CrossMark

The optimal experimental design for multiple alternatives perceptual search

Nonie J. Finlayson[1] · Kelda Manser-Smith[1] · Annika Balraj[3] · Benjamin de Haas[1,2,4] · D. Samuel Schwarzkopf[1,2,5]

Abstract

Perceptual bias is inherent to all our senses, particularly in the form of visual illusions and aftereffects. However, many experiments measuring perceptual biases may be susceptible to nonperceptual factors, such as response bias and decision criteria. Here, we quantify how robust multiple alternative perceptual search (MAPS) is for disentangling estimates of perceptual biases from these confounding factors. First, our results show that while there are considerable response biases in our four-alternative forced-choice design, these are unrelated to perceptual biases estimates, and these response biases are not produced by the response modality (keyboard vs. mouse). We also show that perceptual bias estimates are reduced when feedback is given on each trial, likely due to feedback enabling observers to partially (and actively) correct for perceptual biases. However, this does not impact the reliability with which MAPS detects the presence of perceptual biases. Finally, our results show that MAPS can detect actual perceptual biases and is not a decisional bias towards choosing the target in the middle of the candidate stimulus distribution. In summary, researchers conducting a MAPS experiment should use a constant reference stimulus, but consider varying the mean of the candidate distribution. Ideally, they should not employ trial-wise feedback if the magnitude of perceptual biases is of interest.

Keywords Perceptual bias · Methods · Decision bias · Response bias · Size perception

Visual abilities and experiences of the environment vary considerably between individuals (Halpern, Andrews, & Purves, 1999), and also covary with interindividual differences in brain structure and function (Genç, Bergmann, Singer, & Kohler, 2014; Genç et al., 2011; Schwarzkopf & Rees, 2011;

✉ Nonie J. Finlayson
nonie.j@gmail.com

[1] Experimental Psychology, University College London, 26 Bedford Way, London, UK

[2] Institute of Cognitive Neuroscience, University College London, London, UK

[3] Institute of Biomedical Sciences, The George Washington University, Washington, DC, USA

[4] Department of Psychology, Justus-Liebig University, Giessen, Germany

[5] School of Optometry & Vision Science, University of Auckland, Auckland, New Zealand

Schwarzkopf & Rees, 2013; Verghese et al., 2014). As well as varying between individuals, perceptual biases vary substantially within the visual field for a single person. For example, when presented in the periphery, a stimulus appears smaller and distorted compared with when it is presented in the central visual field (Anstis, 1998; Helmholtz, 1867; Newsome, 1972).

There is also more recent evidence for idiosyncratic patterns of perceptual biases across the visual field in different observers, including the perception of high-level attributes like facial age and gender as well as lower-level features like aspect ratio, spatial frequency, orientation, colour (Afraz, Pashkam, & Cavanagh, 2010), position (Kosovicheva & Whitney, 2017), and size (Moutsiana et al., 2016; Schwarzkopf & Rees, 2013). Most of this spatial heterogeneity in perceptual bias could be due to 'undersampling' of the visual field by neurons tuned to these stimulus features (Afraz et al., 2010). If more neurons selective for female faces have receptive fields covering the upper visual field, then an androgynous face image may appear more female if presented to the upper than to the lower visual field. Similarly, this hypothesis could explain why identification for eye and mouth images is better when they are presented in locations consistent with where they typically appear under natural viewing

conditions (de Haas et al., 2016). In contrast, for the perception of simple stimulus size, perceptual biases are heterogeneous across the visual field, and research suggests this is likely due to idiosyncrasies in the spatial selectivity of neuronal populations (Moutsiana et al., 2016).

We recently developed a method called multiple alternatives perceptual search (MAPS) to estimate these perceptual biases for the perception of simple stimulus size (Finlayson, Papageorgiou, & Schwarzkopf, 2017; Moutsiana et al., 2016). The main purpose of MAPS was to efficiently measure the spatial heterogeneity of perceptual biases across multiple visual field locations. In each trial, observers are presented with four candidate stimuli, circles of various sizes, each in a different visual field quadrant (see Fig. 1a). At the same time, a reference circle whose size is always constant is presented at fixation. Observers are asked to choose the candidate circle whose size appeared to them most like the size of the reference.

This allows us to model the perceptual bias and discrimination ability (uncertainty) to explain their behavioural responses across the whole experiment. Figure 1b shows how this is done, by fitting a Gaussian tuning curve to describe how similar the candidate stimulus at each location appears to reference the stimulus. The model seeks to predict which stimulus location the observer chose on each trial. The red arrows show the perceived similarity for the current stimulus at that location, which is a function of the actual size (on the x-axis) and the parameters of the Gaussian curve. The central tendency of the Gaussian curve reflects the perceptual bias, which would be at zero (along the dotted grey line) if the observer perceived the stimuli size as appearing identical to the

reference size when they were actually the same size. In the case that the modelled distribution were shifted to the right (as in the upper-left location in this example), the observer would report the stimuli appears identical to the reference size when it is actually larger—meaning that their perceptual bias is that the stimuli appear *smaller*. The spread of the distribution denotes the uncertainty and is a measure of how precisely the observer can discriminate the stimuli at a given stimulus location. In the example in Fig. 1b, the model would predict that the observer chose the stimulus in the upper-right location because this produces the greatest similarity signal (the longest red arrow).

In addition to efficiently measuring perceptual biases across multiple visual locations, the MAPS task design also minimizes the influence of decision factors when estimating perceptual biases. This is important because perceptual biases are by their very nature subjective, complicating any inference that can be drawn from psychophysical experiments about perceptual experience (M. J. Morgan, Melmoth, & Solomon, 2013). Even relatively robust designs for measuring the point of subjective equality (the size of the test stimulus that observers choose the test stimulus on 50% of trials) with two-alternative forced-choice procedures can be skewed by nonperceptual factors that are unrelated to the observer's *actual* perceptual experience of the stimulus. For instance, response bias is the tendency for an observer to respond a certain way, regardless of the stimulus, such as having natural tendency to prefer one finger over the other. More importantly, decision-making or cognitive factors can produce demand effects that also skew the results. This may occur when observers are responding in the way they believe the experimenter wants,
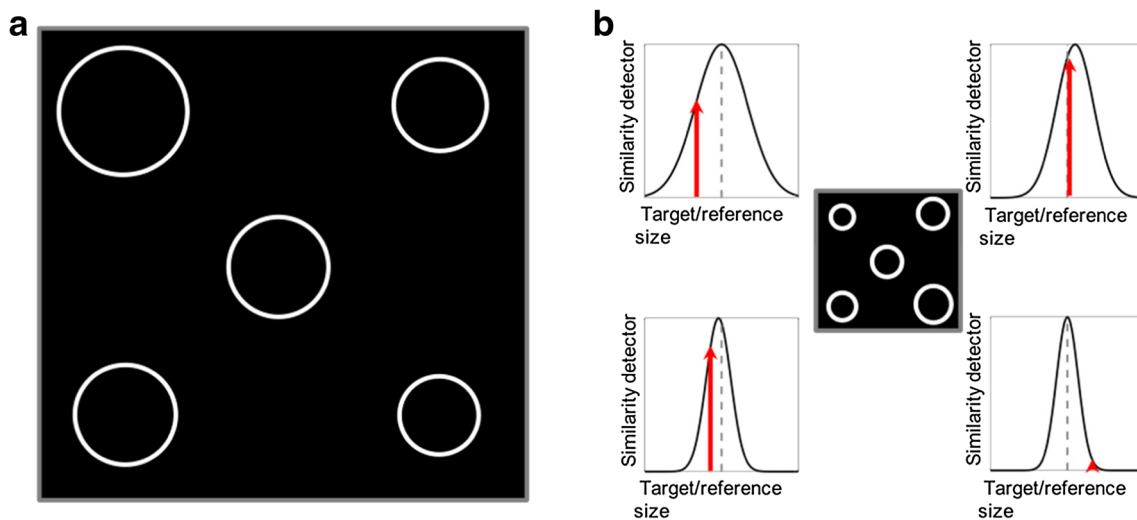

