The stimulus consisted of 152 white dots (diameter = 0.14°), presented within a 6.5° circular aperture centered on the fixation point for 700 milliseconds (42 frames, frame rate = 60 HZ). Dots were grouped in two patches of equal sizes of 56 dots each. Every other frame, the dots of one patch were replaced with a new set of randomly positioned dots. For a coherence value of c', a proportion of c' of the dots from the second patch moved coherently in one direction by a fixed distance of 0.33°, while the remaining dots in the patch moved in random directions by a fixed distance of 0.33°. On the next update, the patches were switched, to prevent participants from tracing the position of specific dots. Frame-specific coherence values were sampled for each screen update from a normal distribution centred around the coherence value c with a standard deviation of 0.07, with the constraint that c' must be a number between 0 and 1.

[1] We reused the original Matlab code that was used for Experiment 1 in Zylberberg et. al. (2012), kindly shared by Ariel Zylberberg.

Stimuli for detection blocks were generated using a similar procedure, with the only difference being that on a random half of the trials coherence was set to 0%, without random sampling of coherence values for different frames (see Fig. 1).

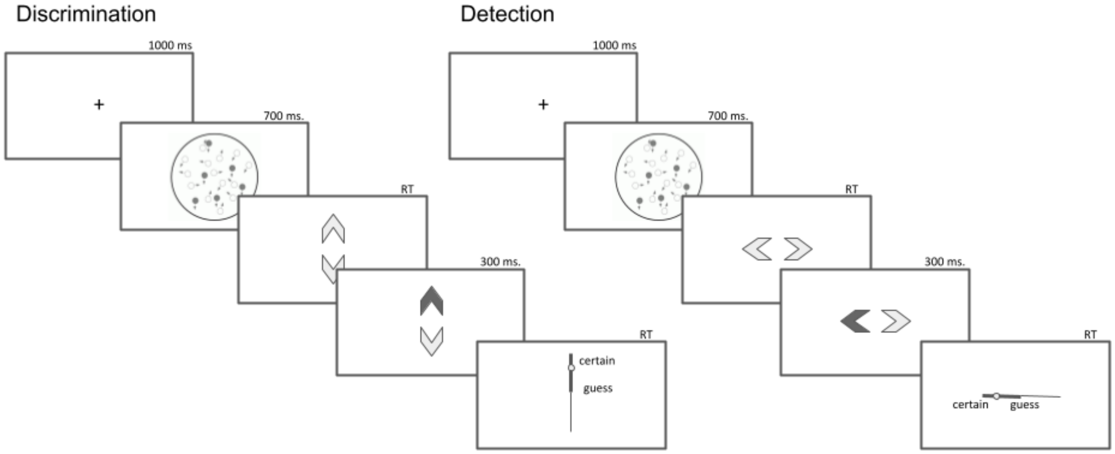

Figure 5.1: Task design for Experiment 1. In both tasks, participants viewed 700 milliseconds of a random dot motion array, after which they made a keyboard response to indicate their decision (motion direction in discrimination, signal absence or presence in detection), followed by a continuous confidence report using the mouse. 5 participants viewed vertically moving dots and indicated their detection responses on a horizontal scale, and 5 participants viewed horizontally moving dots and indicated their detection responses on a vertical scale.

5.2.2 Analysis

Reverse correlation analysis

For the reverse correlation analysis, we followed a procedure similar to the one described in Zylberberg et al. (2012). For each of the four directions (right, left, up and down), we applied two spatiotemporal filters to the frames of the dot motion stimuli as described in previous studies (Adelson & Bergen, 1985; Zylberberg et al., 2012). The outputs of the two filters were squared and summed, resulting in a three-dimensional matrix with motion energy in the specific direction as a function of x, y, and time. We then took the mean of this matrix across the x and y dimensions to obtain an estimate of the overall temporal fluctuations in motion energy in the selected direction. Additionally, for every time point we extracted the variance along the x and y dimensions, to obtain a measure of temporal fluctuations in spatial variance. Using this filter, we obtained trial-wise estimates of temporal fluctuations in the mean and variance of motion energy for upward, downward, leftward and rightward motion. Given a high correlation between our mean and variance estimates, we focused our analysis on the mean motion energy.

In order to distill random fluctuations in motion energy from mean differences between stimulus categories, we subtracted the mean motion energy from trial-specific motion energy vectors. The mean motion energy vectors were extracted at the group level, separately for each motion coherence level and as a function of motion direction. We chose this approach instead of the linear regression approach used by Zylberberg et al. (2012) in order to control for nonlinear effects of coherence on motion energy.

Statistical inference

Statistics were extracted separately for each participant, and group-level inference was then performed on the first-order statistics. T-test Bayes factors were used to quantify the evidence for the null when appropriate, using a Jeffrey-Zellner-Siow Prior for the null distribution, with a unit prior scale (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

5.2.3 Results

Response accuracy

Overall accuracy level was 0.74 in the discrimination and 0.72 in the detection task. Performance for discrimination was significantly higher than for detection ($M_d = 0.02$, 95% CI [0.00, 0.04], $t(9) = 2.43$, $p = .038$). This difference in task performance reflected a slower convergence of the staircasing procedure for the discrimination task during the first session. When discarding all data from the first session and analyzing only data from the last three sessions (1800 trials per participant), task performance was equated between the two tasks at the group level ($M_d = 0.00$, 95% CI [-0.02, 0.02], $t(9) = -0.05$, $p = .962$; $\mathrm{BF}_{01} = 3.24$). In order to avoid conflating true differences between discrimination and detection with more general difficulty effects, the first session was excluded from all subsequent analyses.

Overall properties of response and confidence distributions

In detection, participants were more likely to respond ‘yes’ than ‘no’ (mean proportion of ‘yes’ responses: $M = 0.59$, 95% CI [0.53, 0.64], $t(9) = 3.45$, $p = .007$). We did not observe a consistent response bias for the discrimination data (mean proportion of ‘rightward’ or ‘upward’ responses: $M = 0.52$, 95% CI [0.47, 0.57], $t(9) = 1.00$, $p = .344$).

In detection, participants were generally slower to deliver ‘no’ responses compared to ‘yes’ responses (median difference: 85.37 ms, $t(9) = -3.46$, $p = .007$ for a t-test on the log-transformed response times; see Fig. 5.2, upper panel). No significant difference in response times was observed for the discrimination task (median difference: 6.16 ms, $t(9) = -0.43$, $p = .676$).

Confidence in detection was generally higher than in discrimination ($M_d = 0.06$, 95% CI [0.01, 0.12], $t(9) = 2.49$, $p = .035$; see Fig. 5.2, lower panel). Within detection, confidence in 'yes' responses was generally higher than confidence in 'no' responses ($M = 0.08$, 95% CI [0.03, 0.13], $t(9) = 3.49$, $p = .007$). No difference in average confidence levels was found between the two discrimination responses ($M = 0.02$, 95% CI [−0.03, 0.06], $t(9) = 0.91$, $p = .384$).

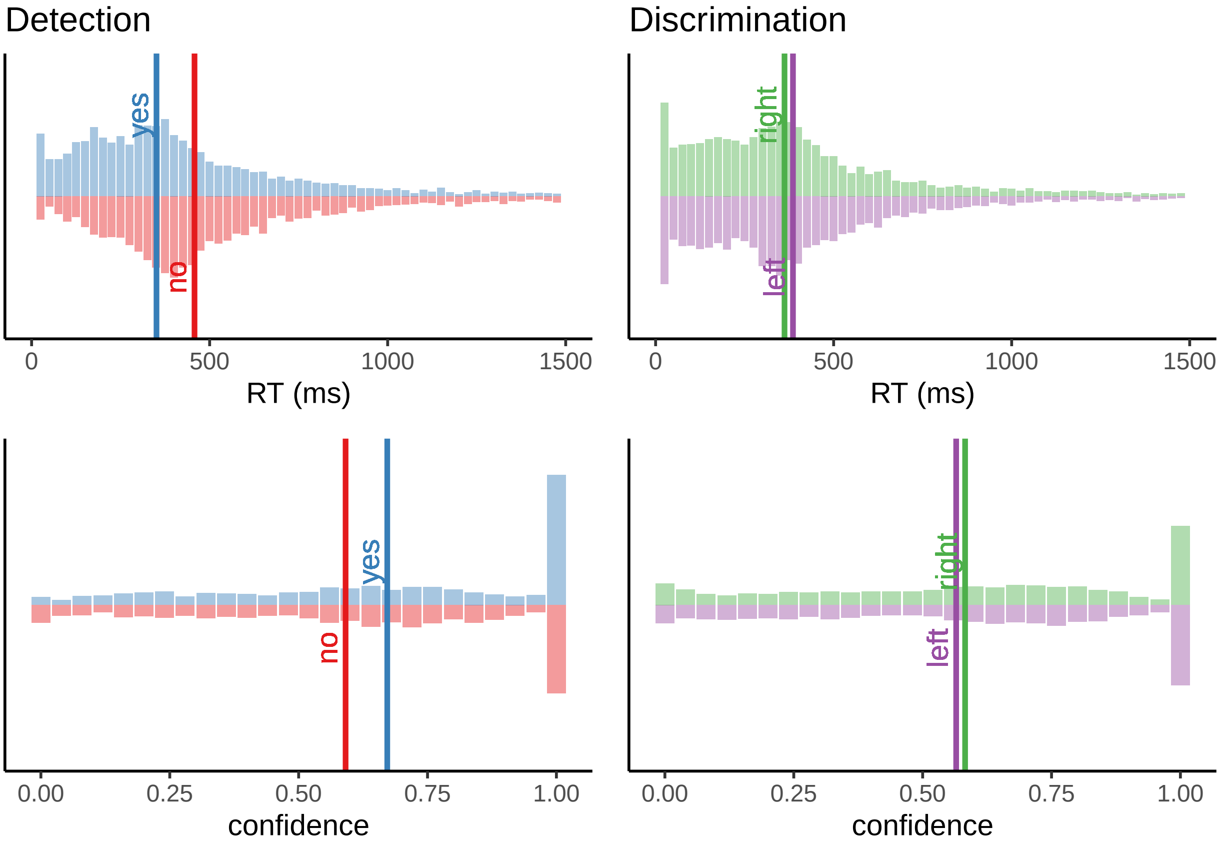

Figure 5.2: Response time (upper panel) and confidence (lower panel) histograms for the detection (left) and discrimination (right) tasks in Experiment 1. Vertical lines represent the median response time and the mean confidence rating for each response.

Response conditional ROC curves

Following Meuwese, Loon, Lamme, & Fahrenfort (2014), we extracted response-conditional type-2 ROC (rc-ROC) curves for the two tasks. Unlike traditional type-I ROC curves that provide a visual representation of subjects' ability to distinguish between two external world states, type 2 ROC curves represent their ability to track the accuracy of their own responses. The area under the response-conditional ROC curve (auROC2) is a measure of metacognitive sensitivity, with higher values corresponding to more accurate metacognitive monitoring.

