A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability

Matjaž Jogan
Department of Psychology and Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA

Alan A. Stocker
Department of Psychology and Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA

Perception is often biased by secondary stimulus attributes (e.g., stimulus noise, attention, or spatial context). A correct quantitative characterization of perceptual bias is essential for testing hypotheses about the underlying perceptual mechanisms and computations. We demonstrate that the standard two-alternative forced choice (2AFC) method can lead to incorrect estimates of perceptual bias. We present a new 2AFC method that solves this problem by asking subjects to judge the relative perceptual distances between the test and each of two reference stimuli. Naïve subjects can easily perform this task. We successfully validated the new method with a visual motion-discrimination experiment. We demonstrate that the method permits an efficient and accurate characterization of perceptual bias and simultaneously provides measures of discriminability for both the reference and test stimulus, all from a single stimulus condition. This makes it an attractive choice for the characterization of perceptual bias and discriminability in a wide variety of psychophysical experiments.

Introduction

The standard two-alternative forced choice (2AFC) method is a ubiquitous choice for measuring detection or discrimination thresholds (Fechner, 1860/1966; Green & Sweets, 1966). During one trial of a typical 2AFC experiment, a subject is asked to make a decision about the perceived difference between two stimuli with regard to a particular stimulus parameter of interest (we will refer to this as the primary parameter). For example, a subject is instructed to indicate which one of two random dot motion stimuli shows an overall motion direction "… that is more counterclockwise than the other" (see Figure 1a).

Over repeated trials with stimulus pairs of varying difference in motion direction, a psychometric function can be calculated that reflects the empirical probability of the subject's choice as a function of stimulus difference. The slope of this sigmoidal function is a direct measure of the subject's discrimination threshold (see Figure 1b, bold gray line). There are many benefits in using the standard 2AFC method. It requires subjects to perform a simple decision task. It provides a threshold measure in physical units unlike scaling methods (Stevens, 1946). Also, by restricting a subject's response to a binary decision, it avoids any contamination of the measured perceptual thresholds with motor noise unlike methods of adjustment. Furthermore, it generally provides a large number of data points, thus allowing a statistically sound analysis and robust fits of the data. Finally, with signal detection theory (SDT) (Green & Sweets, 1966; Macmillan & Creelman, 2005), there exists a well-accepted and simple observer model framework that links the psychometric functions to an internal sensory representation of the stimulus parameter of interest. For all these reasons, the standard 2AFC method has been popular in the perceptual and cognitive neurosciences.

It is often of interest to characterize how secondary stimulus attributes bias the percept of a primary stimulus parameter. This includes contextual effects induced, for example, by the spatial surround (e.g., Tadin, Lappin, Gilroy, & Blake, 2003) or the impact of the illumination spectrum on color perception (e.g., Brainard & Freeman, 1997) but also effects related to the attentional (e.g., Carrasco, Ling, & Read, 2004) or the adaptation state (e.g., Schwartz, Hsu, & Dayan,

Citation: 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):22, 1–18, http://www.journalofvision.org/content/14/3/20, doi:10.1167/14.3.20.

doi: 10.1167/14.3.20 Received May 17, 2013; published March 13, 2014 ISSN 1534-7362 © 2014 ARVO