Fig. 1 a Typical stimulus in MAPS design to measure perceptual biases in size judgments across multiple visual field locations (not shown to scale). Central circle is the reference typically kept constant in size. Observers are instructed to fixate inside this circle. Four circles in the visual field quadrants are candidate stimuli. Sizes are varied, although in some versions of the task one candidate is the correct target (i.e., its size is identical to the reference; here, lower-left circle). Model is fit to

estimate perceptual bias and uncertainty (a measure of discrimination sensitivity) at each of the four candidate locations. b Example analysis of behavioural data from MAPS task. Behavioural responses in each trial were modelled by an array of four 'similarity detectors' tuned to stimulus size. Detector showing strongest output to the stimulus (indicated by red arrows) determined predicted behavioural response in each trial (here, the top-right detector would win). (Colour figure online)

or because of the way the response was obtained. An observer may show a tendency to choose one stimuli over another whenever they are unsure of the correct answer (M. Morgan, Dillenburger, Raphael, & Solomon, 2012; M. J. Morgan et al., 2013). Both of these forms of bias could also occur consciously or unconsciously. Simply through changing the decisional criterion, either by (a) instructing observers to favour one response over another when they are unsure of a correct response, or (b) altering feedback given to explicitly enforce a decision bias, the central tendency of an observer's psychometric function changes and would be falsely interpreted as a change in perception. Moreover, this shift in the psychometric curve for cognitive biases occurs without affecting its slope, indicating that introducing decision bias does not compromise discrimination sensitivity on these tasks. Therefore, shifts in psychometric functions are insufficient evidence for actual differences in perceptual appearance.

This illustrates that traditional psychophysical methods, such as the method of single stimuli or the method of constant stimuli, are inadequate for disentangling the actual perceptual experience from the effect of cognitive factors influencing the perceptual decision. To overcome this problem, a number of procedures have been proposed (Finlayson et al., 2017; Jogan & Stocker, 2014; M. J. Morgan et al., 2013; Moutsiana et al., 2016; Patten & Clifford, 2015). While they differ in terms of experimental design and analytical techniques, they all have in common that the observer must choose between several alternatives to select a candidate stimulus that best matches their perceptual experience. Critically, all of the candidates are subject to the same (hypothesized) perceptual effect, thus forcing the observer to make their choice based on the percept. Moreover, these experiments aim to maximize the difficulty of determining the physically veridical choice.

It is important to address considerations such as decision and response bias when purporting to measure perceptual bias with new methods. For example, take the motion aftereffect, where the viewing of a moving visual stimulus for a time then makes a viewed static image appear to move in the opposite direction. When investigated with traditional methods, this effect is not always accompanied by a sensitivity change (merely a 'bias' change), making it difficult to differentiate between the motion aftereffect as a real illusion or just a response or decision bias (M. Morgan et al., 2012). And yet this is an illusion that everyone can see, highlighting the importance of ensuring the new methods actually measure perceptual bias.

In MAPS experiments seeking to quantify spatial heterogeneity of perceptual biases, decision factors are relatively unlikely to skew measurements: observers should have little reason to make different perceptual decisions for each visual field quadrant. However, the same might not be the case when estimating the strength of a visual illusion, such as the Delboeuf illusion as in (Moutsiana et al., 2016): When viewing a visual illusion, observers may not be naïve to the purpose of the experiment and thus be led to exaggerate or underreport the strength of their actual perceptual experience. Moreover, our previous experiments using the MAPS task (Papageorgiou & Schwarzkopf, 2016) revealed that observers have a pronounced response bias: On average, they chose candidates in the right hemifield more often than candidates in the left hemifield (see Fig. 2). These response biases were unrelated to estimates of perceptual biases and their spatial heterogeneity. Nonetheless, it is important to understand if and how response biases affect the perceptual bias inferred from MAPS.

We therefore carried out a series of experiments to test the effects of decision and response bias in MAPS. All experiments used a standard MAPS design to estimate perceptual bias, discrimination uncertainty, and response bias for judgments on simple visual stimuli (Moutsiana et al., 2016), each with small manipulations. First, exploring response bias, Experiment 1 tested whether or not the response biases that we previously observed with MAPS were produced by the response modality (using a keyboard or mouse). The latter two experiments explored decision effects. Experiment 2 tested the effect of trial-wise feedback on the parameter estimates obtained by MAPS. Experiment 3 tested if observers make their perceptual decision based on a tendency to select the middle (or average) of the range of candidate stimuli. To do this, we induced an artificial perceptual bias by modulating the candidate stimuli by a fixed amount and quantified whether MAPS could estimate these artificial biases reliably, enabling us to disentangle a decision bias towards the average versus actual perceptual bias.
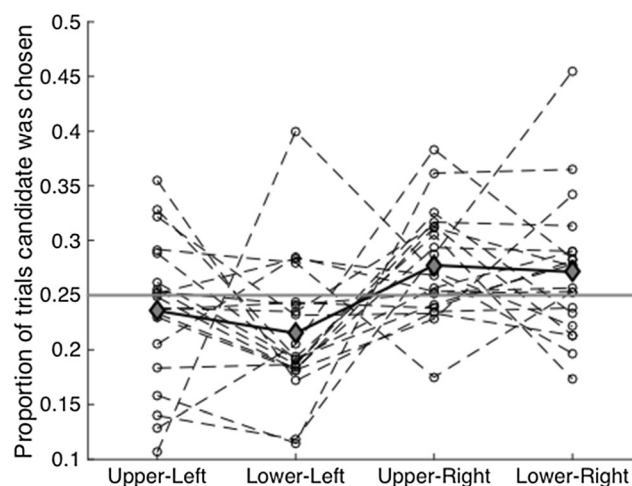

Fig. 2 Response bias in a previous MAPS experiment (Papageorgiou & Schwarzkopf, 2016). Proportion of trials observers selected each of the four candidate locations irrespective of stimulus. Grey filled diamonds and solid black line denote the mean across observers. Open circles and dashed lines denote individual observers