Mean response-conditional ROC curves for the two responses in the discrimination task closely matched ($M = 0.00$, 95% CI $[-0.05, 0.05]$, $t(9) = 0.13$, $p = .900$), indicating that on average, participants had similar metacognitive insight into the accuracy of the two discrimination responses. In contrast, auROC2 estimates for 'yes' responses were significantly higher than for 'no' responses, indicating a metacognitive asymmetry between the two detection responses (group difference in auROC2: $M = 0.11$, 95% CI $[0.03, 0.18]$, $t(9) = 3.28$, $p = .010$).

To better understand the origin of this difference between 'yes' and 'no' curves, we compared the detection auROC2 values with the average discrimination auROC2. We found both a significant increase in auROC2 for 'yes' responses ($M = 0.06$, 95% CI $[0.01, 0.11]$, $t(9) = 2.80$, $p = .021$) and a marginally significant decrease in auROC2 for 'no' responses relative to discrimination ($M = -0.05$, 95% CI $[-0.10, 0.00]$, $t(9) = -2.16$, $p = .059$). In other words, relative to our discrimination benchmark, metacognitive asymmetry in detection was driven by improved metacognitive insight into the accuracy of 'yes' responses, and degraded metacognitive insight into the accuracy of 'no' responses.

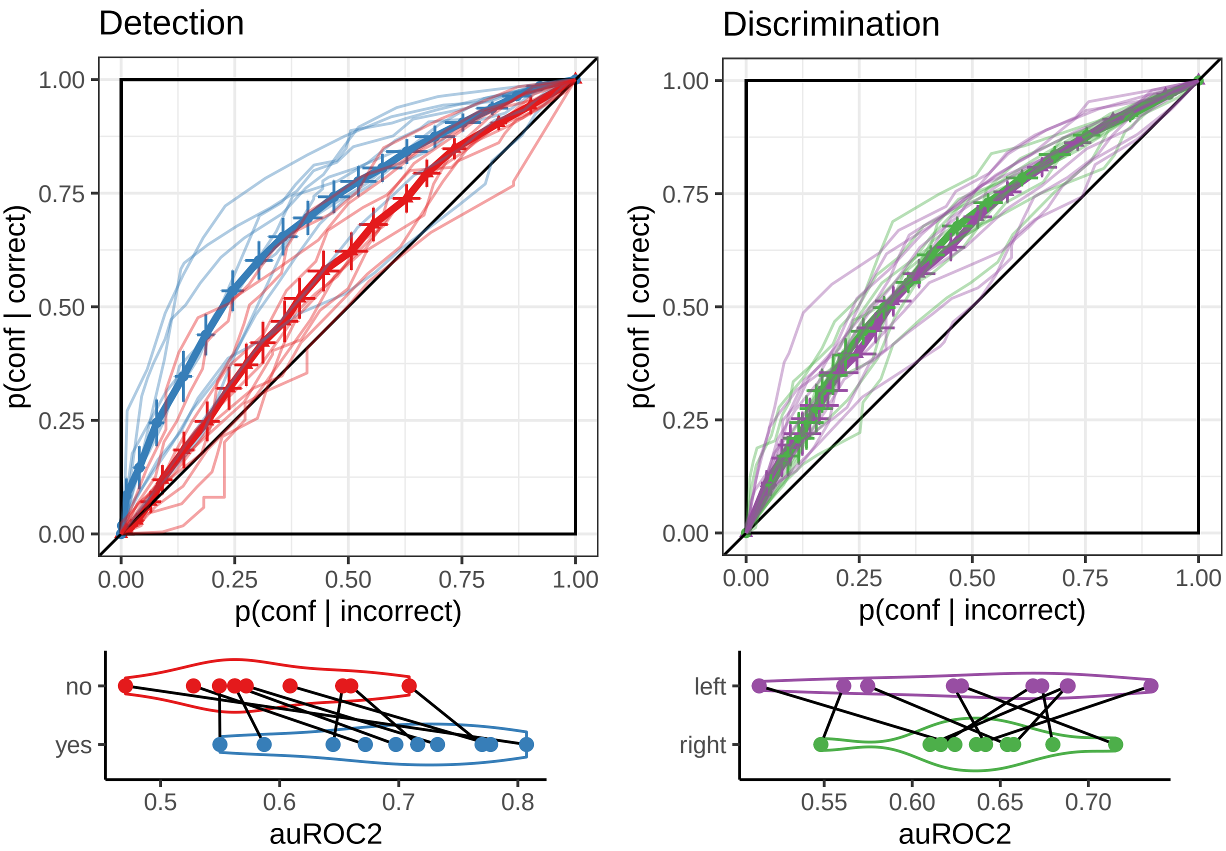

Figure 5.3: Response conditional ROC curves for the two tasks and four responses in Exp. 1. The area under the curve is a measure of metacognitive sensitivity, and the difference in areas between the two responses a measure of metacognitive asymmetry. Lower panel: distributions of the area under the curve for the four responses, across participants. Error bars stand for the standard error of the mean.

A difference in response-conditional auROC estimates can emerge from higher-order differences in metacognitive monitoring for the two responses or from lower-level differences in the perceptual representations of signal and noise (such as in first-order signal detection models where the signal variance is higher; Maniscalco & Lau, 2014). Importantly, a difference can also emerge in first-order signal-detection models that assume equal variance, in the presence of a response bias or insufficient variance in confidence ratings. To test if the metacognitive asymmetry between 'yes' and 'no' responses could be accounted for a by an equal-variance SDT model, we simulated data that was identical to our empirical data except for confidence ratings in correct responses, which were chosen to perfectly agree with the assumptions of an equal-variance SDT model given participants' decision criterion, sensitivity, and their confidence in incorrect responses. We then compared subject-wise differences between the response-conditional auROCs with the differences in this simulated dataset (Mazor, Moran, & Fleming, 2021). The difference in differences was significant, indicating that the observed metacognitive asymmetry could not be accounted for by a first-order equal-variance SDT model ($M = 0.08$, 95% CI [0.02, 0.14], $t(9) = 2.96$, $p = .016$).

Reverse Correlation

Random fluctuations in motion energy made it possible to apply reverse correlation and test which stimulus features are incorporated into decisions and confidence ratings in detection and discrimination. Following Zylberberg et al. (2012), our analysis focused on the first 300 milliseconds since stimulus onset.

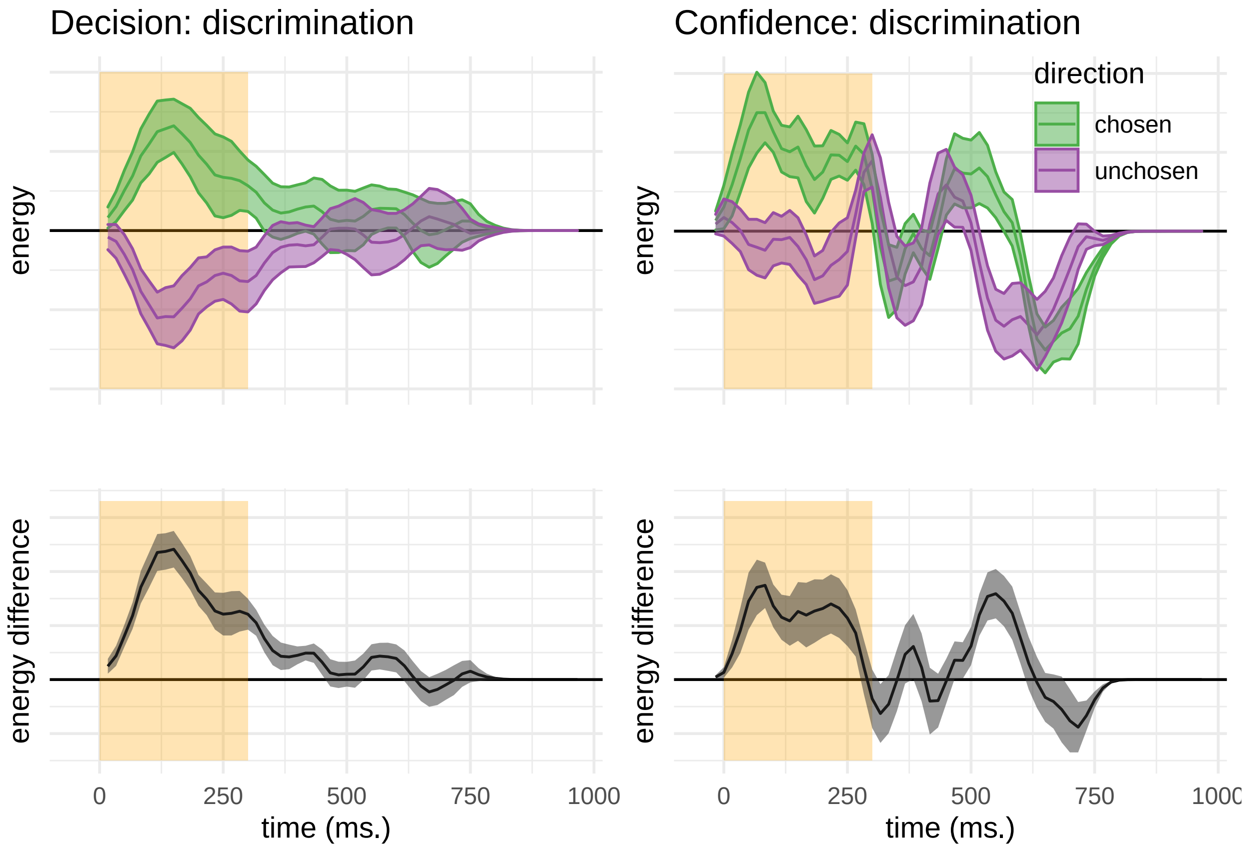

Figure 5.4: Decision and confidence discrimination kernels, Experiment 1. Upper left: motion energy in the chosen (green) and unchosen (purple) direction as a function of time. Lower left: a subtraction between energy in the chosen and unchosen directions. Upper right: confidence effects for motion energy in the chosen (green) and unchosen (purple) directions. Lower right: a subtraction between confidence effects in the chosen and unchosen directions. Shaded areas represent the the mean +- one standard error. The first 300 milliseconds of the trial are marked in yellow

Discrimination Reverse correlation analysis quantified the effect of random fluctuations in motion energy on the probability of responding 'right' and 'left' (or 'up' or 'down'), and the temporal dynamics of decision formation. Similar to the results obtained by Zylberberg et. al., participants' decisions were sensitive to motion energy fluctuations during the first 300 milliseconds of the trial ($t(9) = 7.73$, $p < .001$; see Fig. 5.4, left panels). We note that the symmetry of the two time courses around the x axis does not by itself entail an equal contribution of negative and positive evidence to the final decision, because negative and positive evidence are defined based on participants' decision, making it impossible to test their contribution to decisions without engaging in circular inference. Instead, we tested the contribution of motion energy in the true and opposite directions (defined with respect to the stimulus, not the subject's decision) to discrimination decision. Fluctuations in motion energy in both directions contributed significantly to discrimination decision ($t(9) = 8.38$, $p < .001$), with no significant difference between them ($t(9) = -0.65$, $p = .529$). To conclude, in agreement with the interpretation of Zylberberg et al. (2012), we observed no positive evidence bias in discrimination responses, even when positive and negative evidence were defined with respect to the stimulus itself.

