The optimal experimental design for Multiple Alternatives Perceptual Search

Nonie J. Finlayson^{1*}, Kelda Manser-Smith¹, Annika Balraj³, Benjamin de Haas^{1,2,4}, & D.

Samuel Schwarzkopf^{1,2,5}

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

²Institute of Cognitive Neuroscience, University College London, UK

³Institute of Biomedical Sciences, The George Washington University, D.C., USA

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

⁵School of Optometry & Vision Science, University of Auckland, NZ

* Corresponding author

Keywords: Perceptual bias, methods, decision bias, response bias, size perception

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 non-perceptual 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 MAPS can detect the spatial heterogeneity of actual perceptual biases, and is not a decisional bias towards choosing the target in the middle of the distribution. 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 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 versus mouse). 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.

Introduction

The perceptual quality of the visual environment is not constant across individuals. Visual abilities and experiences vary considerably between different observers (Halpern et al., 1999) and also covary with inter-individual differences in brain structure and function (Genç et al., 2011, 2011, 2014; Schwarzkopf et al., 2011; Schwarzkopf and Rees, 2013; Verghese et al., 2014). Such variability in perceptual biases has been demonstrated particularly for the strength of visual illusions. Recently, a simple photograph of a dress received viral public attention because it illustrated that different individuals could have very distinct perceptual experiences of the same image (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015; Witzel et al., 2017).

As well as varying between individuals, perceptual biases vary substantially within the visual field for a single person. For example, when presented in peripheral vision, a stimulus appears smaller and distorted compared to when presented in the central visual field (Helmholtz, 1867; Newsome, 1972; Anstis, 1998). Furthermore, there is 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, color (Afraz et al., 2010), position (Kosovicheva and Whitney, 2014), and size (Schwarzkopf and Rees, 2013; Moutsiana et al., 2016). Most of this spatial heterogeneity in perceptual bias could be due to “under-sampling” 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, an

androgynous face image may appear more female if presented to the upper than 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 likely due to idiosyncrasies in the spatial selectivity of neuronal populations (Moutsiana et al., 2016).

Perceptual biases are by their very nature subjective. This complicates any inference that can be drawn from psychophysical experiments about perceptual experience (Morgan et al., 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 non-perceptual 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, 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 (Morgan et al., 2012, 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 favor 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 (Morgan et al., 2013; Jogan and Stocker, 2014; Patten and Clifford, 2015; Moutsiana et al., 2016). 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.

Imagine we want to measure the strength of the Ebbinghaus illusion (Figure 1A). The two light grey discs in this example are identical but the disc surrounded by larger inducers (left) appears smaller than the one surrounded by small inducers (right). In a traditional experiment using the method of constant stimuli, the observer might judge whether the light grey disc on the left or the right is larger. One of the light grey discs (e.g., the one surrounded by small inducer stimuli) would be the *reference* stimulus that is always kept constant. The other light grey disc is the *test* stimulus that is varied randomly or by means of a staircase procedure on

every trial. We can then quantify the perceptual bias produced by the Ebbinghaus illusion by determining the point of subjective equality. Decision factors could affect perceptual decisions in this task. For instance, whenever they are uncertain about the stimuli, observers may choose the stimulus surrounded by small inducers more often – *even though the two light grey discs appear exactly the same* – simply because they know that this is what the experimenter hypothesizes.

In contrast, consider a task where observers must judge which of similar candidate stimuli appears the same (or most similar) to a constant reference (Figure 1B). In this scenario, the observer must choose the one that best matches their perceptual experience because both candidates are affected by the contextual effect. Even if they are exaggerating or underestimating the strength of the effect in isolation because of a decision criterion, the comparison between both candidates must always be colored by their *actual* perceptual bias.

Figure 1. Disentangling decision factors from perceptual bias. **A.** Standard stimulus design for measuring Ebbinghaus illusion strength with the method of constant stimuli. The two targets (light grey discs) are identical but the left target should appear smaller than the right one. **B.** Experimental design seeking to minimize decision-making confounds. The observer's task is to judge which of the two light grey discs surrounded by large inducers (the *candidate* stimuli) appears the most similar to the *reference* stimulus surrounded by small inducers.