Experiment 1

Because previous MAPS experiments revealed a consistent response bias (Finlayson et al., 2017; Moutsiana et al., 2016), we tested whether this could have been caused by the response modality. The experimental setup was the same as in previous MAPS tasks (Moutsiana et al., 2016), in which observers are asked to choose which of four eccentric candidate stimuli appeared most like a central reference. One of the four candidates was the target, which was the same size as the reference. Previously, observers were asked to make their perceptual decision (which of four candidate stimuli appeared most like the reference) by pressing one of four buttons on a computer keyboard. For this they used both of their hands, and assigned the index and middle fingers of each hand to the two candidate stimuli in each visual hemifield. This procedure is relatively convenient for participants—however, it could be susceptible to biases related to handedness. For instance, a right-handed person might respond more often with their right hand. Similarly, participants might respond more readily with their middle finger than the index finger. Taken together, this could explain why observers tended to choose candidates in the right hemifield, in particular, those in the upper-right quadrant, more often than other candidates. We therefore compared the results of a MAPS experiment when the observer uses either this standard keyboard procedure to make their perceptual decision or when they used the mouse with their dominant hand.

Methods

Participants

Ten observers (four authors; six female; one left-handed; ages 20–38 years) took part in this experiment. All observers were healthy and had normal or corrected-to-normal vision. The University College London Research Ethics Committee approved all procedures, and observers gave written informed consent prior to participating in this and all following experiments.

Stimuli

The experiment took place in a dark, sound-proof room, where observers were seated in front of a computer screen (Samsung 2233RZ) with a resolution of 1680×1050 pixels and a refresh rate of 120 Hz. Minimum and maximum luminance values were 0.25 and 230 cd/m^2. Head position was held at 48 cm from the screen with a chin rest.

All stimuli used in these experiments were created and displayed using MATLAB (Version 8.5; MathWorks, Natick, MA) and the Psychophysics Toolbox (Version 3; Brainard, 1997; Pelli, 1997). In all four experiments, the stimuli presented comprised light grey (54 cd/m^2) circle outlines presented on a black background. Each stimulus array was made up of five circles (see Fig. 1a): the reference, which was a constant size for all trials (diameter: 0.98°), and was presented in the centre of the screen, and the candidates (the remaining four stimuli), which varied in size independently on every trial. They were presented at the four diagonal polar angles from the reference, with their centres at a distance of 3.92° from fixation.

The size of three of the candidates (diameters of the candidate relative to the diameter of the reference) was chosen from a logarithmic Gaussian distribution centred on zero (diameter equal to the reference). The standard deviation of the Gaussian distribution was 0.3 log units. One candidate, chosen pseudorandomly on each trial, was the 'correct' target, and its diameter was equal to the reference.

Procedure

A blue fixation dot (diameter: 0.2°) was shown in the middle of the screen on a black background for 500 ms at the start of each trial. The stimulus array was then presented for 200 ms, after which the fixation dot was shown. In half the runs of the experiment, observers used both their right and left hand to make their responses by pressing buttons on the keyboard, while in the other half they used the mouse with their dominant hand. In keyboard runs, observers were instructed to make their response by pressing the F, V, K, or M button on the keyboard—corresponding to the four candidates on the screen—which appeared most similar in size to the reference. In mouse runs, observers were instructed to make their response by using the mouse to click at the location where the candidate most similar to the reference had appeared. The mouse pointer was hidden from observers while the stimuli were shown and always appeared centred in the screen when observers could make their response.

After observers had made their response, there was a 'ripple' effect over the target they had chosen—this was made up of three 50-ms frames in which a circle increased in diameter from 0.49° in steps of 0.33° and in luminance. Additional feedback indicated during this 150 ms if observers selected the candidate stimulus that was actually closest in size to the reference stimulus. The fixation dot was green and slightly larger (0.33°) for correct trials and did not change for incorrect trials.

Accuracy (the proportion of trials observers selected the candidate that was the same size as the reference) was on average 48.1% correct for each observer (range 43.4%–51.6%). Although this is above a chance performance of 25%, it means that the task was challenging as observers still frequently made mistakes. This is critical for the MAPS model to fit reliable parameter estimates.

The experimental run comprised 10 blocks of 20 trials each, with the order of the experimental runs using the keyboard or mouse counterbalanced across observers. Observers could start blocks by pressing any button on the keyboard or mouse, respectively. Eight observers completed six runs. Two observers only completed four runs. There was a resting break after the observer completed each run. Between runs, a message on the screen told observers how many runs they had already completed and reminded them of the task and responses.

Analysis

Perceptual biases were estimated by fitting a model to predict any given observer's behavioural response on each trial as previously described (Moutsiana et al., 2016). The model typically predicts the observers' responses with above 50% correct. While this still means many trials are predicted incorrectly (presumably when several candidates are very similar), this is considerably above the chance level of 25%, and repeated experiments show that these perceptual bias estimates are very reliable at repeated tests.

In brief, as seen in Fig. 1b, a Gaussian tuning curve was modelled for each candidate stimulus location, representing an array of four 'neural similarity detectors' (or 'comparators') tuned to stimulus size. Each similarity detector represented the perceptual similarity of the candidate to the reference: The higher the output of the detector, the more likely is it that the candidate stimulus at that location appears the same as the reference. Therefore, for each trial, the similarity detector with the maximal output signals the behavioural choice. The centroid of this Gaussian tuning curve indicates the stimulus size where the candidate stimulus appears identical to the reference, taken to be the perceptual bias. The width (standard deviation) of the Gaussian curve denotes the uncertainty, the reciprocal of discrimination sensitivity for that location. The procedure fits these eight parameters simultaneously (perceptual bias and uncertainty for each of the four candidate locations) to maximize the number of trials that the model could correctly predict the observer's actual behavioural response. To avoid local minima, the model fitting started with the mean across all incorrect trials (when the true target was not chosen) as an approximation for the perceptual bias at each location, and the standard deviation across all choices for a given location as an approximation of the uncertainty. Separately from the model fitting, we quantified the general response bias, the proportion of trials that an observer chose a given candidate location irrespective of the actual stimulus.

Results

Figure 3a shows the response bias measured for each of the four candidate locations. As in our previous experiments

(Papageorgiou & Schwarzkopf, 2016), participants showed a bias towards selecting stimuli on the right-hand side over stimuli presented on the left-hand side of the reference stimuli. More specifically, observers chose the upper-right candidate most frequently, and the lower-left candidate least frequently. Importantly, observers showed the same pattern of response biases when using both the keyboard and mouse to make their responses. To compare the keyboard and mouse conditions statistically, we subtracted the response bias for the two extreme locations, the upper-right and lower-left quadrant, to calculate a response-bias index. There was no significant difference for this index when using the keyboard or the mouse, $t(9) = -0.21$, $p = .838$. The spatial pattern of response biases was also very similar between the two conditions ($R = .62$, $p < .001$).