We then turned to the contribution of motion energy to subjective confidence ratings. The median confidence rating in each experimental session was used to separate all motion energy vectors into four groups, according to decision (chosen or unchosen directions) and confidence level (high or low). Confidence kernels for the chosen and unchosen directions were then extracted by subtracting the mean low confidence vectors from the mean high confidence vectors for both the chosen and unchosen directions. We observed a significant effect of motion energy on confidence within this time window ($t(19) = 2.52$, $p = .021$; see Fig. 5.4, right panels). This effect was significantly stronger for motion energy in the chosen direction, compared to the unchosen direction ($t(9) = 2.81$, $p = .020$). In other words, confidence ratings in the discrimination task were more sensitive to positive evidence than to negative evidence. This is again a successful direct replication of the Positive Evidence Bias observed in Zylberberg et al. (2012).

Detection We next turned to the effects of motion energy on detection responses and confidence ratings. Reverse correlation for detection introduces a challenge: while 'no' responses reflect a belief in the absence of any coherent motion, 'yes' responses can result from three different belief states: participants can detect motion in any of the two directions, or in both. We chose to have two possible motion directions in the detection task in order to prevent participants from making 'no' responses based on significant motion in an unexpected direction. While this choice ensured that participants cannot trivially accumulate evidence for absence, it also made the reverse correlation analysis more difficult, as we did not have full access to participants' beliefs about the stimulus in their 'yes' responses.

As a first approximation, we tested whether sum motion energy along the relevant dimension (horizontal or vertical), regardless of direction (up/down or left/right), affected the probability of a 'yes' response. Sum motion energy did not have a significant effect on participants' responses during the first 300 milliseconds ($t(9) = 1.23$, $p = .249$; see Fig. 5.5, left panel) or at any other time point. The effect of sum motion energy during the first 300 milliseconds on decision confidence was marginally significant ($t(9) = 2.15$, $p = .060$; see Fig. 5.5, right panel). Response-specific effects of sum motion energy on decision confidence were not significant for both responses.

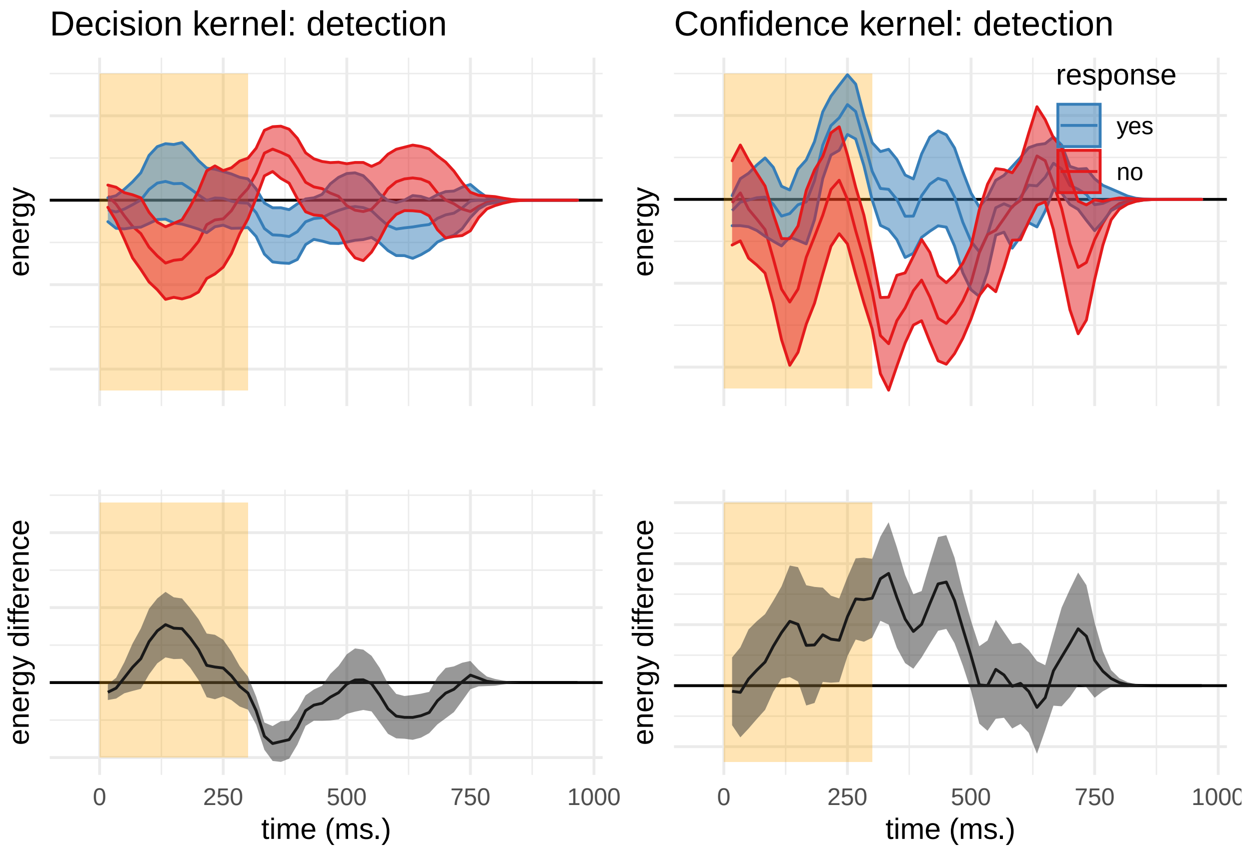

Figure 5.5: Decision and confidence detection kernels, Experiment 1. Upper left: sum motion energy along the relevant dimension in 'yes' (blue) and 'no' (red) responses as a function of time. Lower left: a subtraction between energy in 'yes' and 'no' responses. Upper right: confidence effects for motion energy in 'yes' and 'no' responses. Lower right: a subtraction between confidence effects 'yes' and 'no' responses. Shaded areas represent the the mean +- one standard error. The first 300 milliseconds of the trial are marked in yellow

Detection signal trials

A failure to find significant effects of sum motion energy on detection decision and confidence may be due to the fact that participants were sensitive to relative evidence (e.g., 'more dots are moving to the right') rather than to the sum motion along the relevant axis. However, as we mention above, for any single trial, we cannot tell whether a 'yes' response means 'I perceived coherent motion to the right' or 'I perceived coherent motion to the left'. As a way to approximate participants' perception, we focused on detection signal trials. In these trials, a 'yes' response is most likely to reflect the detection of the true direction of motion. We therefore asked whether fluctuations in the true and opposite directions of motion contributed to detection decision and confidence. This was done by subtracting the motion energy vectors for ‘yes’ and ‘no’ responses in the true and opposite motion directions.

Like discrimination decisions, detection decisions were most sensitive to perceptual evidence in the first 300 milliseconds of the trial (see Fig. 5.6, left panels). However, in contrast to discrimination, a positive evidence bias effect in detection was apparent in the decision itself: when deciding whether a stimulus contained coherent motion, participants were more sensitive to fluctuations in motion energy that strengthened the true direction of motion, in comparison to fluctuations that weakened motion in the opposite direction ($t(9) = 2.31$, $p = .046$).

Motion fluctuations in the first 300 milliseconds of the trial also contributed to confidence in detection ‘yes’ responses (contrasting high and low confidence hit trials; $t(9) = 6.13$, $p < .001$). But unlike in the discrimination task here we found no evidence for a positive evidence bias in confidence ratings ($t(9) = 0.11$, $p = .913$). To reiterate, while detection decisions were mostly sensitive to facilitating fluctuations in motion energy, confidence in detection ‘yes’ responses was equally sensitive to facilitating fluctuations in the true direction of motion, and to interfering fluctuations in the opposite direction of motion. Confidence in ‘miss’ trials was independent of motion energy ($t(9) = 0.16$, $p = .874$). This was true for motion energy in the true direction of motion ($t(9) = 0.12$, $p = .908$) as well as for motion energy in the opposite direction ($t(9) = -0.08$, $p = .941$).

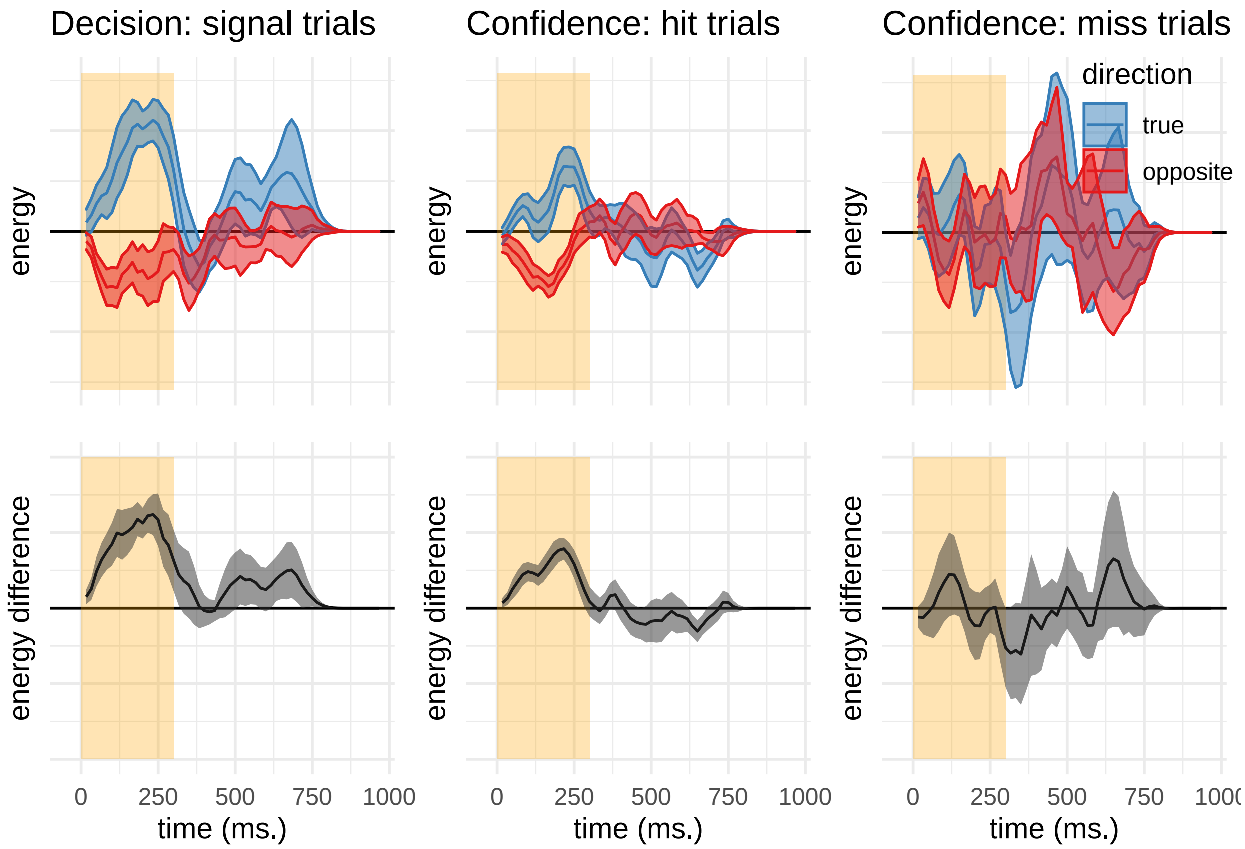

Figure 5.6: Decision and confidence detection kernels in signal trials, Experiment 1. Upper left: difference in motion energy between 'yes' and 'no' responses in the true (blue) and opposite (red) directions as a function of time. Upper middle and right: confidence effects for motion energy in the true and opposite directions for 'yes' and 'no' responses, respectively. Lower panels: the substraction of decision and confidence kernels for the true and opposite directions. Shaded areas represent the the mean +- one standard error. The first 300 milliseconds of the trial are marked in yellow

5.3 Experiment 2

In Exp. 1, we found that detection 'yes' responses are faster and are accompanied by higher subjective confidence than detection 'no' responses. We also replicated the metacognitive asymmetry between detection 'yes' and 'no' responses as measured with response-conditional ROC curves.