We recently developed one such procedure, called Multiple Alternatives Perceptual Search (MAPS), to estimate perceptual biases for the perception of simple stimulus size (Moutsiana et al., 2016; Finlayson et al., 2017). The main purpose of MAPS was to efficiently measure the spatial heterogeneity of perceptual biases across multiple visual field locations. However, the task design also minimizes the influence of decision factors when estimating perceptual biases. In each trial, observers are presented with four candidate stimuli, circles of various

sizes, each in a different visual field quadrant (Figure 2A). 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. We then model the perceptual bias and discrimination ability by fitting a Gaussian tuning curve to the distribution of stimulus values for which the observer chose a particular location to explain their behavioral responses across the whole experiment.

For example, the curve in Figure 2B shows a smoothed histogram of how frequently the observer chose a particular location plotted against the actual stimulus value (here, the size ratio of the candidate stimulus relative to the reference stimulus on a logarithmic scale). The central tendency of this distribution indicates the perceptual bias where the observer perceives stimuli at that location to appear identical to the reference. If this distribution were shifted to the right, the observer would report stimuli that are larger than the reference to appear identical – this means 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 discriminated the stimuli.

Figure 2. A. Typical stimulus in the MAPS design to measure perceptual biases in size judgments across multiple visual field locations (not shown to scale). The central circle is the reference that is typically kept constant in size. Observers are instructed to fixate inside this circle. The four circles in the visual field quadrants are the candidate stimuli. Their sizes are varied although in some versions of the task one candidate is the correct target, that is, its size is identical to the reference (here, the lower left circle). A model is fit to estimate the perceptual bias and uncertainty (a measure of discrimination sensitivity) at each of the four candidate locations. **B.** Smoothed histogram of the distribution of candidate stimulus values in such a MAPS experiment. The x-axis plots the size of the candidates, expressed as the binary logarithm of the candidate diameter divided by the reference diameter. One of the four candidates is the target, which is identical to the reference. This causes the spike in the distribution at 0 log units.

In MAPS experiments, decision factors are relatively unlikely to skew measurements because 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 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 and 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 (Figure 3). These response biases were unrelated to estimates of perceptual biases and their spatial heterogeneity. Nonetheless, it is important to understand if and how the task design affects response biases.

Figure 3. Response bias in a previous MAPS experiment (Papageorgiou and Schwarzkopf, 2016). The proportion of trials observers selected each of the four candidate locations. Grey filled diamonds and solid black line denote the mean across observers. Open circles and dashed lines denote individual observers.

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 decision effects, Experiment 1 tested if observers make their perceptual decision based on a tendency to select the middle (or average) of the range of candidate stimuli. Experiment 1 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 (see below). Experiment 2 explored the effect of trial-wise feedback on the parameter estimates obtained by MAPS.

Experiment 3 tested whether or not the response biases that we previously observed with MAPS were produced by the response modality (using a keyboard or mouse).

Experiment 1

Experiment 1 aimed to test whether the MAPS methodology is influenced by 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 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. 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.

Methods

Participants

Five observers (2 authors; 3 female; ages 21-38 years) took part in this experiment. All observers were healthy and had normal or corrected-to-normal vision. The UCL Research Ethics Committee approved all procedures and observers gave written informed consent prior to participating in this and all following experiments. The relatively small sample size was justified because this was a conceptual (but simpler) replication of the experiments shown in Supplementary Information (Experiments 1a and 1b) 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

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 x 1050 pixels and a refresh rate of 120 Hz. Minimum and maximum luminance values were 0.25 and 230 cd/m². Head position was held at 48cm 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 of light grey (54 cd/m²) circle outlines presented on a black background. Each stimulus array was made up of five circles

(Figure 2A): the reference which was a constant size for all trials (diameter: 0.98°), and was presented in the center 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 centers at a distance of 3.92° from fixation.

There were three stimulus conditions: In a third of trials, 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 centered on 0 (diameter equal to the reference). The standard deviation of the Gaussian distribution was 0.3 log units. One candidate, chosen pseudo-randomly on each trial, was the “correct” target and its diameter was equal to the reference. 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.