Furthermore, it is important to confirm that the parameter estimates for perceptual bias and uncertainty were consistent under the two response modalities. We tested the magnitude of perceptual biases (see Fig. 3b) and uncertainties (Fig. 3c) by averaging them across the four candidate locations and then calculating a paired t test between the two response modalities. There was no significant difference between response modalities, either for perceptual biases, $t(9) = 0.85$, $p = .418$, or uncertainties, $t(9) = 0.58$, $p = .575$.

Next we tested the similarity between the two response modalities by calculating the linear correlations between them after removing the between-subject variance. The pattern of perceptual biases (see Fig. 3d) was strongly correlated when using the keyboard or the mouse ($R = .57$, $p < .001$), indicating that bias was consistent when using both input methods. However, the uncertainties (Fig. 3e) were not significantly correlated between response modalities ($R = .06$, $p = .716$). This indicates that the spatial pattern discrimination ability estimates were not consistent when using the two response modalities. Finally, we also tested the similarity between response and perceptual biases: These two measures were uncorrelated (keyboard: $R = -.12$, $p = .450$; mouse: $R = -.07$, $p = .669$).

Discussion

Experiment 1 showed that the response modality to make a perceptual decision, using either the keyboard or the mouse to select the observer response, did not affect the pattern of response biases. Furthermore, this pattern of results was no different when the one left-handed participant was removed from analyses. This suggests that previous findings—that observers select candidate stimuli in the upper-right quadrant more frequently than the others, especially those in the lower left—is not merely caused by an effect of handedness. The majority of participants tested using MAPS, in the series of experiments presented here and previous experiments (Moutsiana et al., 2016; Papageorgiou & Schwarzkopf, 2016), were right-handed. It could be argued that using both hands to respond using the keyboard biases right-side responses because this
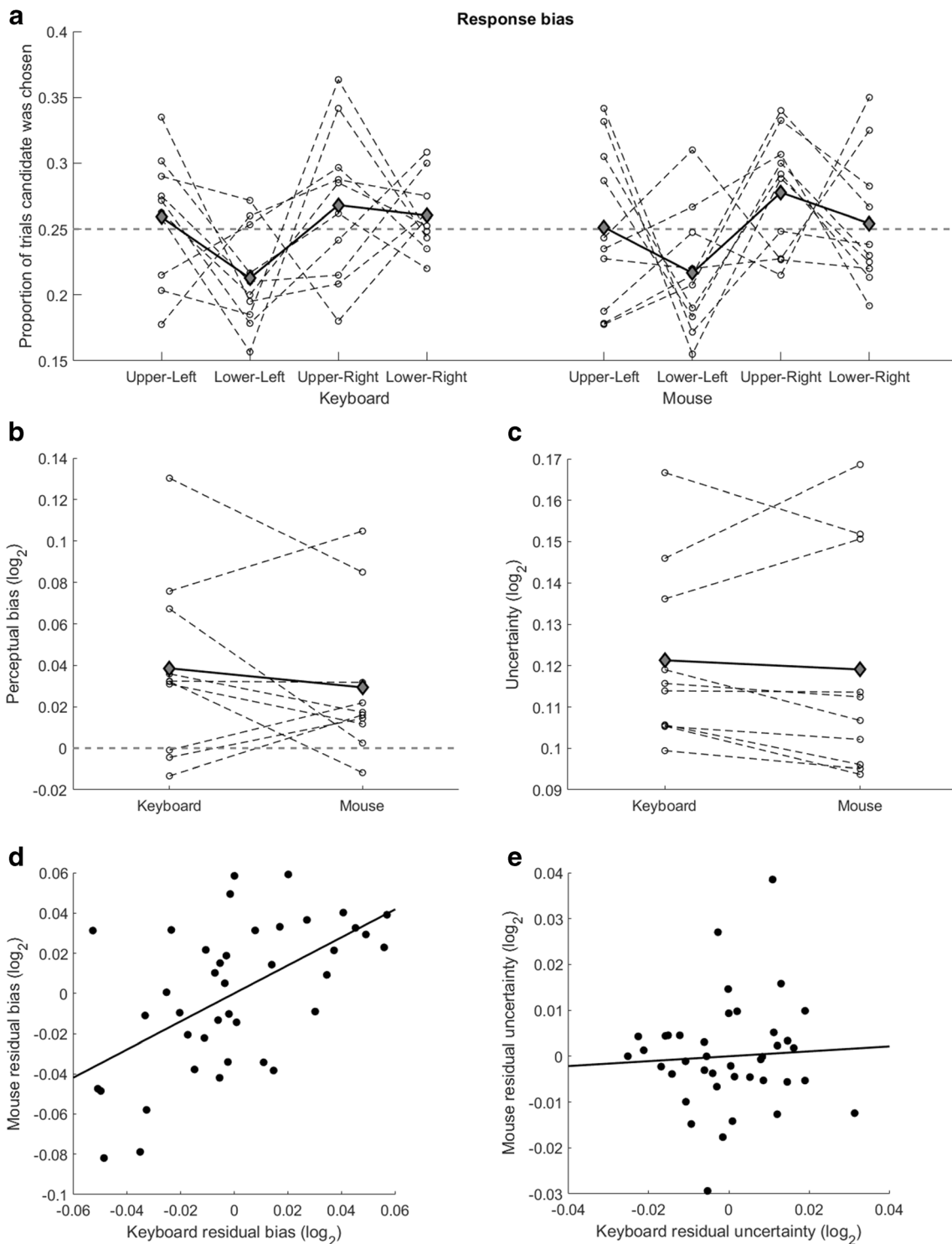
Fig. 3 Results of Experiment 1. **a** Response bias plotted for each of the four candidate stimuli, separately for when observers used the keyboard or the mouse to report their response. Pattern of response biases is very similar regardless of the response modality. **b–c** Perceptual bias (**b**) and uncertainty (**c**) plotted for two response modalities. In **a–c**, grey filled diamonds and solid black line denote the mean across observers. Open circles and dashed lines denote individual observers. **d–e** Perceptual biases (**d**) and uncertainty (**e**) when observers used the mouse plotted against the corresponding value when they used the keyboard. Black circles denote individual locations and observers (after subtracting the mean across candidate locations for each individual observer to remove between-subject variance). Black line indicates best-fitting linear regression

minimizes effort for the observer. However, our results now indicate that this response bias occurs regardless of whether only one or both hemifields are associated with responses using the dominant hand.

Critically, both the magnitude and spatial pattern of perceptual biases were unaffected by the response modality, and we replicated our previous results (Papageorgiou & Schwarzkopf, 2016) showing that there was no relationship between the spatial pattern of response biases and perceptual biases (see Fig. 2). The overall uncertainty was also unaffected by the response modality. This argues against a task difficulty difference between the two response modalities, as we would predict a harder task would widen the overall spread of responses.