Examining random fluctuations in motion energy, we replicated the positive evidence bias in discrimination confidence, such that evidence in support of a decision was given more weight in the construction of confidence than evidence against it. This is consistent with the proposal that participants adopt a detection disposition when rating their confidence in discrimination responses. In detection, decision and confidence were sensitive to fluctuations in motion energy at around the same time window as in discrimination. However, unlike discrimination, in detection a positive evidence bias was apparent in the decision, but not in the confidence kernels. Equal weighting of positive and negative evidence suggests that participants were rating their confidence not in the presence of a signal, but in its category. Furthermore, confidence in detection 'no' responses was not affected by fluctuations in motion energy.

In Experiment 2 we tested the robustness of these findings to a different type of stimuli (flickering patches) and mode of data collection (a ~10 minute online experiment). Specifically, our pre-registered objectives (see our pre-registration document: https://osf.io/8u7dk/) were to first, replicate the positive evidence bias in discrimination, second, replicate the absence of a positive evidence bias in detection confidence ratings, and third, replicate the absence of an effect for positive or negative evidence on confidence in 'no' judgments.

5.3.1 Methods

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). 147 participants were recruited via Prolific, and gave their informed consent prior to their participation. They were selected based on their acceptance rate (>95%) and for being native English speakers. Following our pre-registration, we aimed to collect data until we had reached 100 included participants based on our pre-specified inclusion criteria (see https://osf.io/8u7dk/). Our final data set includes observations from 102 included participants. The entire experiment took around 10 minutes to complete. Participants were paid £1.25 for their participation, equivalent to an hourly wage of £7.5.

Experimental paradigm

The experiment consisted of two tasks (Detection and Discrimination) presented in separate blocks. A total of 56 trials of each task was delivered in 2 blocks of 28 trials each. The order of experimental blocks was interleaved, starting with discrimination.

The first discrimination block started after an introduction section, which included instructions about the stimuli and confidence scale, four practice trials and four confidence practice trials. A second introduction section was presented before the second block. Introduction sections were followed by multiple-choice comprehension questions, to monitor participants' understanding of the main task and confidence reporting interface. To encourage concentration, feedback was given at the end of the second and fourth blocks about overall performance and mean confidence in the task.

Importantly, unlike the lab-based experiment, there was no calibration of difficulty for the two tasks. The rationale for this is that in Experiment 1 participants' perceptual thresholds for motion discrimination were highly similar, and staircasing took a long time to converge. Furthermore, in Exp. 1 we aimed to control for task difficulty, but this introduced differences between the stimulus intensity in detection and discrimination.

To complement our findings, here we aimed to match stimulus intensity between the two tasks, and allow for differences in task performance.

Trial structure In discrimination blocks, trial structure closely followed Experiment 2 from Zylberberg et al. (2012), with a few adaptations. Following a fixation cross (500 ms), a rapid serial visual presentation (RSVP) was be presented (12 frames, presented at 25Hz), consisting of two sets of four adjacent vertical gray bars, displayed to the left and right of the fixation cross (see Fig. 5.7). On each frame, the luminance of the bars was randomly sampled from a Gaussian distribution with a standard deviation of 10/255 units in the standard RGB 0-255 coordinate system. The average luminance of one set of bars was that of the background (128/255). The average luminance of the other set was 133/255, making this patch brighter on average. Participants then reported which of the two sets was brighter on average using the 'D' and 'F' keys on the keyboard. After their response, they rated their confidence on a continuous scale, by controlling the size of a colored circle with their mouse. High confidence was mapped to a big, blue circle, and low confidence to a small, red circle. To discourage hasty confidence ratings, the confidence rating scale stayed on the screen for at least 2000 milliseconds. Feedback about response accuracy was delivered after the confidence rating phase.

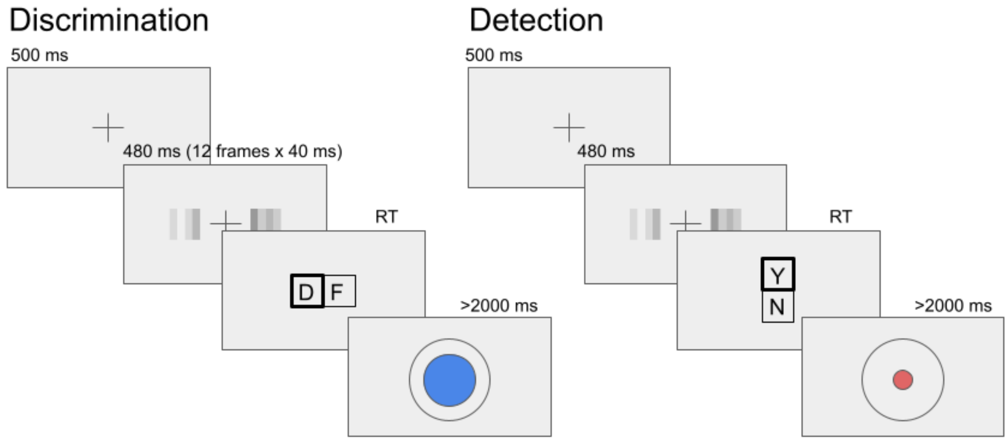

Figure 5.7: Task design for Experiment 2. In both tasks, participants viewed 480 milliseconds of two flicketing patches, after which they made a keyboard response to indicate which of the patches was bright (discrimination) or whether any of the patches was bright (detection).

Detection blocks were similar to discrimination blocks, with the exception that decisions were made about whether the average luminance of either of the two sets was brighter than the gray backgroud, or not. In ‘different’ trials, luminance of the four bars in one of the sets was sampled from a Gaussian distribution with mean 133/255, and the luminance of the other set from a Gaussian distribution with mean 128/255. In ‘same’ trials, the luminance of both sets was sampled from a distribution centered at 128/255. Decisions in Detection trials were reported using the ‘y’ and ‘n keys (‘y’ for ‘yes’ and ‘n’ for ‘no’). Confidence ratings and feedback were as in the discrimination task.

5.3.2 Results

Response accuracy

Overall accuracy level was 0.85 in the discrimination and 0.67 in the detection task. Performance for discrimination was significantly higher than for detection ($M_d = 0.18$, 95% CI [0.16, 0.20], $t(101) = 18.01$, $p < .001$). Unlike in Experiment 1, where we aimed to control for task difficulty, here we decided to match stimulus intensity between the two tasks, so a difference between detection and discrimination performance was expected (Wickens, 2002, p. 104).

Overall properties of response and confidence distributions

Similar to Exp. 1, participants were more likely to respond ‘yes’ than ‘no’ in the detection task (mean proportion of ‘yes’ responses: $M = 0.54$, 95% CI [0.53, 0.56], $t(101) = 4.78$, $p < .001$). We did not observe a consistent response bias in discrimination (mean proportion of ‘right’ responses: $M = 0.50$, 95% CI [0.48, 0.51], $t(101) = -0.62$, $p = .537$).

Participants were also slower to deliver ‘no’ responses compared to ‘yes’ responses (median difference: 77.12 ms, $t(101) = -6.84$, $p < .001$ for a t-test on the log-transformed response times; see Fig. 5.8, upper panel). No significant difference in response times was observed for the discrimination task (median difference: 10.90 ms, $t(101) = -1.40$, $p = .165$).

Confidence in detection was generally lower than in discrimination, consistent with lower accuracy in this task ($M_d = -0.09$, 95% CI [−0.11, −0.07], $t(101) = -8.41$, $p < .001$; see Fig. 5.8, lower panel). Within detection, confidence in ‘yes’ responses was generally higher than confidence in ‘no’ responses ($M = 0.10$, 95% CI [0.07, 0.12], $t(101) = 8.15$, $p < .001$). No difference in average confidence levels was observed between the two discrimination responses ($M = 0.00$, 95% CI [−0.02, 0.02], $t(101) = -0.03$, $p = .974$).

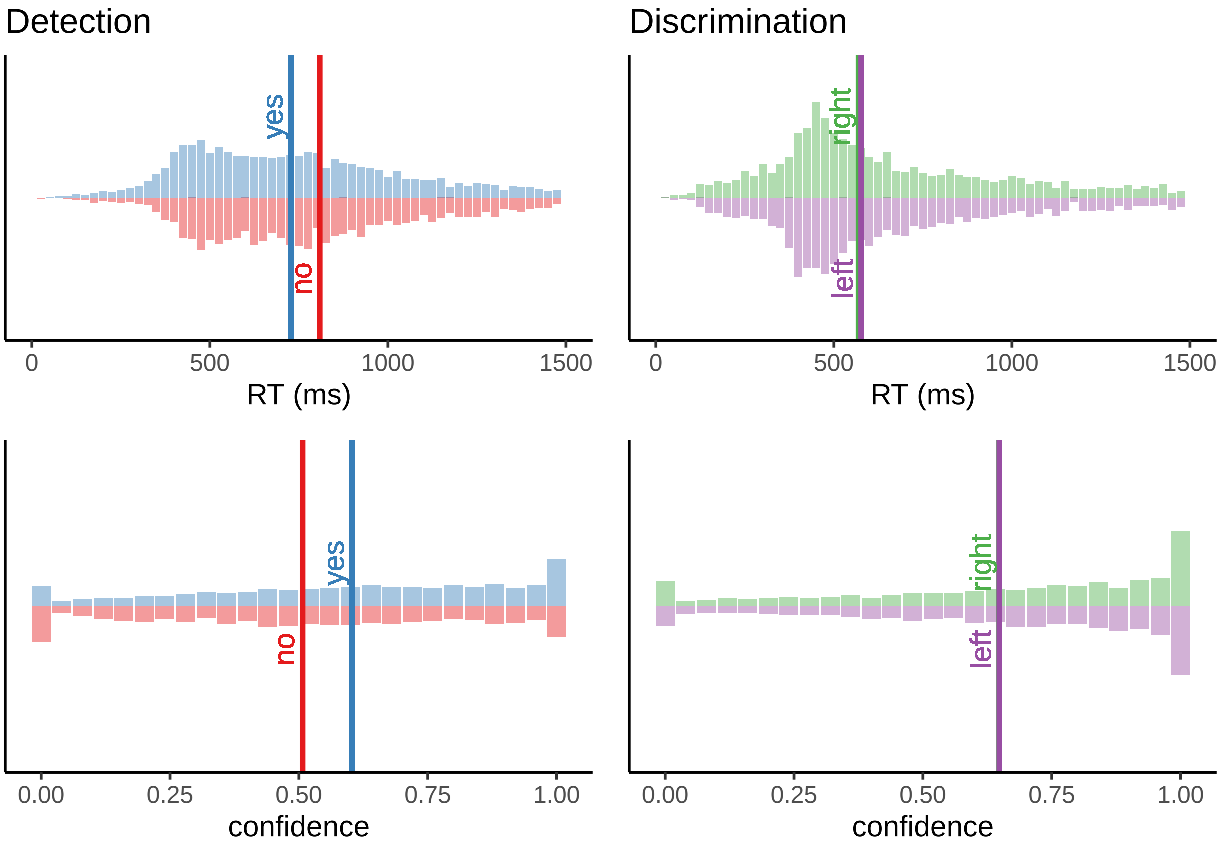

Figure 5.8: Response time (upper panel) and confidence (lower panel) histograms for the detection (left) and discrimination (right) tasks in Experiment 2. Vertical lines represent the median response time and the mean confidence rating for each response.