Procedure

A blue fixation dot (diameter: 0.2°) was shown in the middle of the screen on a black background for 500ms at the start of each trial. The stimulus array was then presented for 200ms, after which the fixation dot was shown. 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– indicating which appeared most similar in size to the reference. After observers had made their response, there was a ‘ripple’ effect over the target they had

chosen – this was made up of three 50ms frames in which a circle increased in diameter from 0.49° in steps of 0.33° and in luminance. Additional feedback indicated during this 150ms 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 150 blocks of 12 trials each, and the three experimental conditions were randomly interleaved across trials. There was a resting break after the observer completed each block. Between blocks a message on the screen told observers how many blocks they had already completed, and reminded them of the task and responses. Observers could start blocks by pressing any button on the keyboard.

Analysis

Perceptual biases were estimated by fitting a model to predict any given observer's response on each trial as previously described (Moutsiana et al., 2016). In brief, a Gaussian tuning curve was modeled for each candidate stimulus location, which represented the output of a 'detector' in response to the presented stimulus. This detector represents 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 detector with the maximal output signals the behavioral choice. The centroid of this Gaussian tuning curve indicates the perceptual bias where the candidate stimulus appears identical to the reference. 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 behavioral 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. In addition, we quantified the response bias, the proportion of trials that an observer chose a given candidate location irrespective of the actual stimulus.

Individual observer statistics

In addition 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 behavioral responses 1000 times and then fit the MAPS model to these data. For each of these 1000 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 1000

slopes 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, 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 (Figure 4A). This was supported by a significant main effect of the candidate distribution in a one-way repeated-measures analysis of variance ($F(2,8)=66.18$, $p<0.001$). We also conducted an analysis testing the significance of this relationship for each observer using a permutation analysis. 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: $p=0.116$, S2: $p=0.003$, S3: $p=0.007$, S4: $p=0.001$, S5: $p=0.011$).

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). Perceptual biases measured between the three conditions were all very strongly correlated (smaller vs larger candidates: $R=0.81$, $p<0.001$; smaller candidates vs control: $R=0.75$, $p<0.001$; larger candidates vs control: $R=0.88$, $p<0.001$). Uncertainty between the three conditions was also generally correlated, albeit less strongly (smaller vs larger candidates: $R=0.41$, $p=0.069$; smaller candidates vs control: $R=0.54$, $p=0.013$; larger candidates vs control: $R=0.61$, $p=0.004$).

Discussion

The results of Experiment 1 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. We carried out two additional experiments (1a and 1b) in which we varied two of the candidate stimuli on some trials (Supplementary Information). These corroborated the findings of Experiment 1 showing that MAPS reliably detects the pattern of artificially induced perceptual biases.

Figure 4. Results of Experiment 1. Perceptual biases were plotted against the stimulus condition. The 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 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.

Experiment 2

Experiment 1 suggested that observers did not simply always tend to select the middle of the candidate range. However, previous research showed that providing feedback on each trial could also influence an observer's decision criterion (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 the analysis 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.

Methods

Participants

Eight observers (4 authors; 5 female; ages 21-38 years) took part in this experiment. All observers were healthy and had normal or corrected-to-normal vision.

Procedure

The procedure was essentially the same as in Experiment 1. 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. 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.

Stimuli

The stimuli were the same as in Experiment 1, except there was no modulation of the candidate distribution and that in runs without feedback the fixation dot stayed blue when observers selected the correct target.

Results

When observers received no feedback, the magnitude of perceptual bias (Figure 5A) estimates (averaged across the four locations in each observer) was significantly (paired t-test: $t(7)=-2.8$, $p=0.026$) greater ($M=0.06$) than when they received feedback ($M=0.01$, a comparable effect to the equivalent condition in previous experiments, 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 (Figure 5B). However, discrimination sensitivity was at similar levels regardless of whether or not observers received feedback (feedback: $M=0.1$; no feedback: $M=0.1$; $t(7)=0.04$, $p=0.968$). As before, we also tested the similarity of spatial patterns of these parameters between the two conditions. The perceptual biases were highly correlated ($R=0.76$, $p<0.001$). In contrast, the discrimination sensitivities were unrelated in the two conditions ($R=-0.09$, $p=0.624$).

Figure 5. 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.

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 modeled 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 due to the fact that 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.

Experiment 3

Because our previous experiments revealed a consistent response bias, we tested if this could have been caused by the response modality. 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 (4 authors; 6 female; 1 left-handed; ages 20-38 years) took part in this experiment. All observers were healthy and had normal or corrected-to-normal vision.