Experiment 2

Experiment 1 shows that the MAPS perceptual bias is unlikely to be due to response bias, but what about decision bias? Previous research has shown that providing feedback on each trial could influence an observer's decision criterion (M. Morgan et al., 2012). We used trial-wise feedback in all our MAPS experiments to date. The four-alternative forced-choice task and the brief stimulus duration are challenging even though observers perform well in excess of the chance level of 25% correct. In fact, for the purpose of measuring perceptual bias it is important that observers make a lot of mistakes, that is, there must be many trials in which they would not report perceiving the correct target as identical to the reference. However, without any feedback about their task performance observers might not be motivated to perform reliably. Nevertheless, it is also possible that because observers received feedback about whether or not they selected the correct target, they might have been conditioned to actively correct for perceptual biases in their choices, and in turn reduce estimates of their perceptual biases. In Experiment 2, we tested whether or not trial-wise feedback influences the results of MAPS.

Method

Participants

Eight observers (four authors; five female; ages 21–38 years) took part in this experiment. All observers were healthy and had normal or corrected-to-normal vision.

Stimuli and procedure

The stimuli and procedures were the same as in Experiment 1, using only the keyboard to respond. Each observer completed two runs of the experiment. In one run, feedback was given on correct trials as in all the other experiments. In the other run, no feedback was given, and the fixation dot stayed blue when observers selected the correct target. The order of these two runs was counterbalanced and randomly assigned across observers. There were 400 trials per run, and these were subdivided into 20 blocks of 20 trials each.

Results

When observers received no feedback, the magnitude of perceptual bias (see Fig. 4a) estimates (averaged across the four locations in each observer) was significantly, $t(7) = -2.80$, $p = .026$, greater ($M = 0.06$) than when they received feedback ($M = 0.01$, a comparable effect to the equivalent condition in our previous experiments here and in other studies, e.g., Experiment 1: $M = 0.02$).

One explanation for this effect of feedback could have been that observers performed worse without feedback, either because they were less motivated or because they lacked any information to monitor their performance. Therefore, we also compared the uncertainties (as a measure of discrimination sensitivity) between conditions (see Fig. 4b). However, discrimination sensitivity was at similar levels regardless of whether or not observers received feedback (feedback: $M = 0.1$; no feedback: $M = 0.1$), $t(7) = 0.04$, $p = .968$. As before, we also tested the similarity of spatial patterns of these parameters between the two conditions. The perceptual biases were highly correlated ($R = .76$, $p < .001$). In contrast, the discrimination sensitivities were unrelated in the two conditions ($R = -.09$, $p = .624$), and there were no differences in response choice between the two feedback conditions ($ps \geq .289$).

Discussion

When giving feedback to observers on each trial about whether or not their choice was the target, estimates of perceptual biases were considerably lower than when observers received no feedback. In contrast, the discrimination sensitivity (expressed as the uncertainty parameter modelled by MAPS) was unaffected by feedback. This result suggests that observers did not simply perform worse without feedback. The reduction of perceptual bias magnitudes with feedback are likely because observers actively corrected for their perceptual biases in the experiment. Observers may have been learning through feedback which sizes in each location were associated with a 'correct' response, and trying to maximize 'correct' feedback rather than following the instructions to base their decision on the perceived closest candidate in size.

Importantly, as in all the other experiments, the spatial pattern of perceptual biases was highly preserved regardless of whether feedback was given. This demonstrates that MAPS could indeed detect the underlying spatial heterogeneity of perceptual biases in each observer. Feedback only affected the overall magnitude of these biases, but not their spatial relationship. Thus, feedback may have indeed introduced a
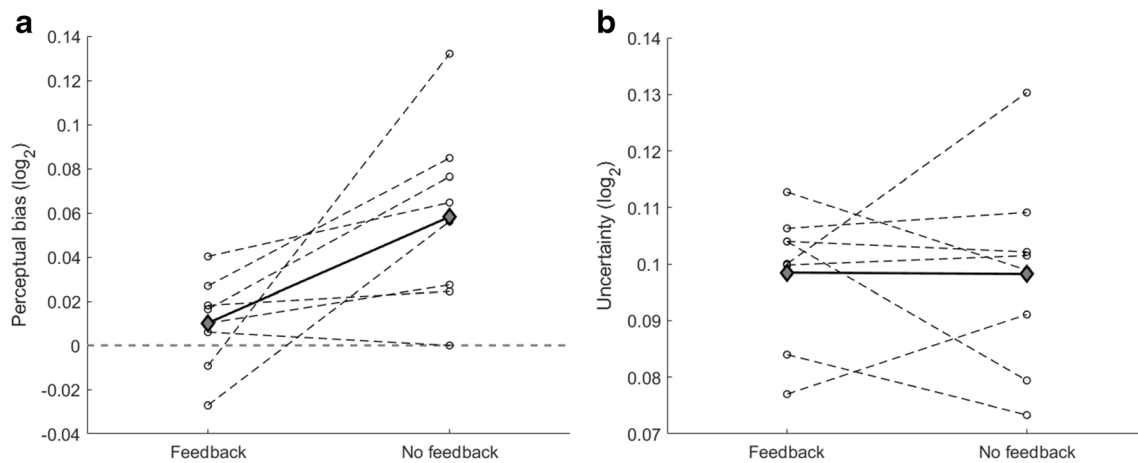
Fig. 4 Results of Experiment 2. Perceptual biases (**a**) and uncertainty (**b**) were plotted separately for runs with or without feedback in each trial. Grey filled diamonds and solid black lines denote the mean across observers. Open circles and dashed lines denote individual observers

decision bias, but this did not cancel out the underlying differences in perceptual biases across the four locations.

Experiment 3

Experiment 3 aimed to test if MAPS is influenced by a particular form of decision bias. In particular, if observers always choose the candidate that appears to be in the middle of the range (or the average) of candidate stimuli on a given trial, they would on average tend to select the correct target. The experimental setup was the same as in previous experiments. However, in some randomly interleaved trials, a fixed offset was either added or subtracted to all the candidate stimulus sizes, including to the target. This effectively shifted the distribution of stimulus values. Critically, we did not add this shift to the independent variable in the MAPS model. Thus, if an observer were to always choose the average candidate, they would now select a candidate that was also either larger or smaller than the reference. Because the shift is not included in the MAPS model, the estimated perceptual biases in this case would remain unchanged. Conversely, if observers adjusted their choice based on their *actual* percept, the estimated perceptual bias should shift to *counteract* the fixed offset. Our analysis was deliberately designed such that evidence for an effect of the shifted distributions supports the interpretation that MAPS measures perceptual experience. Because the main purpose of this experiment was to test the reliability of perceptual bias estimates, we focused our analysis on this parameter. Additionally, to further test the effects of feedback, we conducted two experiments, with (Experiment 3a) and without (Experiment 3b) feedback. Based on Experiment 2, providing feedback should counteract the effects of artificial shifts due to participants actively correcting for their perceptual biases, but we should see the full effects of the artificial shifts in Experiment 3b without feedback.