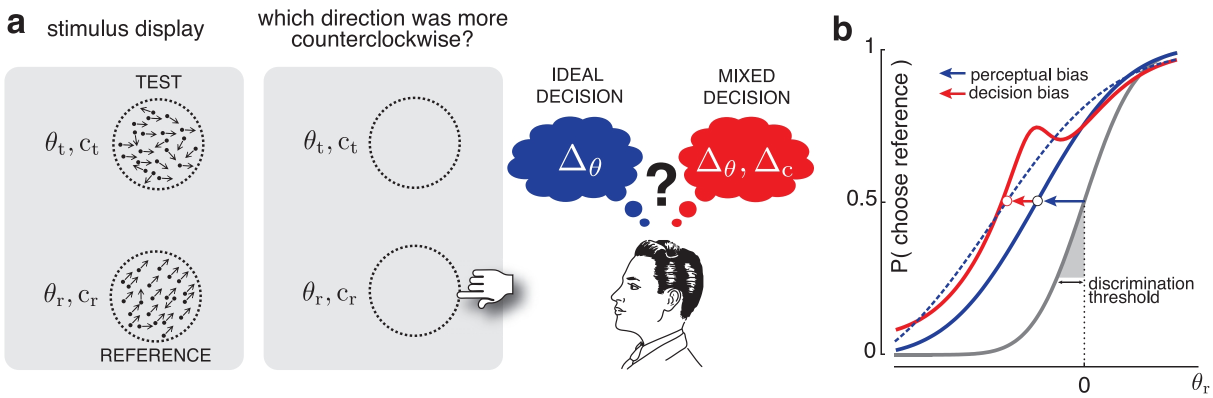

Figure 1. Measuring perceptual bias with the standard 2AFC method. (a) Example experiment: In order to measure how dot coherence *c* may bias the perceived direction of motion θ of a random-dot stimulus, a subject is presented with a pair of stimuli (test/reference) that differ in both stimulus parameters, i.e., direction θ and dot coherence *c*. The subject's task is to judge the relative difference in motion direction θ (primary parameter) between the two stimuli by indicating which stimulus' motion direction is e.g., "… *more counterclockwise than the other.*" Ideally, the subject's choice is exclusively based on the perceived difference in motion directions Δ_θ between the stimuli. However, it cannot be ruled out that the subject may employ an alternative, mixed decision strategy that also takes into account the difference in motion coherence Δ_c. (b) Ideally, perceptual bias is quantified as the shift of the psychometric function due to the difference in dot coherence between test and reference (bold blue line) relative to the psychometric function for identical dot coherences (bold gray line). A mixed decision strategy, however, can result in a distorted psychometric function (bold red line, prediction from a mixed strategy decision model; see Methods). A naïve fit of the distorted psychometric function (dashed blue line) then leads to an error in measured perceptual bias. We refer to this error as decision bias.

2007) of the observer. Furthermore, Bayesian modeling approaches to perception (e.g., Knill & Richards, 1996) have an interest in a precise quantification of perceptual biases as a function of stimulus uncertainty, when uncertainty is either implicitly modulated by stimulus contrast (e.g., Stocker & Simoncelli, 2006) or explicitly by adding stimulus noise (e.g., Girshick, Landy, & Simoncelli, 2011; Putzeys, Bethge, Wichmann, Wagemans, & Goris, 2012; Webb, Ledgeway, & McGraw, 2010). Many studies that have investigated the effect of these secondary stimulus attributes have relied on the standard 2AFC method for measuring perceptual biases. However, we argue that the standard 2AFC method is not well constrained for this purpose and can lead to incorrect estimates of perceptual biases. We illustrate the problem by returning to our example experiment depicted in Figure 1. Let us assume that the experimenter wants to measure how a change in dot coherence biases a subject's perceived motion direction of a random dot motion stimulus. Using the standard 2AFC method, the subject is presented with two stimulus alternatives that *have to differ* in their coherence levels. Therefore, when performing the task, the subject could (unintentionally or unconsciously) base his or her decision not only on the perceived difference in motion direction, but also in dot coherence. Intuitively, such a mixed decision strategy seems particularly likely in conditions in which the perceived difference in motion direction is small, and thus the task is difficult (Morgan, Dillenburger, Raphael, & Solomon, 2012). In general, it is not feasible for the experimenter to infer the exact decision strategy of a subject in every trial, making a correct interpretation of the data very difficult if not impossible. Ignoring the ambiguity, however, can lead to substantial misinterpretations of the measured psychometric functions and thus to errors in estimating perceptual biases.

In order to get a qualitative sense of these errors, we considered a simple model for a mixed decision strategy (see Methods for details). We assumed that the frequency with which a subject makes a decision based on the difference in dot coherence rather than motion direction is proportional to the relative uncertainty in each perceived difference. As shown in Figure 1b, such a mixed decision strategy leads to distortions of the psychometric function predominantly in the range in which the perceived difference in motion direction is small. Fitting the distorted psychometric function with a sigmoidal function (e.g., a cumulative Gaussian) can lead to significant errors in estimating perceptual bias as well as discrimination threshold; in our example, the experimenter would overestimate both. The sign and the magnitude of the errors depend on the instruction (*i.e.*, "… *more/less counterclockwise* …") and the specifics of the observer's mixed strategy. In the following, we will refer to the estimation error as *decision bias*. Because decision biases most likely occur in trials that are difficult for the observer, i.e., around perceptual equality, it has been suggested early on to discount the reliability of difficult trials (Fechner, 1860/1966). This would require providing subjects with a third choice (*e.g.*, "*I don't know*"), and several methods have been proposed to explicitly model such choices of indecision (García-Pérez & Alcala-Quintana, 2011; Kaernbach, 2001; Watson, Kellogg, Kawanishi, & Lucas, 1973). While discounting those trials that are most likely corrupted might help to reduce decision bias, it will not prevent it. Furthermore, this approach discounts those trials that are particularly informative in constraining the psychometric function, which is not very efficient.

We considered it more promising to try to prevent decision bias in the first place rather than to work around it. In this paper, we present a new 2AFC method that is designed to prevent decision biases by making sure that a subject's choice is constrained to stimuli that are identical in all but the primary stimulus parameter. By adding a second reference stimulus, the method resolves the ambiguity in the decision process while still retaining all the advantages of a forced choice procedure. We experimentally validated our new method with a visual motion-discrimination task and demonstrate that it not only successfully avoids decision biases, but is also efficient in characterizing discrimination thresholds compared to the standard 2AFC method.