Procedure & Stimuli

The procedure and stimuli were the same as previous experiments, except as detailed below. 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 centered in the screen when observers could make their response.

The order of the experimental runs using the keyboard or mouse was 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.

Results

Figure 6A shows the response bias measured for each of the four candidate locations. As in our previous experiments (Papageorgiou and 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 (paired t-test: $t(9)=-0.21$, $p=0.838$). The spatial pattern of response biases was also very similar between the two conditions ($R=0.62$, $p<0.001$).

Figure 6. Results of Experiment 3. **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. The 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). The black line indicates the best-fitting linear regression.

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 (Figure 6B) and uncertainties (Figure 6C) 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 ($t(9)=0.85$, $p=0.418$) or uncertainties ($t(9)=0.58$, $p=0.575$).

Next, as in previous experiments, 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 (Figure 6D) was strongly correlated when using the keyboard or the mouse ($R=0.57$, $p<0.001$), indicating that bias was consistent when using both input methods. However, the uncertainties (Figure 6E) were not significantly correlated between response modalities ($R=0.06$, $p=0.716$). 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. As previously observed, these two measures were uncorrelated (keyboard: $R=-0.12$, $p=0.45$; mouse: $R=-0.07$, $p=0.669$).

Discussion

Experiment 3 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. This suggests that our previous findings – 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 and Schwarzkopf, 2016), were right-handed. It could be argued that using both hands to respond using the keyboard biases right-side responses because this 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 and Schwarzkopf, 2016) showing that there was no relationship between the spatial pattern of response biases and perceptual biases (Figure 3). 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.

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 how reliable

MAPS is at detecting the spatial heterogeneity of perceptual biases, how parameter estimates depend on whether feedback is given on each trial, and we probed whether previously observed response biases in MAPS were caused by the response modality.

MAPS reliably detects spatial heterogeneity of perceptual biases

In our first experiment, we tested if observers were always prone to selecting candidate stimuli from the middle of the range. 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. Two additional experiments (Supplementary Information) conceptually replicated this design and corroborated this finding.

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. 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 fully 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 probably 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. This may be particularly important when testing inexperienced observers. Another way to achieve this could be to provide participants with a summary accuracy after each block instead of giving them feedback about each trial.

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 and 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 final 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 and 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- and right-handed observers to establish whether behavioral 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 modeled 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 and Schwarzkopf, 2016). Our results here were mixed, with Experiment 1 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 and 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.

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. This may be because of a stronger bias to select the middle of the range of candidate stimuli. Therefore, 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 most of the experiments 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.

Acknowledgements

This work was supported by a research grant from the European Research Council (310829).

Data availability

Data and stimulus materials are available at: osf.io/gxnvt (Version 2).

References

- Afraz A, Pashkam MV, Cavanagh P (2010) Spatial heterogeneity in the perception of face and form attributes. *Curr Biol CB* 20:2112–2116.
- Anstis S (1998) Picturing peripheral acuity. *Perception* 27:817–825.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- de Haas B, Schwarzkopf DS, Alvarez I, Lawson RP, Henriksson L, Kriegeskorte N, Rees G (2016) Perception and Processing of Faces in the Human Brain Is Tuned to Typical Feature Locations. *J Neurosci Off J Soc Neurosci* 36:9289–9302.