Method

Participants

Both Experiments 3a and 3b had five observers (3a: two authors, three female, ages 21–38 years; 3b: one author, four female, ages 26–36 years). All observers were healthy and had normal or corrected-to-normal vision. These relatively small sample sizes were justified because this was a conceptual (but simpler) replication of the experiments shown in the Supplementary Information (Experiments S1 and S2) with larger sample sizes. Moreover, in addition to group-level inferential statistics, we also tested the significance of the difference in perceptual biases for each participant separately.

Stimuli and procedure

The stimuli and procedures were the same as in Experiment 1, except as described below. There were three stimulus conditions: In a third of trials, the size of three of the candidates was chosen from a logarithmic Gaussian distribution centred on zero, as in previous experiments. In the other two thirds of trials, the sizes of the four candidate stimuli was reduced/enlarged by subtracting/adding a constant 0.05 log units. This means that the 'target' stimulus that would otherwise have been equal to the reference now had a diameter of 0.95° or 1.01° visual angle, respectively. However, the analysis ignored the changes in the candidate stimuli.

Feedback was similar to that in Experiment 2, with Experiment 3a providing participants with feedback and Experiment 3b providing no feedback to participants. Due to the candidate size shift, there was not always a candidate with the exact size of the reference; therefore, feedback for a 'correct' response was instead given to the candidate closest in size to the reference.

The experimental run comprised 150 blocks of 12 trials each, and the three experimental conditions were randomly interleaved across trials.

Individual observer statistics

In additional to classical inferential statistics, we conducted a statistical analysis of the hypothesized effect for each individual observer. For this we ran a permutation test separately for each of the three experimental conditions in which we shuffled the order of behavioural responses 1,000 times, and then fit the MAPS model to these data. For each of these 1,000 permutations, we then calculated a linear regression to estimate the slope with which perceptual biases changed depending on how the candidate distribution was shifted. We then determined the significance of this relationship by quantifying the proportion of these 1,000 slopes from shuffling that were at least as large as the slope estimated for the actual biases. This analysis ignored the sign of slopes and thus constitutes a two-tailed test whether or not the slope was significantly different from zero.

Results

To test how shifting the distribution of candidate stimuli affected estimates of perceptual bias, we first collapsed and averaged bias estimates across the four locations for each observer. This showed that perceptual biases estimates counteracted the shifts in the candidate stimulus distribution: If candidates were smaller on average than the reference, then relative to the no-shift control condition the perceptual biases estimates were larger, and, vice versa, when candidates were larger on average than the reference, the perceptual bias

estimates were reduced (see Fig. 5). This was supported by a significant main effect of the candidate distribution in a one-way repeated-measures analysis of variance, both with feedback, Experiment 3a: $F(2, 8) = 66.18$, $p < .001$, and without feedback, Experiment 3b: $F(2, 8) = 108.20$, $p < .001$. We also conducted an analysis testing the significance of this relationship for each observer using a permutation analysis. In Experiment 3a, the relationship was significant for all but one of the five observers, probably because that observer showed only a modest difference between the no-shift condition and the condition with larger candidates (S1: $p = .116$, S2: $p = .003$, S3: $p = .007$, S4: $p = .001$, S5: $p = .011$). In Experiment 3b, the relationship was significant for all five observers (S1: $p = .001$, S2: $p = .033$, S3: $p = .002$, S4: $p = .001$, S5: $p = .001$).

To ensure that the estimation of the heterogeneity of perceptual biases was reliable, we also tested how similar the spatial pattern of perceptual biases was between the three experimental conditions: smaller candidates, larger candidates, and control. To do this, we calculated the linear correlation between perceptual biases for different conditions after removing the between-subject variance (i.e., the mean across candidates for each observer was subtracted before calculating the correlation).

For Experiment 3a, perceptual biases measured between the three conditions were all very strongly correlated (smaller vs larger candidates: $R = .81$, $p < .001$; smaller candidates vs control: $R = .75$, $p < .001$; larger candidates vs control: $R = 0.88$, $p < .001$). Uncertainty between the three conditions was also generally correlated, albeit less strongly (smaller vs larger candidates: $R = .41$, $p = .069$; smaller candidates vs control: $R = .54$, $p = .013$; larger candidates vs control: $R = .61$, $p = .004$).
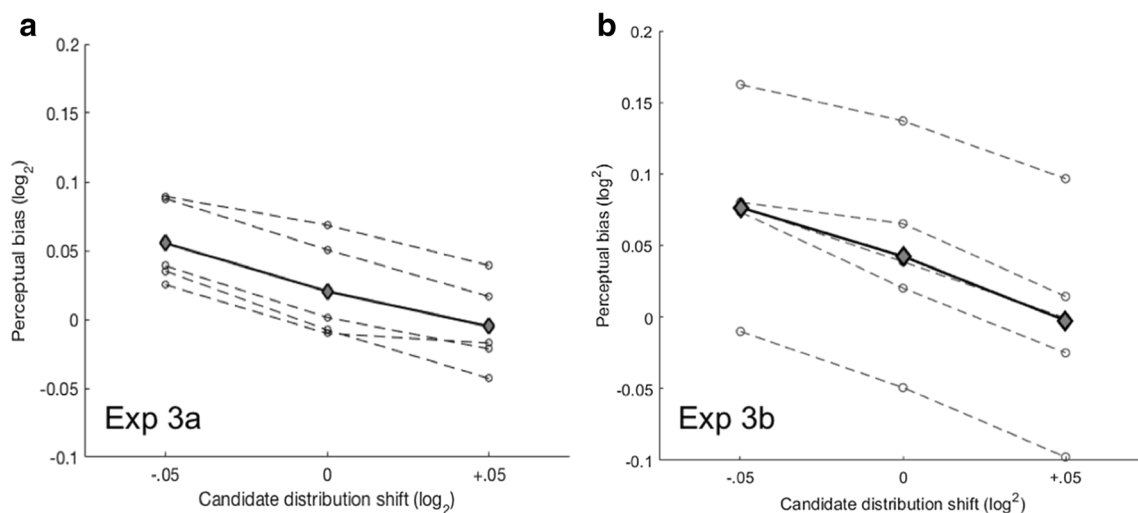

Fig. 5 Results of Experiment 3a (a) and 3b (b). Perceptual biases were plotted against the stimulus condition. Candidate distribution was shifted to be on average 0.05 log units smaller, the same, or larger than the reference. If observers select the candidate that appears most like the constant reference, then their perceptual bias should compensate for the shift of the candidate distribution (see text). Grey filled diamonds and solid black lines denote the mean across observers. Open circles and dashed lines denote individual observers

We found the same results for bias and uncertainty in Experiment 3b: Perceptual biases measured between the three conditions strongly correlated (smaller vs larger candidates: $R = .86$, $p < .001$; smaller candidates vs control: $R = .86$, $p < .001$; larger candidates vs control: $R = .72$, $p < .001$), and uncertainty was less strongly correlated (smaller vs larger candidates: $R = .88$, $p = .052$; smaller candidates vs control: $R = .57$, $p = .319$; larger candidates vs control: $R = .89$, $p = .043$).