Response conditional ROC curves

In contrast to the results of Experiment 1, auROC2 for 'yes' and 'no' responses were not significantly different (group difference in area under the response-conditional curve, AUROC2: $M = 0.02$, 95% CI $[-0.02, 0.06]$, $t(58) = 1.13$, $p = .264$; see Fig. 5.9). In the Discussion, we discuss a candidate explanation for this null finding. Importantly, similar metacognitive sensitivity for 'yes' and 'no' responses should not affect the interpretation of our reverse correlation findings.

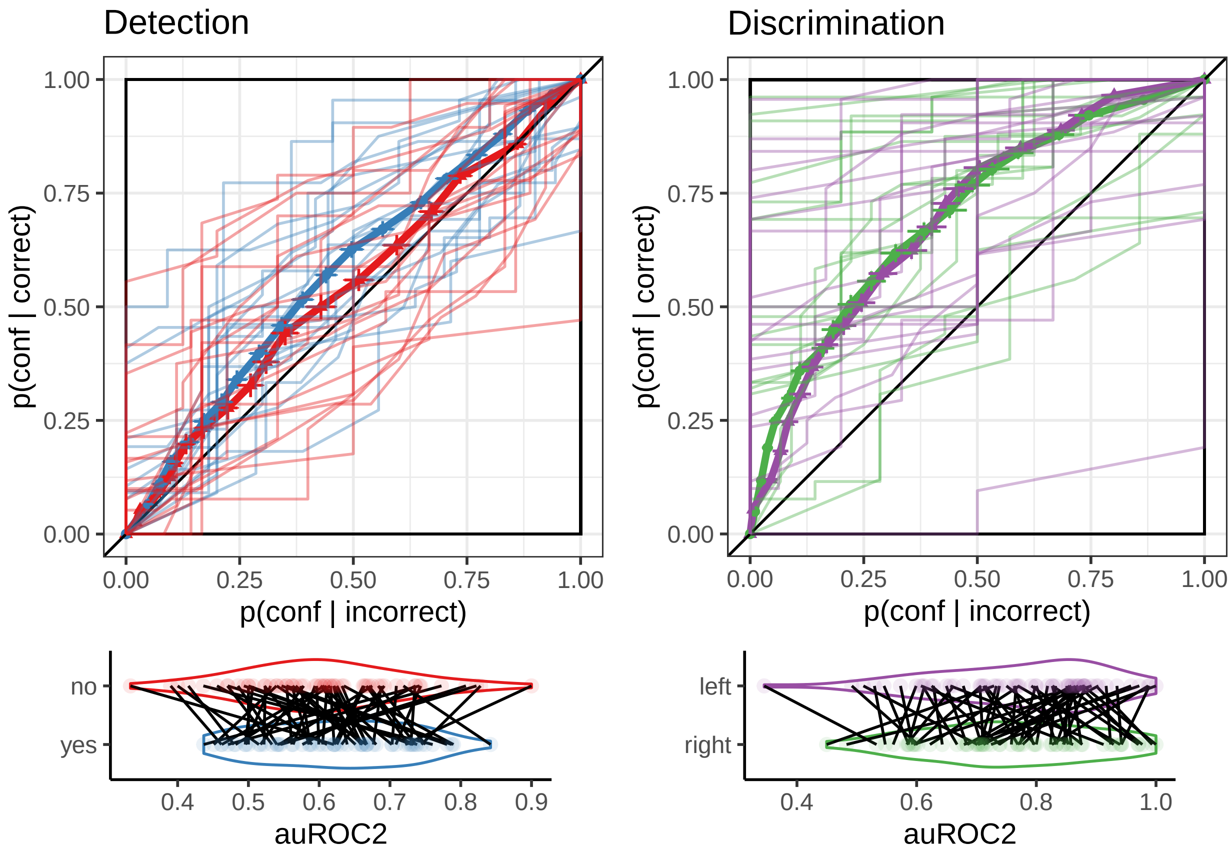

Figure 5.9: Response conditional ROC curves for the two tasks and four responses in Exp. 2. The area under the curve is a measure of metacognitive sensitivity. Lower panel: distributions of the area under the curve for the four responses, across participants. Error bars stand for the standard error of the mean.

Reverse Correlation

Stimuli in Exp. 2 consisted of two flickering patches, each comprising 4 gray bars presented for 12 frames. Together, this summed to 96 random luminance values per trial, which we subjected to reverse correlation analysis, following the analysis of Exp 2. in Zylberberg et al. (2012).

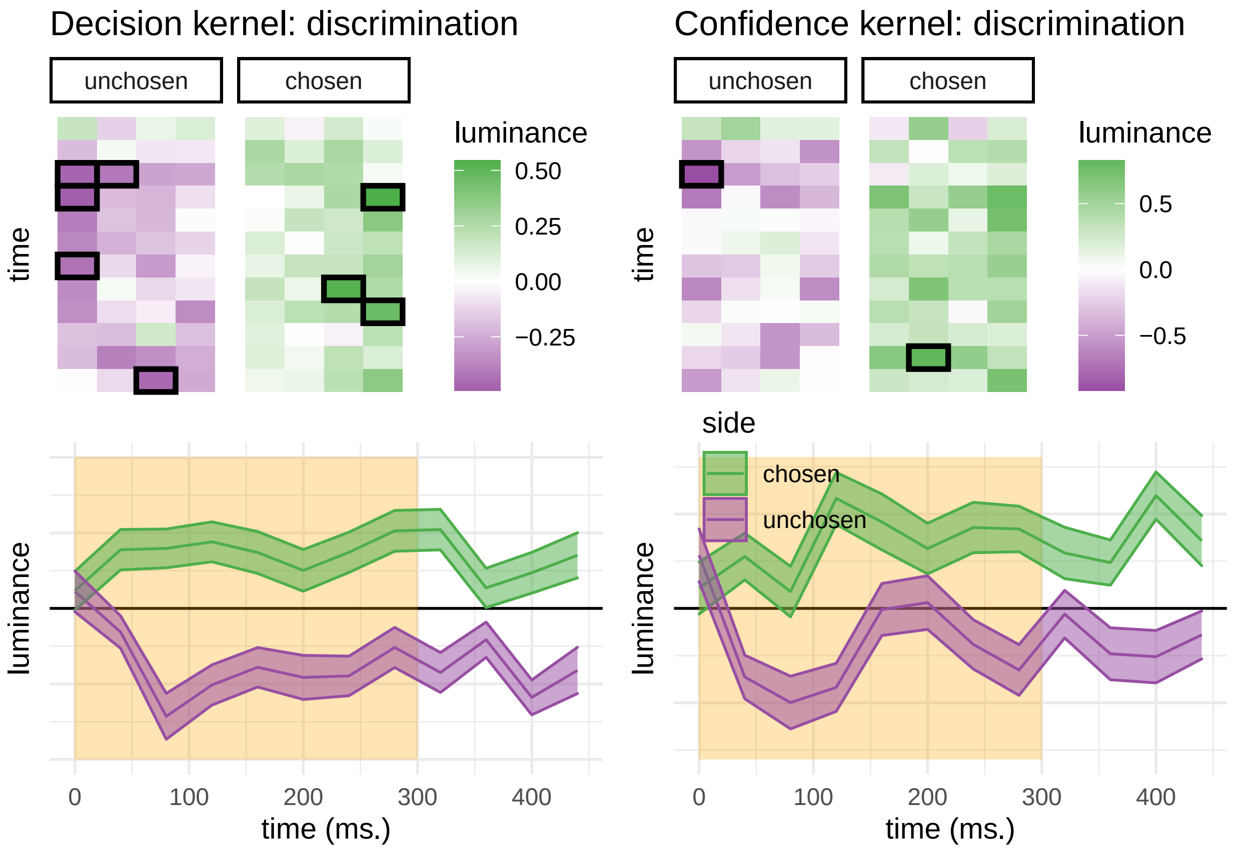

Figure 5.10: Decision and confidence discrimination kernels, Experiment 2. Upper panels: decision (left) and confidence (right) kernels for the flickering patch stimuli. Black frame signify a significant effect at the 0.05 significance level controlling for family-wise error rate across the 48 (12 timepoint x 4 positions) comparisons. Lower panels: decision and confidence kernels, averaged across the four bars to yield a single timecourse for the chosen (green) and unchosen (purple) stimuli. Shaded areas represent the the mean +- one standard error. The first 300 milliseconds of the trial are marked in yellow

Discrimination decisions First, we asked whether random fluctuations in luminance had an effect on participants' discrimination responses. Similar to the results obtained by Zylberberg et. al., discrimination decisions were sensitive to motion energy fluctuations during the first 300 milliseconds of the trial ($t(101) = 10.98$, $p < .001$; see Fig. 5.10, left panels). As per our comment in section 5.2.3, in order to test for decision biases we need to divide evidence not based on participants' decision, but based on the true signal. Participants' decisions were significantly more sensitive to fluctuations in luminance in the foil compared with the signal stimulus within the first 300 miliseconds of the trial ($t(100) = -2.29$, $p = .024$).

Discrimination confidence We observed a significant effect of motion energy on confidence within the first 300 milliseconds of the stimulus ($t(100) = 7.14$, $p < .001$; see Fig. 5.10, right panels). Replicating Zylberberg et al. (2012), this effect was significantly stronger for motion energy in the chosen direction, compared to the unchosen direction ($t(100) = 2.56$, $p = .012$).

Detection decisions We pooled luminance values from both right and left stimuli and contrasted the resulting values as a function of detection response. The sum luminance had a significant effect on participants' responses during the first 300 milliseconds ($t(101) = 6.10$, $p < .001$; see Fig. 5.11, left panel), suggesting that participants were sensitive to sum evidence (overall luminance) in their detection responses.