- Finlayson NJ, Papageorgiou A, Schwarzkopf DS (2017) A new method for mapping perceptual biases across visual space. *J Vis* 17:5–5.
- Gegenfurtner KR, Bloj M, Toscani M (2015) The many colours of “the dress.” *Curr Biol CB* 25:R543-544.
- Genç E, Bergmann J, Singer W, Kohler A (2014) Surface Area of Early Visual Cortex Predicts Individual Speed of Traveling Waves During Binocular Rivalry. *Cereb Cortex N Y N* 1991.
- 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. *Front Hum Neurosci* 5:161.
- Halpern SD, Andrews TJ, Purves D (1999) Interindividual variation in human visual performance. *J Cogn Neurosci* 11:521–534.
- Helmholtz H (1867) *Handbuch der Physiologischen Optik*.
- Jäkel F, Wichmann FA (2006) Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers. *J Vis* 6:1307–1322.
- Jogan M, Stocker AA (2014) A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *J Vis* 14:20.
- Kosovicheva A, Whitney D (2014) Stable individual distortions in the perceived locations of static stimuli. *J Vis* 14:1183–1183.
- Lafer-Sousa R, Hermann KL, Conway BR (2015) Striking individual differences in color perception uncovered by “the dress” photograph. *Curr Biol CB* 25:R545-546.
- Morgan M, Dillenburger B, Raphael S, Solomon JA (2012) Observers can voluntarily shift their psychometric functions without losing sensitivity. *Atten Percept Psychophys* 74:185–193.
- Morgan MJ, Melmoth D, Solomon JA (2013) Linking hypotheses underlying Class A and Class B methods. *Vis Neurosci* 30:197–206.
- Moutsiana C, de Haas B, Papageorgiou A, van Dijk JA, Balraj A, Greenwood JA, Schwarzkopf DS (2016) Cortical idiosyncrasies predict the perception of object size. *Nat Commun* 7:12110.
- Newsome LR (1972) Visual angle and apparent size of objects in peripheral vision. *Percept Psychophys* 12:300–304.
- Papageorgiou A, Schwarzkopf DS (2016) Data & Materials of Papageorgiou & Schwarzkopf 2016. Available at: https://figshare.com/articles/Data_Materials_of_Papageorgiou_Schwarzkopf_2016/4203075 [Accessed March 10, 2017].

- Patten ML, Clifford CWG (2015) A bias-free measure of the tilt illusion. *J Vis* 15:8.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* 10:437–442.
- Rubichi S, Nicoletti R (2006) The Simon effect and handedness: evidence for a dominant-hand attentional bias in spatial coding. *Percept Psychophys* 68:1059–1069.
- Schwarzkopf DS, Rees G (2013) Subjective size perception depends on central visual cortical magnification in human v1. *PLoS One* 8:e60550.
- Schwarzkopf DS, Song C, Rees G (2011) The surface area of human V1 predicts the subjective experience of object size. *Nat Neurosci* 14:28–30.
- Verghese A, Kolbe SC, Anderson AJ, Egan GF, Vidyasagar TR (2014) Functional size of human visual area V1: A neural correlate of top-down attention. *NeuroImage* Available at: <http://www.sciencedirect.com/science/article/pii/S1053811914001220> [Accessed March 3, 2014].
- Winkler AD, Spillmann L, Werner JS, Webster MA (2015) Asymmetries in blue-yellow color perception and in the color of “the dress.” *Curr Biol* CB 25:R547-548.
- Witzel C, Racey C, O’Regan JK (2017) The most reasonable explanation of “the dress”: Implicit assumptions about illumination. *J Vis* 17:1.

Supplementary Information

We carried out two additional experiments that were conceptually closely related to Experiment 1 in that they involved changes to the distribution of candidate stimuli. Effectively, we introduced an artificial perceptual bias by changing the values of two of the four candidate stimuli by a fixed amount. This change was not incorporated into the MAPS model. This way we could quantify whether the estimate of perceptual bias for these candidate locations counteracted the artificial biases.

Experiment 1a: Artificial size bias

This was a basic MAPS experiment to measure the perceptual heterogeneity of size judgments. However, in some trials we added a fixed amount to one candidate stimulus and subtracted a fixed amount from another.

Methods

Participants

Fifteen participants (3 authors; 7 female; 1 left-handed; ages 20-37) took part in this experiment.

Procedure

The procedure was identical to the basic MAPS experiment on size perception as described in the main text. Within a single session, each observer completed four experimental runs comprising 200 trials each that were subdivided into 10 blocks of 20 trials. Half of the experimental runs participants performed measured their baseline biases. The other half of the runs contained artificially induced biases: two of the four candidate stimuli were altered subtly, one by adding and one by subtracting a fixed amount from these candidates. Which two targets were altered was counterbalanced across observers, as was the order of experimental runs.

Stimuli & Task

The stimuli were generated in the same way as in the main experiments, and the task was identical. However, in the experimental runs with artificially induced biases, we added 0.1 log units to one candidate stimulus and subtracted 0.1 log units from another.

