Experiment 3: pre-registration

Objectives

The objectives of this experiment are to replicate the results of Exp. 2 in a confirmatory factorial design. Specifically, we aim to:

1. Replicate the positive evidence bias in discrimination: stronger effect of positive compared to negative evidence on decision confidence.
2. Replicate the positive evidence bias in detection judgments: stronger effect of positive compared to negative evidence on detection judgments.
3. Replicate the symmetric evidence weighting in detection confidence judgments: similar effect of positive compared to negative evidence on detection confidence judgments.

Participants

Participants will be recruited through Prolific. We will aim for a net sample size of 100 participants who will meet our inclusion criteria for both tasks. Task-wise analysis will be applied to the data of all participants that meet our task-specific inclusion criteria.

Experimental Paradigm

The experiment will follow a similar structure to Exp. 2, with several differences, highlighted in yellow.

The experiment will consist of two tasks (Detection and Discrimination) presented in separate blocks. A total of 56 trials of the dicrimination task and 112 trials of the detection task will be delivered in 6 blocks of 28 trials each (~5 seconds per trial; ~2 minutes and 20 seconds per block). The order of experimental blocks will be interleaved, starting and ending with detection. We include more detection trials in this experiment to allow our analysis to focus on detection hit responses without compromising statistical power.

The first detection block will start after an Introduction section. The introduction will include instructions about the stimuli and confidence scale, four practice trials and four confidence practice trials (~3 minutes). A second introduction section will be presented before the second block (first discrimination block; ~2 minutes). Introduction sections will be followed by multiple-choice comprehension questions, to monitor participants' understanding of the main task and confidence reporting interface. To encourage concentration, feedback will be given at

the end of the second and fourth blocks about overall performance and mean confidence in the task. Overall, the total experiment is expected to take about 20 minutes:

	Duration
Introduction (1+2)	3 minutes
168 trials in 6 blocks	15 minutes
Inter-block feedback and rest	2 minutes
Total	**20 minutes**

There will be no calibration of difficulty for the two tasks.
The experiment will terminate after 2 blocks if accuracy for both of the tasks falls below 55%, or after an incorrect response for one of the two multiple choice questions. In such an occasion, data will not be used for the main analysis.

Trial Structure

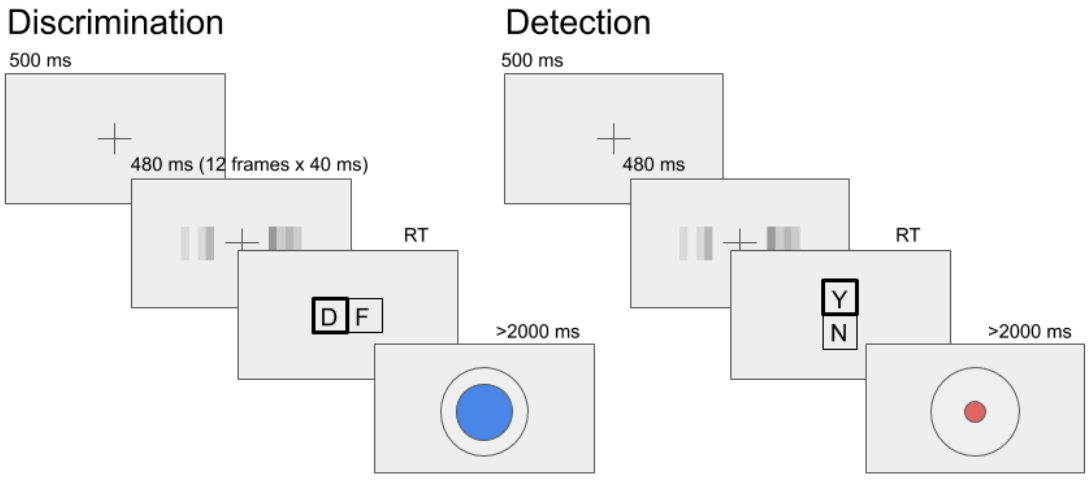

Figure 2: Trial strucure for the discrimination and detection tasks.