Results

The setup of the new method is as follows: Instead of being presented with one reference stimulus (and the test) as in the standard 2AFC method, a subject is now presented with two reference stimuli. The subject's task is to decide which one of the two references is "*perceptually closer*" to the test with regard to the primary stimulus parameter. By constraining the two references to differ only in the primary parameter, we prevent the subject from applying a mixed strategy in making that decision. Yet we still enforce a comparison to a test stimulus that can be different in terms of the secondary stimulus parameter. Figure 2a illustrates the new method for the random dot motion stimulus example introduced earlier: A subject is presented with a display that contains two reference stimuli with motion directions θ_{r_1} and θ_{r_2}, respectively, and identical dot coherence c_r, and a test stimulus with motion direction θ_t and dot coherence c_t. Because the subject is forced to make a decision between the two reference stimuli that have identical coherence levels, the decision is limited to a comparison of perceived motion directions and thus is free of decision bias.

Analogous to the standard 2AFC method, an empirical choice probability can be computed over repeated presentations of a particular stimulus triple (the two references and the test). However, by varying both references in motion direction, we no longer obtain a one-dimensional psychometric function but rather probability values at discrete locations in the two-dimensional parameter space defined by the two reference motion directions (see Figure 2b). We refer to this table as the *psychomatrix*. Efficient sampling of this two-dimensional space is important, yet commonly used adaptive staircase procedures are not easily extendable to two dimensions. We thus resorted to an adaptive Bayesian estimation technique that optimally selects the reference values θ_{r_1} and θ_{r_2} for the current trial based on the outcomes of previous trials by maximizing the expected information gain (Kontsevich & Tyler, 1999; Kujala & Lukka, 2006; Lesmes, Jeon, Lu, & Dosher, 2006; Sims & Pelli, 1987). This technique and intrinsic symmetries in the psychomatrix allow a sparse and very efficient sampling of the stimulus space (orange squares in Figure 2b; see Methods for details).

From the recorded psychomatrix, a subject's perceptual bias and discriminability can be extracted by fitting the empirical probability values with a two-dimensional probability surface. We extended standard SDT to obtain a functional description of this probability surface (Figure 2c). Following the theory, we treat the perceived motion directions of the two reference and the test stimuli $(\hat{\theta}_{r_1}, \hat{\theta}_{r_2}, \hat{\theta}_t)$[1] over repeated trials as noisy samples of the conditional probability distributions $p(\hat{\theta}_{r_1}|\theta_{r_1}), p(\hat{\theta}_{r_2}|\theta_{r_2})$, and $p(\hat{\theta}_t|\theta_t)$. These perceptual distributions are defined by the particulars of the assumed underlying perceptual model. For example, they can represent the perceptual characteristics of an ideal Bayesian observer (Stocker & Simoncelli, 2006). The new method is not limited to any specific perceptual model. Thus, for the purpose of this paper and for reasons of simplicity, we assumed that these distributions are Gaussians and, in the case of the two reference stimuli, have identical and fixed widths and are centered on the true stimulus values θ_{r_1} and θ_{r_2}, respectively. This left us with three free model parameters: Standard deviations σ_r and σ_t, which represent the perceptual noise levels of the reference and the test stimuli, respectively, and $\langle\hat{\theta}_t\rangle$, which represents the mean perceived motion direction of the test stimulus, i.e., the point of subjective equality (PSE).

With these assumptions, we can compute the probability that the second reference is perceptually closer to the test than the first reference, i.e., $P(|\Delta(\hat{\theta}_{r_1}, \hat{\theta}_t)| > |\Delta(\hat{\theta}_{r_2}, \hat{\theta}_t)|)$ by considering all possible relative orders of the three stimuli (see Methods for details). Figure 2d shows the predicted decision probability surface. It forms two edges along each diagonal. The surface is point symmetric around the intersection of the edges that represents $\langle\hat{\theta}_t\rangle$. Symmetry originates from the fact that for each pair of reference stimuli there is a point symmetric pair that has the exact same perceptual distances to the test and thus leads to the