Critically, we also tested how well the biases MAPS estimated compensated for the artificial shift of the candidate distribution. Because participants have a natural perceptual bias that makes peripheral targets appear smaller, we first subtracted the bias estimate for zero candidate shift from the estimates for shifts to smaller or larger candidates. Then we conducted a t test to test whether the bias estimates for smaller and larger candidate shifts differed from the actual shift of ±0.05. In Experiment 3a, estimates of perceptual bias were significantly different from the artificial shift both for smaller ($M = .035$), $t(4) = -4.09$, $p = .015$, and for larger candidates ($M = -.026$), $t(4) = 4.7$, $p = .009$. However, in Experiment 3b, where no trial-wise feedback was given, the estimates of perceptual biases were greater than in Experiment 3a, and there were no significant differences between the estimates and the artificial shift for smaller ($M = 0.034$), $t(4)=-2.45$, $p = .071$, or larger ($M = -0.045$), $t(4) = 2.34$, $p = .08$, candidates. It is also worth noting that Experiment 2 directly compared the magnitude of bias estimates with and without feedback, finding a significant difference between the two conditions.

Finally, we also tested whether there was any difference in the goodness of fit of the model (quantified by the accuracy of the winning model in each observer to predict the observer's behavioural response on each trial). The model always predicted trials considerably above the chance level of 25%, ranging on average between prediction accuracies of 51% and 53% in Experiment 3a and 56% and 57% in Experiment 3b. Critically, there was no difference in model accuracy between the different distribution shift conditions either in Experiment 3a, $F(2, 8) = 0.27$, $p = .769$, nor in Experiment 3b, $F(2, 8) = 0.14$, $p = .876$.

Discussion

The results of Experiment 3 suggest that MAPS estimates are largely unaffected by a simple decision bias in which observers select the average candidate (or the middle of the range). When we shifted the distribution of candidate stimuli, observers apparently made their perceptual decision based on which candidate they perceived to be most like the reference stimulus. Importantly, perceptual bias estimates were also highly consistent between the different conditions. This demonstrates that even though the magnitude of bias estimates was modulated by the candidate distribution, MAPS was nonetheless reliable at detecting the underlying pattern of actual perceptual biases in each observer.

Importantly, when trial-wise feedback was given, perceptual bias estimates were smaller and significantly different from the actual introduced shift in the candidate distribution. However, without trial-wise feedback, bias estimates were greater and not significantly different from the actual shifts. Because the feedback given was in relation to the candidate *closest* in size to the reference, and not always an exact match in size, this may have affected participants' responses, such as reducing the ability of participants to use this feedback to correct for their perceptual bias. However, as we still saw a reduction in the overall bias in the feedback versus the no-feedback experiment, this is unlikely to be the case. The difference in estimated perceptual biases between the feedback and no-feedback experiments also mirrored the magnitude of that difference in Experiment 2.

We also carried out two additional experiments (S1 and S2) in which we varied two of the candidate stimuli on some trials (see the Supplementary Information). These corroborated the findings of Experiment 3 showing that MAPS reliably detects the pattern of artificially induced perceptual biases.

General discussion

In this series of experiments, we tested the robustness of perceptual bias and discrimination sensitivity estimates obtained by the MAPS method. Specifically, we tested whether previously observed response biases in MAPS were caused by the response modality, how parameter estimates depend on whether feedback is given on each trial, and how reliable MAPS is at detecting the spatial heterogeneity of perceptual biases.

Response bias is not due to response modality

Our previous work had shown that the four-alternative forced-choice design of this task produces pronounced response biases with observers on average selecting the upper-right candidate most often and the lower-left candidate least often. It is common to see pronounced response biases in designs with more than one alternative to choose from (Jäkel & Wichmann, 2006). In our task, this could have been due to the fact that observers reported their perceptual decision using both hands on the keyboard, and that their handedness could have interacted with response frequency. However, our first experiment demonstrated that when observers made their responses using a computer mouse with their dominant hand, the pattern of response biases remained unchanged. This demonstrates that these response biases are not simply caused by using both hands on the keyboard.

It could nevertheless be the case that hand dominance influences results on this task. Other psychological effects, such as a hemifield attentional bias, may still influence results and

interact with handedness. Previous research has found that handedness does in fact influence attention when perceiving stimuli—observers were most likely to attend to stimuli presented on the side of their dominant hand (Rubichi & Nicoletti, 2006). As only one left-handed participant was present in this study, we cannot draw conclusions as to whether this has occurred here. Further experiments could be conducted to directly compare groups of left-handed and right-handed observers to establish whether behavioural responses in MAPS are influenced by handedness.

Importantly, the magnitude and spatial pattern of perceptual biases were both unaffected by the response modality. This demonstrates MAPS is a reliable method for measuring perceptual biases. Similarly, the (mean group) level of discrimination sensitivity (expressed as the uncertainty modelled by MAPS) was consistent between response modalities. However, the spatial pattern of uncertainties was not consistent. Our previous work found a high test–retest reliability for MAPS estimates of uncertainty even under different environmental conditions (ambient music), although they may be more variable than estimates of perceptual bias (Papageorgiou & Schwarzkopf, 2016). Our results here were mixed, with Experiment 3a showing consistent spatial patterns of uncertainty, unlike when comparing feedback conditions or response modality conditions. It is possible that the differences in uncertainty may vary less than perceptual biases and therefore may need more power to be detected reliably. Critically, both our previous research (Papageorgiou & Schwarzkopf, 2016) and our present experiment show that there is no relationship between response biases and perceptual biases. This means that estimates of perceptual biases obtained with MAPS are not trivially caused by response biases.

Trial-wise feedback reduces perceptual bias estimates

In our second experiment, we compared the effect of providing observers with feedback on their performance on each trial. We found that perceptual bias estimates were considerably greater when they received no feedback, and this pattern of results was replicated between Experiments 3a and 3b. This difference was not simply explained by overall worse performance because their discrimination sensitivity (expressed as the uncertainty parameter fit by the MAPS model) was unaffected by feedback. Importantly, as in our other experiments, the spatial patterns of perceptual biases were very consistent between experimental runs with and without feedback. This suggests that MAPS reliably detects the natural spatial heterogeneity of perceptual biases. Only the magnitude of biases is reduced when feedback is given. This implies that the use of feedback can be justified when the main purpose of the experiment is to establish the spatial pattern of perceptual biases. However, when the exact magnitude of biases is critical (for example, when designing an isomeric stimulus based on the

observer's subjective perceptual experience) feedback is counterproductive. To date, most of our MAPS experiments employed feedback in order to motivate observers to perform the task adequately and to avoid a situation where they simply respond in a haphazard manner. Experiments 2 and 3b here demonstrate that participants are able to respond adequately without feedback, and when the strength of biases is important it would be wise to forgo feedback. However, if feedback is required for some reason, one way to achieve this could be to provide participants with a summary accuracy after each block instead of giving them feedback about each trial.