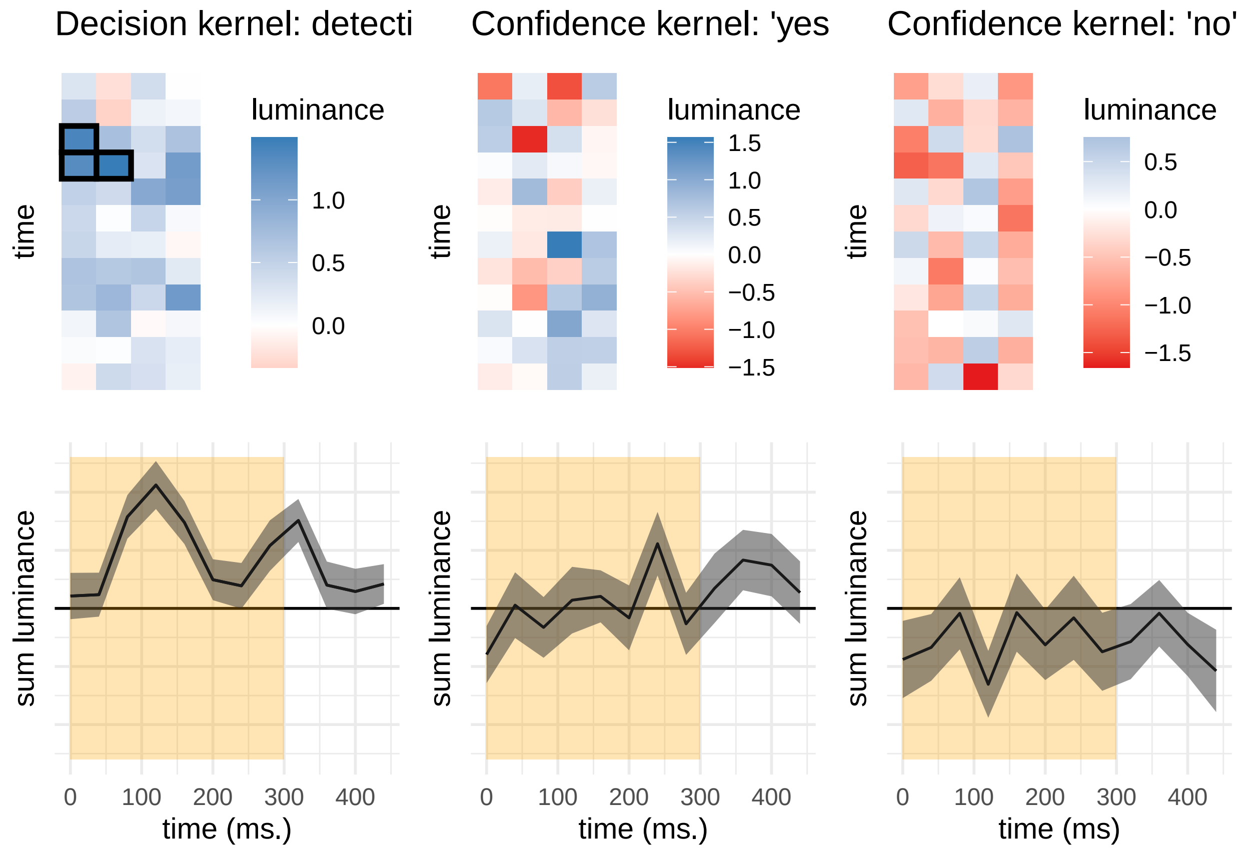

Figure 5.11: Decision and confidence detection kernels, Experiment 2. Upper panels: decision (left) and confidence (right) kernels for the flickering patch stimuli, showing the effect of overall luminance (across both stimuli) on decision and confidence. Black frame signify a significant effect at the 0.05 significance level controlling for family-wise error rate across the 48 (12 timepoint x 4 positions) comparisons. Lower panels: decision and confidence kernels, averaged across the four bars to yield a single timecourse for the difference in luminance effects in 'yes' and 'no' responses. Shaded areas represent the the mean +- one standard error. The first 300 milliseconds of the trial are marked in yellow

We then asked if overall luminance had an effect on decision confidence, such that participants are more confident in their 'yes' responses for brighter displays, and more confident in their 'no' responses for darker displays. Interestingly, and in contrast with our hypothesis, sum luminance had no effect on decision confidence in 'yes' responses ($t(99) = -0.02$, $p = .983$), but had a significant effect on confidence in 'no' responses ($t(99) = -2.43$, $p = .017$; see Fig. 5.11, middle and right panels). As we show below, confidence in 'yes' responses was sensitive to the relative evidence for the two stimulus categories, rather than to the overall luminance of the screen. Our next analysis of detection signal trials diverged from our pre-registered plan. For the pre-registered analysis, see Appendix section 6.4.

5.3.3 Detection signal trials

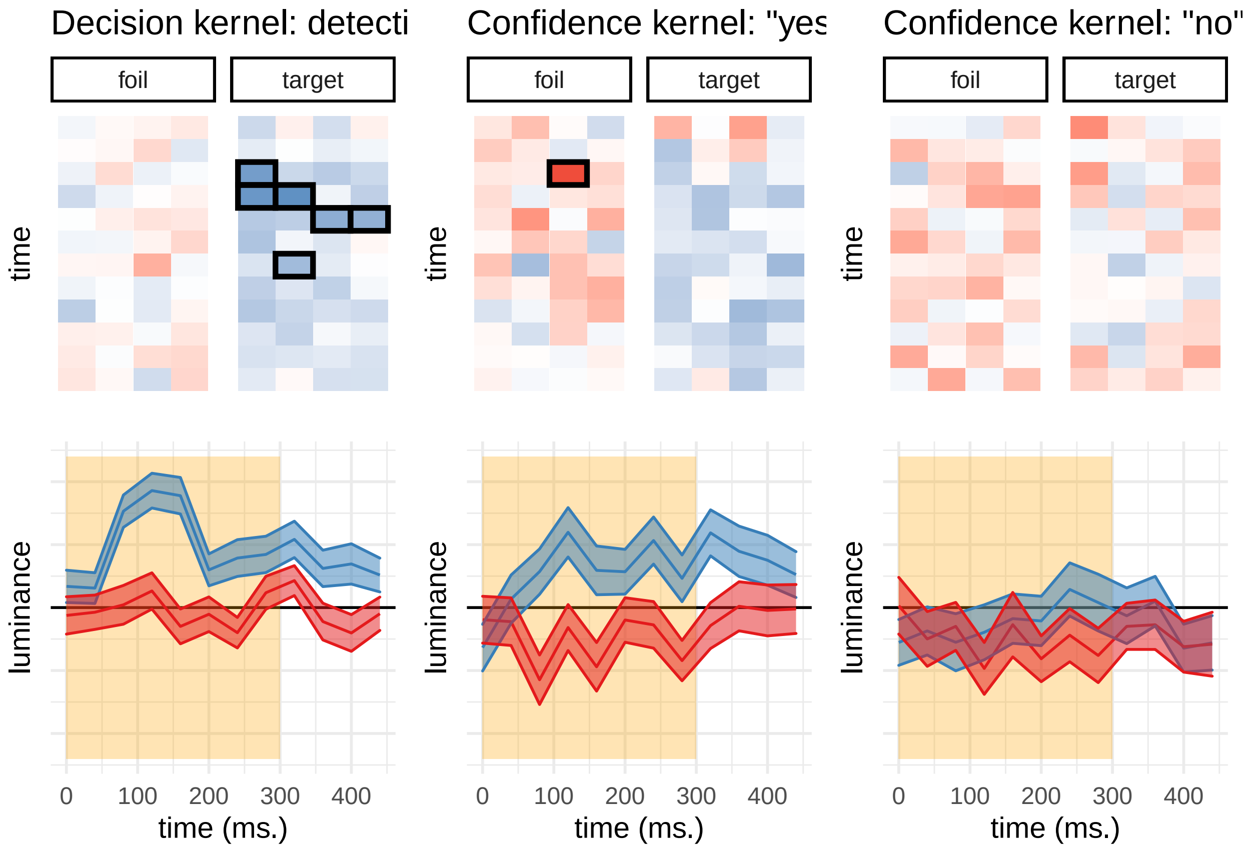

Figure 5.12: Decision and confidence kernels for detection signal trials, Experiment 2. Upper left: mean difference in luminance between ’yes’ and ’no’ responses for the target stimulus and foil stimuli. Upper middle and right panels: mean effect of luminance on confidence in the target and foil stimuli, in ’yes’ and ’no’ responses. Lower panels: the effects of luminance on decision and confidence, averaged across the four spatial locations. Shaded areas represent the the mean +- one standard error. The first 300 milliseconds of the trial are marked in yellow

We next focused on detection signal trials. In these trials, we could separate stimuli to a signal channel (the bright stimulus) and a noise channel (the foil), and ask how random variability in luminance in each channel affected detection decision and confidence. As in Exp. 1, a positive evidence bias effect in detection was apparent in the decision itself: when deciding whether one of the flickering patches was brighter, participants were sensitive to positive noise in the bright patch, but not to negative noise in the foil patch ($t(101) = 6.10$, $p < .001$). Random fluctuations in luminance in the first 300 milliseconds of the trial also contributed to confidence in detection ‘yes’ responses (hit trials; $t(99) = 5.08$, $p < .001$). Similar to the results of Exp. 1, detection confidence was not susceptible to a positive evidence bias ($t(99) = -0.12$,

$p = .901$). To reiterate, while detection decisions were mostly sensitive to facilitating noise, confidence in detection 'yes' responses was equally sensitive to facilitating noise in the target stimulus, and to interfering noise in the foil stimulus.

Consistent with the results of Exp. 1, confidence in 'miss' trials was independent of the contrast in luminance between the right and left stimuli ($t(98) = 1.26$, $p = .210$). However, as described in section 5.3.2, confidence in 'no' responses was sensitive to the overall luminance of the display. A negative effect of luminance on confidence in 'no' responses was significant for the foil stimulus ($t(98) = -2.64$, $p = .010$), and marginally significant for the target stimulus ($t(98) = -1.67$, $p = .099$). Importantly, for both stimuli higher confidence was associated with lower luminance values, consistent with our observation that confidence in detection 'no' responses was based on the overall darkness of the display, rather than on relative evidence.

5.4 Discussion

In two experiments, we compared participants' decisions and confidence ratings in discrimination and detection, matched for difficulty (Exp. 1) and signal strength (Exp. 2). In order to measure the contribution of perceptual evidence to confidence in detection and discrimination confidence ratings, we followed Zylberberg et al. (2012) and applied reverse correlation to noisy stimuli in perceptual decision making tasks. We fully replicated the main results of Zylberberg and colleagues: decision and confidence were affected mostly by perceptual evidence in the first 300 milliseconds of the trial, peaking at around 200 milliseconds. We also successfully replicated the positive-evidence bias: confidence in the discrimination task was more affected by supporting evidence than by conflicting evidence, giving rise to a 'positive evidence bias'. A positive evidence bias in discrimination confidence judgments may indicate that participants adopt a detection disposition in their metacognitive monitoring, and focus on sum evidence rather than relative evidence.