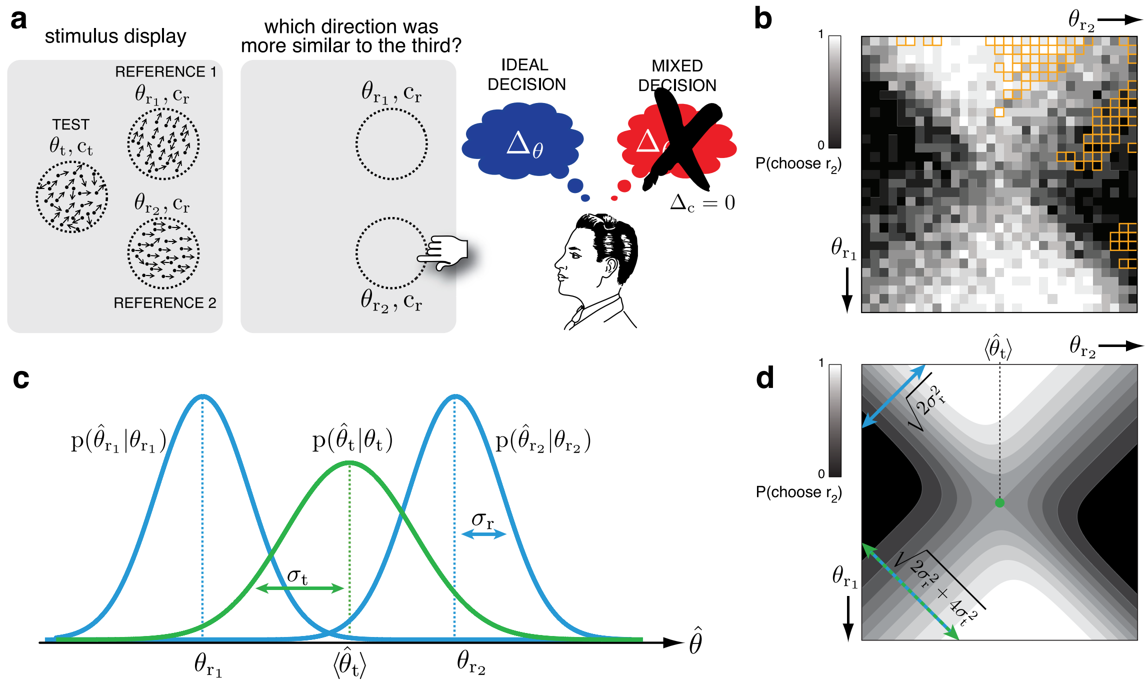

Figure 2. New 2AFC method. (a) In each trial, a subject is presented with two reference stimuli and a test. The two reference stimuli are constrained to differ only in the primary parameter (here: motion direction θ). The test stimulus, however, can differ from the references not only in the primary but also the secondary stimulus parameter (here: dot coherence c). After the simultaneous presentation of all three stimuli, the subject is informed which two were the references. The subject's task is to make a decision based on the relative difference in the primary stimulus parameter between the test and each of the two references, for example, to choose the reference with the smallest perceptual distance to the test (i.e., minimal $|\Delta(\theta_r, \theta_t)|$). We argue that this choice is free of decision bias because both references are identical in all stimulus parameters other than their motion direction. (b) The empirical choice probabilities over repeated trials at discrete value pairs $(\theta_{r_1}, \theta_{r_2})$ of the reference stimuli form a *psychomatrix*. Probability values are rendered in gray scale. Intrinsic symmetries and the application of an adaptive method (see Methods) allow an efficient sampling of this parameter space (orange squares). (c) An observer model based on SDT links the empirical probabilities of the psychomatrix to the perceptual distributions. Each curve represents the distribution of the percept of the test and the two reference stimuli as a function of the primary stimulus parameter. Under the assumption of Gaussian sensory noise (assumed to be identical and constant for both references) the observer model is fully determined by the standard deviations σ_r, σ_t and the expected percept of the test $\langle\hat{\theta}_t\rangle$ (d) The two-dimensional probability surface of the psychomatrix $P(|\Delta(\hat{\theta}_{r_1}, \hat{\theta}_t)| > |\Delta(\hat{\theta}_{r_2}, \hat{\theta}_t)|)$ as predicted by the observer model (see Methods for details).

same decision probability. Each line on the surface represents reference pairs that lead to equal decision probabilities (iso-probability lines). The observer is at chance ($p = 0.5$) along the diagonals where the references are either identical (positive diagonal) or are equally distant to the perceived test but on opposite sides (negative diagonal). Normal to each diagonal, the probability values transition from high to low and vice versa. The slopes of these transitions directly reflect the noise levels of the test and reference stimuli. Moving perpendicular to the positive diagonal (blue arrow) increases the distance between the two references without changing their mean, and thus the slope in this direction reflects how well the subject can discriminate the two references as governed by their noise parameter σ_r. Moving perpendicular from a point on the negative diagonal changes the mean of the two references without changing their relative distances to each other. Changes in probability in this direction reflect how well the subject can discriminate the position of the test relative to the mean of the references, which is determined by both noise parameters σ_r and σ_t. Note that the psychomatrix fully constrains all three free model parameters (the noise parameters of the reference and the test stimuli, σ_r and σ_t, as well as the PSE of the test stimulus, $\langle\hat{\theta}_t\rangle$). The same is not true for the standard 2AFC method with which the noise parameters of the test and the reference are jointly reflected in the slope of the psychometric function and thus cannot be disambiguated without additional measurements.