Discrimination

Trial structure will closely follow Experiment 2 from Zylberberg and colleagues (2012), with a few adaptations. Following a fixation cross (500 ms), a rapid serial visual presentation (RSVP) will 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. On each frame, the luminance of the bars will be 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 will be that of the background (128/255). The average luminance of the other set will be 133/255, making this patch brighter on average.

Participants will then report which of the two sets was brighter on average than the gray background using the ‘D’ and ‘F’ keys on the keyboard. After their response, they will rate their confidence on a continuous scale, by controlling the size of a colored circle with their mouse. High confidence will be indicated by making the circle bigger, and low confidence will be indicated by making it smaller. To discourage hasty confidence ratings, the confidence rating scale will stay on the screen for at least 2000 milliseconds. Feedback about response accuracy will be delivered after the confidence rating phase.

Detection

Detection trials will be similar to discrimination trials, except that decisions will be 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 will be sampled from a Gaussian distribution with mean 133/255, and luminance of the other set will be sampled from a Gaussian distribution with mean 128/255. In ‘same’ trials, luminance of both sets will be sampled from a distribution centered at 128/255. Decisions in Detection trials will be reported using the ‘y’ and ‘n keys (‘y’ for ‘yes’ and ‘n’ for ‘no’). Confidence ratings and feedback will be as in the discrimination task.

In both tasks, on a random 50% of the trials (*high-luminance trials*), the luminance of both sets of bars will be increased by 2/255 for the entire duration of the display.

	Signal presence	Luminance	Number of trials	Used in contrasts
Discrimination	1	high	28	1,3
Discrimination	1	standard	28	1,3
Detection	1	high	28	2,3,4
Detection	1	standard	28	2,3,4

Detection	1	high	28	2
Detection	1	standard	28	2

Table 2: experimental trials. Overall luminance will be manipulated orthogonally to signal presence and side. The rightmost column refers to the contrasts in the next section.

Analysis

1. Replicate the positive evidence bias in discrimination. This will be achieved by contrasting confidence ratings in discrimination high-luminance trials and standard-luminance trials.
   Power: With 100 participants, we will have 95% statistical power to detect an effect of 0.36 s.d.
2. Replicate the positive evidence bias in detection judgments. This will be achieved by contrasting the proportion of 'yes' responses in detection high-luminance and standard-luminance trials.
   Power: With 100 participants, we will have 95% statistical power to detect an effect of 0.36 s.d.
3. Replicate the weaker positive evidence bias in detection compared to discrimination confidence ratings. This will be achieved by looking at the interaction between luminance (high vs. standard) and task (discrimination correct responses vs. detection hits) on confidence (by performing a t-test on the contrast (discrimination high luminance - discrimination low luminance)-(detection high luminance - detection low luminance)).
   Power: With 100 participants, we will have 95% statistical power to detect an effect of 0.36 s.d.
4. Replicate the absence of a positive evidence bias in detection confidence ratings.
   This will be achieved by extracting a Bayes Factor testing the null hypothesis that the difference in confidence ratings for high vs. standard luminance trials equals zero. We will test this against a null Cauchy distribution with a scaling parameter equal to the standardized effect size observed in discrimination (Hypothesis 1). This reflects a prior belief that the positive evidence bias in detection is expected to be about the same magnitude as in discrimination.

Participant Exclusion

Participant exclusion will be decided separately for the discrimination and detection tasks. Task-wise exclusion will be applied to participants that:

1. Perform with accuracy of below 55%.

2. Use the maximum confidence rating in more than 70% of the trials.

In addition, we will exclude participants from both tasks if they fail one or more of our multichoice task comprehension checks. Data collection will terminate once we obtain full datasets (detection and discrimination) from 100 participants.

Trial Exclusion

Trials with response time below 200 ms will be excluded from all analysis.