MAPS reliably detects spatial heterogeneity of perceptual biases

In our third set of experiments, we tested a particular kind of decision bias, in which observers were always prone to selecting candidate stimuli from the middle of the range, might affect the results in MAPS. The results clearly argued against this: When candidates were on average smaller than the reference because the stimulus distribution was shifted, the perceptual bias estimates were larger. Vice versa, when candidates were on average larger, perceptual bias estimates were smaller. Although Experiment 3a (with feedback) appeared to underestimate the candidate distribution shift, Experiment 3b (without feedback) demonstrated that this was likely due to feedback, with a comparable underestimation of perceptual bias as seen in Experiment 2.

Conclusions

In this series of experiments, we demonstrated that MAPS is a robust method for estimating the spatial heterogeneity of perceptual biases. The use of the keyboard with two hands is no worse than the use of the mouse with one hand. Moreover, the spatial pattern of perceptual biases is very consistent irrespective of the exact methods used, suggesting that the variability of actual perceptual biases within observers can be detected reliably. However, the magnitude of perceptual biases is susceptible to experimental choices. Trial-wise feedback can reduce perceptual bias estimates and thus may not always be optimal. Future MAPS experiments could use a random offset on the distribution of candidate stimuli in order to encourage observers to make their decision based on their perceptual experience rather than any decision criterion. Our present findings however suggest that at least for testing the spatial heterogeneity of size perception, such a variable candidate distribution is not necessary. An alternative approach could be to vary the reference stimulus between trials. However, preliminary data on this suggest that observers may not actually use such a roving reference but rely on an implicit representation of the reference they formed in memory. This

problem would probably be exacerbated when testing against an absolute reference like 'vertical orientation' or 'static stimulus'.

We originally designed MAPS to map the spatial heterogeneity of perceptual biases across multiple visual field locations. However, it can also be adapted for measuring perceptual biases, such as perceptual distortions, illusions, or aftereffects in situations when the stimulus location is not of interest. The bias estimates can be averaged across the candidate locations, as was done in Experiment 3 here, or trials could be pooled based on which candidate was chosen. In such a design, it may be advisable to only use two candidates instead of four, as this would boost the efficiency of the experiment. However, decision criteria may be more pronounced and it may therefore also be advisable not to provide any feedback, use no correct target, and include a random offset for the two candidates on each trial.

Data availability Data and stimulus materials are available at: osf.io/gxnvt (Version 3).

References

Afraz, A., Pashkam, M. V., & Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Current Biology: CB*, *20*(23), 2112–2116. doi:https://doi.org/10.1016/j.cub.2010.11.017

Anstis, S. (1998). Picturing peripheral acuity. *Perception*, *27*(7), 817–825.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

de Haas, B., Schwarzkopf, D. S., Alvarez, I., Lawson, R. P., Henriksson, L., Kriegeskorte, N., & Rees, G. (2016). Perception and processing of faces in the human brain is tuned to typical feature locations. *The Journal of Neuroscience*, *36*(36), 9289–9302. doi:https://doi.org/10.1523/JNEUROSCI.4131-14.2016

Finlayson, N. J., Papageorgiou, A., & Schwarzkopf, D. S. (2017). A new method for mapping perceptual biases across visual space. *Journal of Vision*, *17*(9), 5. doi:https://doi.org/10.1167/17.9.5

Genç, E., Bergmann, J., Singer, W., & Kohler, A. (2014). Surface area of early visual cortex predicts individual speed of traveling waves during binocular rivalry. *Cerebral Cortex (New York, N.Y.: 1991)*. doi:10.1093/cercor/bht342

Genç, E., Bergmann, J., Tong, F., Blake, R., Singer, W., & Kohler, A. (2011). Callosal connections of primary visual cortex predict the spatial spreading of binocular rivalry across the visual hemifields. *Frontiers in Human Neuroscience*, *5*, 161. doi:https://doi.org/10.3389/fnhum.2011.00161

Halpern, S. D., Andrews, T. J., & Purves, D. (1999). Interindividual variation in human visual performance. *Journal of Cognitive Neuroscience*, *11*(5), 521–534.

Helmholtz, H. (1867). Handbuch der Physiologischen Optik [Handbook of physiological optics]. Leipzig, Germany: Leopold Voss.

Jäkel, F., & Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers. *Journal of Vision*, *6*(11), 1307–1322. doi:https://doi.org/10.1167/6.11.13

Jogan, M., & Stocker, A. A. (2014). A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *Journal of Vision*, *14*(3), 20. doi:https://doi.org/10.1167/14.3.20

Kosovicheva, A., & Whitney, D. (2017). Stable individual signatures in object localization. *Current Biology*, *27*(14), R700–R701. doi:https://doi.org/10.1016/j.cub.2017.06.001

Morgan, M., Dillenburger, B., Raphael, S., & Solomon, J. A. (2012). Observers can voluntarily shift their psychometric functions without losing sensitivity. *Attention, Perception, & Psychophysics*, *74*(1), 185–193. doi:https://doi.org/10.3758/s13414-011-0222-7

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. *Visual Neuroscience*, *30*(5/6), 197–206. doi:https://doi.org/10.1017/S095252381300045X

Moutsiana, C., de Haas, B., Papageorgiou, A., van Dijk, J. A., Balraj, A., Greenwood, J. A., & Schwarzkopf, D. S. (2016). Cortical idiosyncrasies predict the perception of object size. *Nature Communications*, *7*, 12110. doi:https://doi.org/10.1038/ncomms12110

Newsome, L. R. (1972). Visual angle and apparent size of objects in peripheral vision. *Perception & Psychophysics*, *12*(3), 300–304. doi:https://doi.org/10.3758/BF03207209

Papageorgiou, A., & Schwarzkopf, D. S. (2016). *Data & materials of Papageorgiou & Schwarzkopf* 2016. Retrieved from https://figshare.com/articles/Data_Materials_of_Papageorgiou_Schwarzkopf_2016/4203075

Patten, M. L., & Clifford, C. W. G. (2015). A bias-free measure of the tilt illusion. *Journal of Vision*, *15*(15), 8. doi:https://doi.org/10.1167/15.15.8

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Rubichi, S., & Nicoletti, R. (2006). The Simon effect and handedness: Evidence for a dominant-hand attentional bias in spatial coding. *Perception & Psychophysics*, *68*(7), 1059–1069.

Schwarzkopf, D. S., & Rees, G. (2013). Subjective size perception depends on central visual cortical magnification in human V1. *PLOS ONE*, *8*(3), e60550. doi:https://doi.org/10.1371/journal.pone.0060550

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, *14*(1), 28–30. doi:https://doi.org/10.1038/nn.2706

Verghese, A., Kolbe, S. C., Anderson, A. J., Egan, G. F., & Vidyasagar, T. R. (2014). Functional size of human visual area V1: A neural correlate of top-down attention. *NeuroImage*. doi:https://doi.org/10.1016/j.neuroimage.2014.02.023