Some assumptions of the above observer model may not generally hold, in particular, the assumption that the noise levels are independent of the stimulus value. If the noise characteristics are known, one can always find an appropriate perceptual space such that the noise level is constant in that space. Even if not, the model still provides a good generic description of the decision probabilities of the psychomatrix very much like a cumulative Gaussian often provides a simple yet good parametric description of a typical psychometric function. We verified this with a set of numerical simulations in which we simulated an observer's behavior, assuming stimulus-dependent noise, but then fitted the resulting psychomatrix with a constant noise model. Although we found small estimation errors for perceptual bias under these conditions, they were significantly smaller than the errors obtained with the standard 2AFC method under the same conditions (see Figure S2, *Supplemental Data*). Of course, observer models with more sophisticated noise characteristics can always be implemented although they might require a joint fit across different stimulus combinations.

Experimental validation and comparison

We experimentally validated our new method with the visual motion discrimination experiment shown in Figure 2. The goal was twofold: First, we wanted to demonstrate that our method is practical and allows us to correctly measure perceptual bias and discriminability. Second, we also wanted to provide a quantitative comparison with the standard 2AFC method. In particular, we wanted to test whether the standard 2AFC method indeed can lead to decision biases and show that those biases do not occur with our new method.

Thirteen human subjects participated in the experiment. All subjects first performed the experiment using the standard 2AFC method and then again using our new method. The task was to make a forced choice based on the relative perceived directions of motion between a test and reference stimuli (see Figures 1a and 2a). The experiment contained two sets of stimulus conditions: *balanced conditions* in which dot coherences of test and reference stimuli were identical and *unbalanced conditions* in which they were different. Trials of balanced and unbalanced conditions were randomly interleaved. Stimulus presentation time, spatial eccentricity, and the number of trials per condition were identical for both methods. We expected subjects to be prone to decision bias in the unbalanced stimulus conditions when performing the standard 2AFC task but that such bias would not occur with our new method. Because it is difficult to distinguish whether a shift of the psychometric function is due to decision or perceptual bias (see Figure 1b), we designed the experiment such that the *expected perceptual bias had to be zero*. Specifically, although test directions were uniformly sampled from all directions, we combined all the data, thus averaging out any potential perceptual bias due to symmetry across the cardinal directions (Dakin & Alais, 2010). The rationale was that any remaining and significant bias therefore would represent decision bias. In order to control for the direction of a potential decision bias, we divided the subjects into two groups, each obtaining a different task instruction for the standard 2AFC task: Subjects of the first group (*CW*) were instructed to choose the stimulus whose direction of motion was more clockwise; subjects of the second group (*CCW*) were asked to pick the one whose direction was more counterclockwise. Subjects were not aware of the separation of the subject pool into two groups or of the fact that there were two different instructions.

We found that subjects indeed showed significant decision biases in the unbalanced stimulus conditions when performing the standard 2AFC task. Figure 3a shows the measured bias as a function of test coherence c_t for the average subject of each of the two groups. As expected, the decision bias of each group is in opposite directions and grows with increasing difference in dot coherence between test and reference stimulus. This behavior is highly significant and matches the prediction of a mixed decision strategy model. It suggests that subjects, when in doubt, choose the more reliable stimulus (i.e., the one with a higher coherence level). While small in absolute size, these decision biases are of the same order of magnitude as the average expected perceptual biases in perceived motion direction (Dakin & Alais, 2010) and thus would significantly affect such measurements. We found substantial differences across individual subjects: Some showed strong biases, and others did not. Figure 3c illustrates this by plotting each subject's total number of errors according to whether the errors consisted of picking the reference or the test stimulus. A mixed decision strategy predicts that the errors would be more frequent in picking the stimulus with a higher coherence level. Although overall subjects clearly exhibited such behavior, some subjects did not. We think this is further evidence that the standard 2AFC method is not well constrained in unbalanced stimulus conditions and permits the subjects to employ rather arbitrary decision strategies that cannot be fully controlled. We did not, however, find any subject that applied the opposite strategy, i.e., preferring the stimulus with lower coherence level. In comparison, data collected with our new 2AFC method do not exhibit any significant decision bias (Figure 3b).