In both experiments, evidence accumulation for detection responses had a similar temporal profile to that of discrimination. However, detection decisions but not confidence ratings showed a positive evidence bias: when making a detection response participants mostly ignored random fluctuations in stimulus energy that were not aligned with the true, presented signal, but these fluctuations were later taken into account when rating their confidence. In both experiments, relative evidence contributed to decision confidence in 'yes' responses, but was ignored in 'no' responses. Finally, in Experiment 2, but not in Experiment 1, sum evidence (the overall luminance of the display) significantly contributed to confidence in 'no' responses. Below we explore the predictions of several Bayes-rational models and their alignment with our observations.

5.4.1 Model 1: a rational agent + symmetric evidence structure

The first model made optimal decisions based on the likelihood ratio between the two hypotheses. This model had full access to the stimulus. Stimuli were modeled as ordered pairs of numbers, corresponding to the two sensory channels (for example, right and left motion, or right and left flickering patch). For simplicity, we ignored the temporal and spatial dynamics of evidence accumulation in our simulations, and focused on the general patterns of evidence weightings instead. In noise trials, both numbers were modeled as sampled from a normal distribution with mean 0 and standard deviation 1 ($E_n \sim \mathcal{N}(0,1)$). In signal trials, one of the two numbers was sampled from a normal distribution with mean 1 ($E_s \sim \mathcal{N}(1,1)$). The agent observes the two numbers, and decides (based on the likelihood ratio, and having full access to the true underlying distributions) if a stimulus was present or not (detection), or which of the two numbers was sampled from the signal distribution (detection). Their confidence is then proportional to the log likelihood ratio between the two hypotheses (signal presence of absence, or signal 1 or 2).

This model makes accurate predictions for the contribution of positive and negative perceptual evidence to discrimination and detection decisions: equal in discrimination, but asymmetric for detection (see Fig. 5.13. However, its predictions for confidence ratings are the exact opposite of what we observe in our data. The model predicts a positive evidence bias in detection confidence ratings, but we find symmetrical confidence kernels for detection confidence. In discrimination, where the model predicts equal contribution of positive and negative evidence to confidence, we find a significant positive evidence bias.

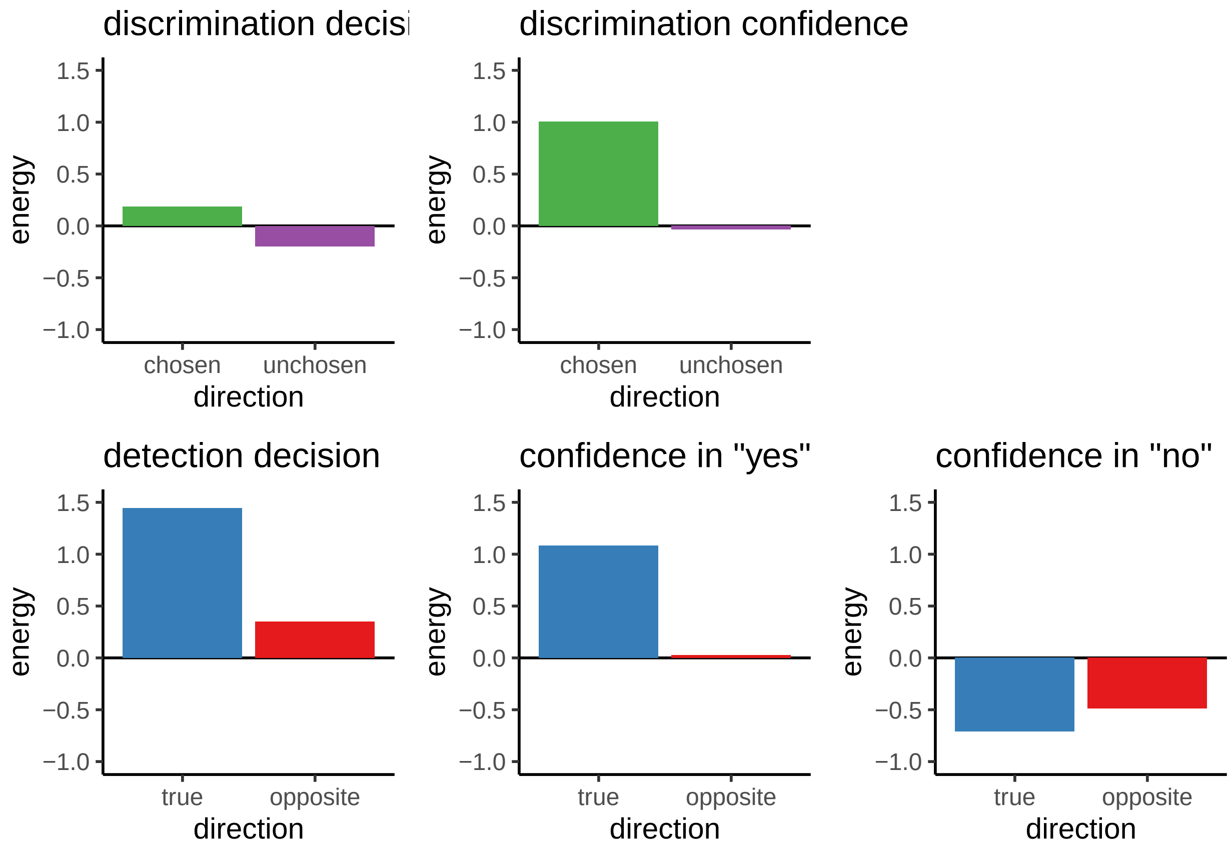

Figure 5.13: Simulated reverse-correlation analysis in Model 1. A bias emerges in detection, but not in discrimination confidence ratings - the opposite of what we observe.

5.4.2 Model 2: a rational agent + symmetric evidence structure

One possible driver of the positive evidence bias in confidence ratings is higher informational value in signal than in noise, such that giving more weight to information from this channel is rational. This is the case in unequal-variance SDT settings, where signal is sampled from a wider range of values than noise. As an example, if noise is sampled from a Gaussian distribution with mean 0 and variance 1 and signal from a Gaussian distribution with mean 2 and variance 3, sampling the value 6 is much more informative than sampling the value -2, because the first is only likely if sampled from the signal distribution (likelihood ratio $>$ 1,000,000), but the second is likely under both distributions (likelihood ratio $=$ 1). Similarly, if the representation of coherent motion is more variable across trials than the representation of random motion, participants would be rational to give more weight to evidence for coherent motion in one channel than evidence for its absence in the other channel.

Higher variability in the representation of signal is often built into the experiment itself. For example, in our Exp. 1, following Zylberberg et al. (2012), the number of coherently moving dots was itself randomly determined, sampled from a Gaussian distribution once in every four frames. This means that there were two sources of variability for the true direction of motion (variability in the direction of randomly moving dots and variability in the number of coherently moving dots), but only one source of variability for the opposite direction (variability in the direction of randomly moving dots). But even when signal is not made more variable by design, the representation of signal is expected to be more variable based on the Weber-Fechner law (Fechner & Adler, 1860) and from the coupling between firing rate and firing rate variability implied by the Poisson form of neuronal firing distributions.

To obtain qualitative predictions, we simulated an unequal-variance first-order SDT model (full simulation details, including the source python code are available in appendix **??**). This model was identical to model 1 with one exception. In this model the artificial agent had access only to a degraded version of the two sensory samples, corrupted by additional noise. To model the unequal variance nature of the perception of signal and noise, this perceptual noise was sampled from a normal distribution with mean 0 and a standard deviation proportional to the magnitude of the sensory sample ($x' = x + \epsilon; \epsilon \sim \mathcal{N}(0, 0.5 \times x)$). The had full knowledge of this generative model for extracting a Log Likelihood Ratio in the process of making a decision and rating their confidence.

This simulation gave rise to a pronounced positive evidence bias in discrimination confidence ratings and in detection decisions (see Fig. **??**. Simulated agents were more sensitive to variations in the signal channel for deciding whether a signal was present or not, and when rating their confidence in discriminating between two stimulus classes. However, in contrast with the observed data, our unequal-variance model also predicted a positive evidence bias in detection confidence ratings and an effect of relative evidence on confidence in 'no' responses, which we do not observe in the actual data.

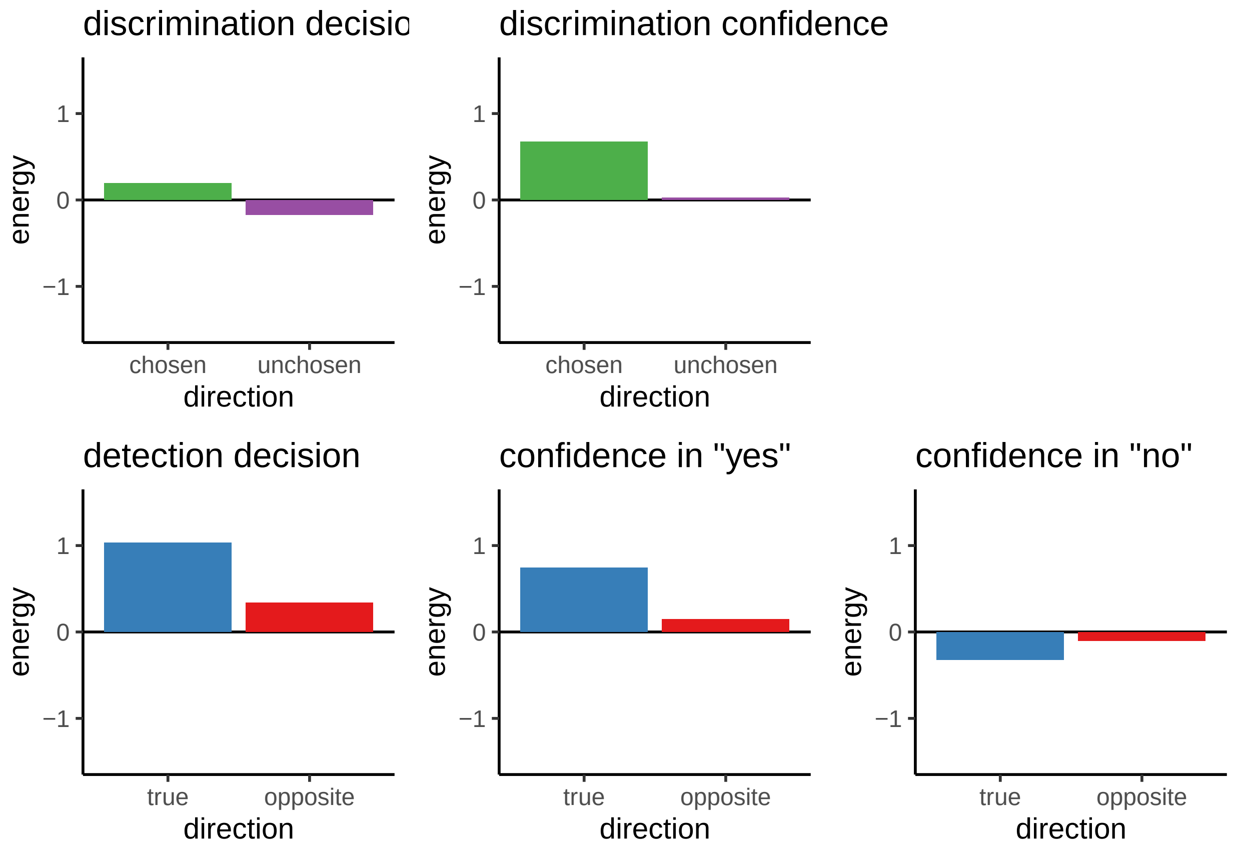

Figure 5.14: Simulated reverse-correlation analysis in Model 2. A bias emerges in detection as well as in discrimination confidence ratings, in contrast to our finding of symmetrical confidence kernels in detection.

