Experiment 4: pre-registration

Objectives

The objectives of this experiment are to test the generalizablity of the results from Exp. 2 and 3 to a different stimulus feature (hue, rather than intensity). By using an iso-luminant color space and counterbalancing the color-mapping between participants, we aim to control for nonlinearities in perceived luminance, which may produce differences in the weighting of positive and negative evidence. Specifically, we aim to:

1. <u>Replicate the positive evidence bias in discrimination</u>: a stronger effect of positive compared to negative evidence on discrimination confidence.
2. <u>Replicate the positive evidence bias in detection judgments</u>: a stronger effect of positive compared to negative evidence on detection judgments.
3. <u>Test for the presence of a positive-evidence bias in detection confidence</u>: a stronger effect of positive compared to negative evidence on detection confidence.

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. 3, 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 [detection, discrimination, detection, discrimination, detection, 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. The introduction will include instructions about the stimuli and confidence scale, four practice trials, and four confidence practice trials (~3 minutes). A second introduction will be presented before the second block (first discrimination block; ~2 minutes). Introductions will be followed by multiple-choice comprehension questions, to monitor participants' understanding of the main task and confidence reporting interface. Subjects will be allowed to repeat the multiple choice questions up to three times. To encourage concentration, feedback will be given after every block about overall performance and mean confidence in correct and incorrect responses. 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 if subjects fail to respond correctly within the three allowed repetitions of one of the two multiple choice questions. We will then not use the data from the first two blocks for the main analysis.

Trial Structure

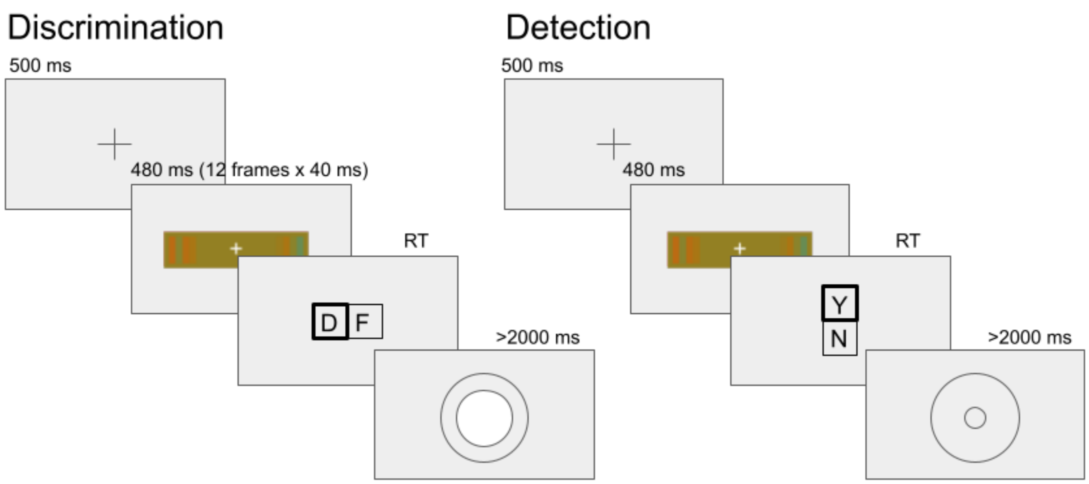

Figure 2: Trial structure for the discrimination and detection tasks.

Discrimination

Trial structure will closely follow Experiment 2 from Zylberberg and colleagues (2012), with a few adaptations. Following a fixation cross (500 ms) rapid serial visual presentation (RSVP) will be presented (12 frames, presented at 25Hz), consisting of two sets of four adjacent vertical colored bars, displayed to the left and right of the fixation cross. On each frame, the hue of the bars will be randomly sampled from a Gaussian distribution in the CIE L*a*b* colour space, centred in L = 54, a = 21.5 and b = 11.5, with a radius of 49 (Schrugin et al., 2020). For half of the participants, the hues in “ordinary patches” will be sampled around an orientation of 1.85, with a standard deviation of 0.35 and the hues in “special patches” will be sampled around an orientation of 2.1 with a standard deviation of 0.35. For the first group, special patches are a little more orange than ordinary patches, and for the second group special patches are a little more green than ordinary patches. To make sure ordinary patches are perceived as the absence of signal relative to the background, the RSVP display will be overlaid on top of a rectangle that will have the mean color of ordinary patches.

Participants will then report which of the two sets was special 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 gray 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 one of the patches was special, or not. In 'different' trials, the hue of the four bars in one of the sets will be sampled from the ordinary distribution, and luminance of the other set will be sampled from the special distribution. In 'same' trials, luminance of both sets will be sampled from the ordinary distribution. 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-evidence trials*), the luminance of both sets of bars will be shifted in the direction of the "special" distribution by 0.1 radians.

Analysis

1. Replicate the positive evidence bias in discrimination. This will be achieved by contrasting confidence ratings in discrimination high-evidence trials and low-evidence 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-evidence and low-evidence trials.
   Power: With 100 participants, we will have 95% statistical power to detect an effect of 0.36 s.d.
3. Compare positive evidence bias in detection compared to discrimination confidence ratings. This will be achieved by looking at the interaction between evidence (high vs. low) and task (discrimination correct responses vs. detection hits) on confidence (by performing a t-test on the contrast (discrimination high evidence - discrimination low evidence)-(detection high evidence- detection low evidence)).
   Power: With 100 participants, we will have 95% statistical power to detect an effect of 0.36 s.d.
4. Test for a positive evidence bias in detection 'yes' responses confidence ratings.
   This will be achieved by extracting a Bayes Factor testing the null hypothesis that the difference in confidence ratings for high vs. low evidence 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 an 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 upon the last try. Data collection will terminate once we obtain full datasets (detection and discrimination) from 100 participants, up to a budget of £600.

Trial Exclusion

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

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. Nature human behaviour, 4(11), 1156-1172.