Data from the balanced conditions allowed us to quantitatively compare the discrimination thresholds measured with the standard and our new 2AFC method. We expected that with all other stimulus parameters equal (i.e., presentation duration, eccentricity, size, etc.), adding a second reference would require the visual system to distribute its attentional resources across more stimuli and thus would decrease

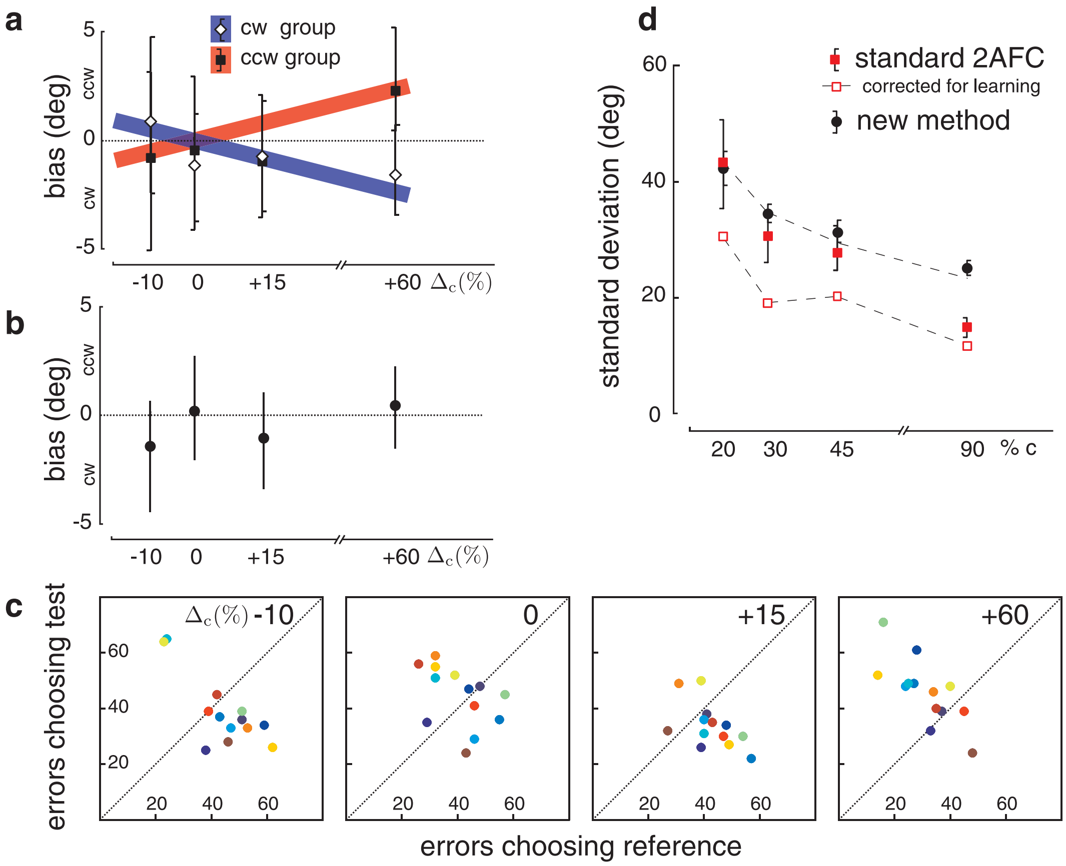

Figure 3. Experimental validation and comparison to the standard 2AFC method. (a) Decision bias of the standard 2AFC method as a function of dot coherence difference Δ_c between test and reference stimuli. Shown is the bias for the average subject of each group. As predicted, the decision bias depends on the task instruction and is larger the larger the coherence difference. The bias pattern is highly significant ($p < 10^{-3}$). (b) The new method correctly estimates perceptual bias to be zero and shows no significant decision bias (average subject). (c) Error trials for individual subjects. Each panel shows the total number of errors in choosing the reference over the test and vice versa for each subject (individual dot color) and for each unbalanced condition (Δ_c). The overall shift in relative error frequency with dot coherence difference reflects the decision bias shown in (a). However, there is substantial variation across subjects, indicating that individual subjects may have applied different decision strategies. (d) Discrimination thresholds of the average subject obtained with the standard and the new 2AFC method (both indicated as standard deviations σ of assumed underlying Gaussian noise distributions). Since subjects first performed the standard 2AFC experiment, perceptual learning effects are likely to have led to smaller estimates for the new method. Indeed, noise levels were significantly lower for the standard 2AFC method when computed only over the second half of the trials (dashed lines; see also Supplemental Data). All error bars indicate 95% bootstrap intervals.

the accuracy with which each stimulus was represented and memorized (Pelli, 1985). As a consequence, we expected discrimination thresholds to be higher for the new method. We found that thresholds were comparable when computed over all the data except at the lowest coherence level. However, because subjects performed the standard 2AFC experiments first, it was likely that perceptual learning led to the relatively small thresholds for the new method (see Figure S1, *Supplemental Data*). Limiting the analysis to data from the second half of the trials resulted in slightly but significantly smaller thresholds for the standard 2AFC method, whereas the values for the new method did not significantly change (Figure 3d, dashed lines).