Results and Discussion

The results suggest that MAPS reliably detected the spatial pattern of the artificially induced biases in this experiment (Figure S1). There was a significant difference in the perceptual bias estimates ($F(3,42)=15.9$, $p<0.001$) depending on the sign and magnitude of the artificially induced biases (for this analysis, we averaged the perceptual bias estimates for the two candidate locations that did not contain an artificial bias: “unbiased”). Importantly, as expected the direction of the perceptual bias estimates went in the opposite direction of the artificial biases. So when candidates were made to look larger than they should, the perceptual bias estimate was reduced and, vice versa, if the artificial bias made the candidate appear smaller, the perceptual bias estimate was shifted to be larger. These results conceptually replicate the findings of Experiment 1 because the estimates of perceptual biases adjust for the artificial shift in the candidate distribution. However, in the present experiment two opposing shifts occurred within the same trial.

Figure S1. Results of Supplementary Experiment 1a. Perceptual bias was plotted for the three manipulations of candidate locations, that is, candidates that were shifted to be 0.1 log units smaller or larger than they should be nominal as well as candidates that were unbiased. Control refers to experimental runs in which all four candidate locations were unbiased. Grey filled diamonds and solid black lines denote the mean across observers. Open circles and dashed lines denote individual observers.

Experiment 1b: Artificial orientation bias

All of our experiments so far had only tested size judgments, so we decided to conduct an experiment to generalize the findings to another visual attribute. Therefore, we conducted a basic MAPS experiment to measure the perceptual heterogeneity of orientation judgments. As in Experiment 1a, in some trials we added a fixed amount to one candidate stimulus and subtracted a fixed amount from another.

Methods

Participants

Ten participants (1 author; 4 female; 2 left-handed; ages 21-32) took part this experiment.

Procedure

Observers were recruited for two sessions on separate days comprising four runs each. Half of the runs contained artificial biases. Each run consisted of 200 trials, subdivided into 10 blocks of 20 trials.

Stimuli

The stimuli contained five small Gabor patches (sinusoidal grating with wavelength: 0.62° ; contrast: 50% convolved with a Gaussian envelope with standard deviation: 0.49°) presented on a grey background (54cd/m^2). The central patch was the reference and was oriented either at 30° or 120° counterclockwise from horizontal. From trial to trial the target orientations varied relative to the reference by drawing orientation offsets from the reference from a Gaussian distribution with standard deviation of 15° . The artificial bias added or subtracted from two candidate stimuli in the artificial bias runs was 5° . The 'ripple' effect indicating the observer's behavioral response was a non-Cartesian grating, a concentric sinusoidal grating (wavelength: 2.48°) convolved with a Gaussian envelope increasing from 0.49° in steps of 0.33° and decreasing in contrast from 50% in steps of 10%.

Task

The task in this experiment was conceptually the same as for the size perception experiments. Observers were presented with a stimulus array comprising five Gabor patches. The central patch indicated the reference orientation and was constant throughout the experiment but it was counterbalanced across observers to be either 30° or 120° counterclockwise from horizontal. Observers were instructed to indicate which of the candidates appeared to be most parallel to the reference patch.

Results and Discussion

The results suggest that MAPS reliably detected the spatial pattern of the artificially induced biases also for orientation judgments (Figure S2). There was a significant difference in the perceptual bias estimates ($F(3,27)=39.4$, $p<0.001$) depending on the sign and magnitude of the artificially induced biases (for this analysis, we again averaged the perceptual bias estimates for the two candidate locations that did not

contain an artificial bias). Importantly, as expected the direction of the perceptual bias estimates went in the opposite direction of the artificial biases. So when candidates were made to look larger than they should, the perceptual bias estimate was reduced and, vice versa, if the artificial bias made the candidate appear smaller, the perceptual bias estimate was shifted to be larger. These results also conceptually replicate the findings of Experiment 1 and Experiment 1a and generalize these findings to a different visual attribute.

Figure S2. Results of Supplementary Experiment 1b. Perceptual bias was plotted for the three manipulations of candidate locations, that is, candidates that were shifted to be 5° more clockwise or counter-clockwise than they should be nominal as well as candidates that were unbiased. Control refers to experimental runs in which all four candidate locations were unbiased. Grey filled diamonds and solid black lines denote the mean across observers. Open circles and dashed lines denote individual observers.