5.4.3 Model 3: confidence decision cross

Models 1 and 2 described the behaviour of a rational agent but were unsuccessful in accounting for the mismatch between decision and confidence kernels. Model 3 drops the rationality assumption. This model is identical to Model 1 when it comes to the modeling of perceptual samples and the decision process. However, when coming to rate its confidence in a discrimination judgment, this model extracts the Log Likelihood Ratio not between stimulus category 1 and 2, but between signal presence or absence. Similarly, confidence in discrimination judgments is based on the Log Likelihood Ratio between the presenec of stimulus 1 or 2.

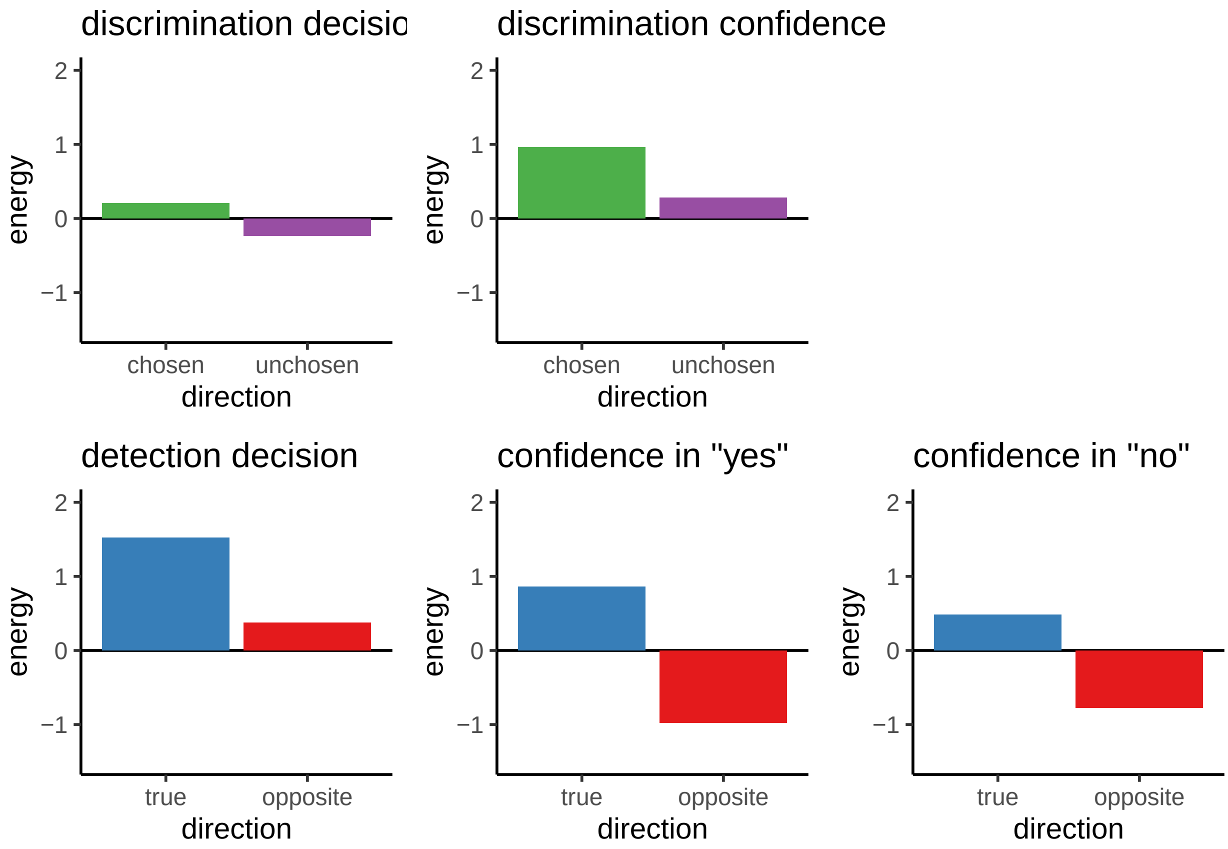

Figure 5.15: Simulated reverse-correlation analysis in Model 3. A positive evidence bias emerges in discrimination confidence ratings, and a negative evidence bias emerges in detection confidence ratings. This is in contrast to our finding of symmetrical confidence kernels in detection.

5.4.4 Evidence for absence

The results of the two experiments were highly similar, with two exceptions. First, the observed metacognitive asymmetry between confidence judgments for detection responses in Experiment 1 was not replicated in Experiment 2. In the second experiment, participants had similar metacognitive insight into their judgments about target presence and absence. Second, in Exp. 1 we found no effect of stimulus energy on confidence judgments in detection ‘no’ responses, whereas in Experiment 2 participants were more confident in the absence of the stimulus when overall stimulus energy was low. We suggest that these two observations may be related, and that the difference may lie in the availability of evidence for absence in the two experiments.

In Exp. 2, signal presence was defined as one of the flickering patches being brighter than the gray background. This meant that participants could be highly confident in the absence of a signal when both stimuli were particularly dark. This is what we observe in our reverse correlation analysis of detection ‘no’ responses (Fig. 5.11 and Fig. 5.12, right panels). In contrast, in Exp. 2 the presence of a signal could mean coherent motion to one of two opposite directions. This means that evidence for absence was never available: the opposite of the presence of rightward motion is leftward motion, not random motion. Indeed, motion energy had no effect on confidence in 'no' responses in Exp. 1 (Fig. 5.5 and Fig. **??**, right panels).

The availability of positive evidence for signal absence may have boosted metacognitive sensitivity for detection 'no' responses in Exp. 2. Interestingly, however, even in Experiment 2, overall confidence in absence was lower than in presence with a similar effect size to that of Exp. 1 (mean differences of 0.08 and 0.10 of the confidence scale in Exp. 1 and 2, respectively), and 'yes' responses were faster on average (median differences -85.37 and -77.12). This may hint to the fact that RT and confidence differences between judgments of presence and absence are unrelated to the informational asymmetry between evidence for presence and for absence.

In summary, in two experiments we replicated the positive evidence bias for discrimination confidence judgments and found a similar bias in detection decisions. A first-order unequal variance framework accounted for this, but failed to account for the absence of a positive evidence bias for confidence judgments in signal presence: participants were more confident in the presence of a signal not only when the true signal was stronger, but also when the opposite signal was weaker. Our findings hint at a qualitative difference in the way subjects evaluate evidence for presence, absence, stimulus class.

In both experiments, detection 'yes' responses were faster on average, and accompanied by higher levels of subjective confidence compared with detection 'no' responses. In contrast, discrimination responses were similar at the group level. These behavioural asymmetries are in line with the classic interpretation of detection responses: 'yes' responses reflect the successful accumulation of evidence for signal presence, and 'no' responses reflect a failure to accumulate such evidence rather than the successful accumulation of evidence for signal absence.

General Discussion

In this thesis I investigated inference about absence in visual perception, and its relation with self-modeling and default-mode reasoning. In chapter 2 I focused on the first few-trials in a visual search task to trace the origins of the metacognitive knowledge that allows subjects to efficiently decide that visual items are missing from a display. In Chapter 5 I used reverse correlation to ask what information is incorporated into confidence judgments in decisions about the presence and absence of a stimulus. Then, in chapter **??** I used functional imaging to compare the neural processes governing metacognitive evaluation of decisions about stimulus type and stimulus presence or absence. Finally, in chapter @(ref:asymmetry) I borrowed ideas from the visual search literature to ask at what cognitive level does the metacognitive asymmetry between judgments of presence and absence emerge.

In what follows I will provide a summary of the results from all chapters. I will then evaluate my original proposal, that inference about absence critically relies on self-knowledge, in light of my findings. Specifically, I will examine alternative interpretations and first-order accounts of inference about absence. Before concluding, I will briefly describe two directions for future research that build on and extend my work here.

5.5 Summary of results

5.6 What I didn't find

5.6.1 Chapter 1: no correlation with explicit metacognition

5.6.2 Chapter 2: no effect of confidence in signal presence

mention project with Roy

5.6.3 Chapter 3: small differences in brain activity between inference about absence and presence

5.6.4

5.7 Future directions

5.7.1 Failures of a self-model

5.8 Conclusion

Chapter 6

Signal Detection Theory

Placeholder

6.1 ROC and zROC curves

6.2 Unequal-variance (uv) SDT

6.3 SDT Measures for Metacognition

6.4 Pseudo-discrimination analysis

6.4.1 Exp. 1

6.4.2 Exp. 2

6.5 Unequal-variance model

6.5.1 Discrimination

Generative model

Inference

6.5.2 Detection

Generative model

Inference

6.6 Confidence button presses

6.7 zROC curves

6.8 Global confidence design matrix

6.9 Effect of confidence in our pre-specified ROIs

6.10 SDT variance ratio correlation with the quadratic confidence effect

6.11 Correlation of metacognitive efficiency with linear and quadratic confidence effects

6.12 Confidence-decision cross classification

6.13 Static Signal Detection Theory

6.13.1 Discrimination

Generative model

Inference

6.13.2 Detection

Generative model

Inference

6.14 Dynamic Criterion

6.14.1 Discrimination

Generative model

Inference

6.14.2 Detection

Generative model

Inference

6.15 Attention Monitoring

6.15.1 Discrimination

Generative model

Inference

6.15.2 Detection

Generative model

Inference

References

Placeholder

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Josa a*, *2*(2), 284–299.

Fechner, G. T., & Adler, H. E. (1860). Elements of psychophysics [elemente der psychophysik]. *Leipzig, Germany: Breitkopf and Ha Rtel*.

Koizumi, A., Maniscalco, B., & Lau, H. (2015). Does perceptual confidence facilitate cognitive control? *Attention, Perception, & Psychophysics*, *77*(4), 1295–1306.

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: Meta-d', response-specific meta-d', and the unequal variance sdt model. In *The cognitive neuroscience of metacognition* (pp. 25–66). Springer.

Mazor, M., Moran, R., & Fleming, S. (2021). Stage 1 registered report: Metacognitive asymmetries in visual perception.

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.

Rausch, M., Hellmann, S., & Zehetleitner, M. (2018). Confidence in masked orientation judgments is informed by both evidence and visibility. *Attention, Perception, & Psychophysics*, *80*(1), 134–154.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. *Elife*, *9*, e60705.

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

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*, 79.