In order to demonstrate the new method's ability to measure nonzero perceptual biases, we fit the data for individual motion directions (i.e., we fit a full psychomatrix and extracted the three parameters σ_r, σ_t, and $\langle\hat{\theta}_t\rangle$ for each motion direction independently). Because the experiment was originally not designed for this local analysis of direction bias, we had to take measures to increase the number and quality of trials per condition. We thus limited the analysis to the combined data of the three most reliable and least noisy subjects and for the least noisy stimulus condition only. In addition, we

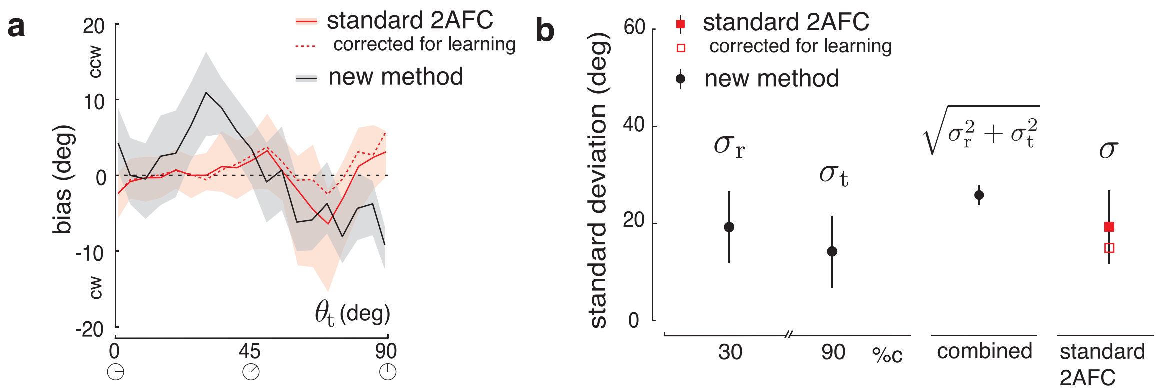

Figure 4. Biases in perceived motion direction measured with the new and the standard 2AFC method. (a) Black lines show the measured perceptual bias with the new 2AFC method for a test ($c_t = 90\%$) relative to a reference stimulus ($c_r = 30\%$) as a function of test motion direction θ_t. Data were collapsed assuming perceptual biases to be symmetric along the cardinal directions. Gray shaded areas represent the 95% interval over 1,000 bootstrap samples. Red lines show the bias measured with the standard 2AFC method. The measurement is likely be confounded with decision bias. (b) Simultaneously estimated noise parameters σ_r and σ_t with the new method averaged over all directions θ_t. The comparison with the overall noise level measured with the standard 2AFC method (σ, standard deviation of cumulative Gaussian) shows that noise estimates with the new method are higher.

increased the number of trials for each direction by assuming that perceptual biases are symmetric across both cardinal axes, thus collapsing trials from all four quadrants of orientation. Furthermore, we combined trials over a sliding window of 15°. Figure 4a shows the extracted perceptual direction bias of a test stimulus with $c_t = 90\%$ coherence relative to reference stimuli with $c_r = 30\%$ as a function of test direction θ_t. It indicates that the direction of the reference stimulus had to be more counterclockwise to perceptually match the direction of the test on the interval (0°, 45°) (positive relative bias) and more clockwise on the interval (45°, 90°) (negative relative bias). This pattern matches some previous reports (Dakin & Alais, 2010; Loffler & Orbach, 2001; Rauber & Treue, 1998) and is similar to what has been reported for the relative perceptual bias in perceived orientation (Girshick et al., 2011; Tomassini, Morgan, & Solomon, 2010). We forgo here a more detailed discussion of perceptual bias in motion direction as it would go beyond the scope of this paper. Figure 4b shows the values of the jointly estimated noise parameters σ_r and σ_t averaged over all directions.

For comparison, we performed the same local bias analysis of the data from the standard 2AFC experiment (same subset of subjects). The extracted pattern does not show the cardinal bias effect with the same clarity and symmetry as the pattern characterized with the new method, indicating that the estimated values are likely to be affected by decision bias (Figure 4a). In contrast to our new method, the standard 2AFC method does not allow a simultaneous fit of both the test and the reference noise parameters in unbalanced stimulus conditions. Rather, the slope of a psychometric function in these conditions represents the combined noise level. In Figure 4b, we compare the average standard deviation of the cumulative Gaussian fits for individual test motion directions with the combined values from the new method. As with the balanced stimulus conditions, overall noise levels for the standard 2AFC method are slightly smaller than those measured with the new method.

Maximal reference range of the psychomatrix

The adaptive allocation algorithm places the references at values that are most informative with regard to the parameters of the underlying observer model. These are typically values along the upper and lower end of the two edges of the probability surface along the diagonals of the psychomatrix (see Figure 2b). As long as the range of the psychomatrix (i.e., the extrema of the reference values) is large enough relative to the slope of these edges, the fit of the psychomatrix is well behaved. However, if the range is too small, then the most informative values lie outside the range of the psychomatrix, which can lead to fits of the model parameters that are not well constrained and are unstable. We systematically analyzed this problem and simulated the motion discrimination experiment for different reference ranges using a model observer with fixed and known noise levels σ_r and σ_t. The resulting placement distributions of the reference stimuli differ for different reference ranges. As the range is successively reduced, more and more trials are placed at the boundaries of the psychomatrix. In the extreme case (±25°) all trials are placed at a single location (Figure 5a). We can define a boundary index (BI) that indicates the total fraction of trials placed at the boundary values of the psychomatrix. Figure 5b, c shows the fit quality

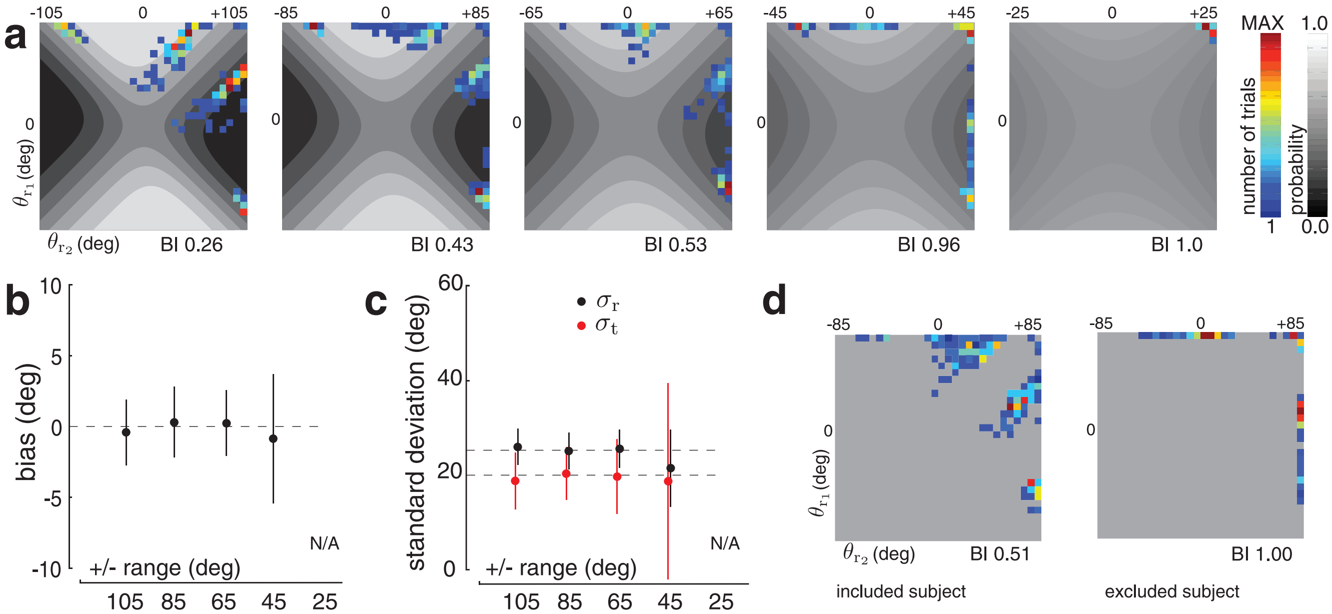

Figure 5. Reference range of the psychomatrix. (a) The placement of the reference values depends on the chosen range of the psychomatrix. Shown are the probability surface and the distribution of the reference values for a model observer in a simulated run of our experiment. The observer's noise levels were set to $\sigma_r = 25°$, $\sigma_t = 20°$, and the average perceived test value $\langle\hat{\theta}_t\rangle$ was assumed to be zero. Each panel shows the probability surface and the reference placements for a different reference range of the psychomatrix with decreasing ranges from left to right. As the range decreases, the adaptive method allocates more trials on the borders of the psychomatrix as indicated by the BI, the total fraction of trials with reference pairs placed at the boundary of the psychomatrix. The fit perceptual biases (b) and noise parameters (c) for each of the simulated psychomatrices (mean and standard deviation over repetitions of the simulated experiment). The fits become less reliable if the range is too small, and the trials are pushed toward the border values of the psychomatrix. (d) The actual trial placement for two subjects (unbalanced condition, $\Delta_c = 60\%$). The perceptual bias over motion direction (Figure 4) was measured by combining the data from subjects with a BI $<$ 0.9.

for perceptual bias and noise parameters, respectively, for each of the range conditions. As the range is large enough and the BI is small, the fit values well reflect the model values. However, as the chosen range becomes too small and the BI approaches one, the fits are no longer well constrained, leading to increasingly biased and noisy estimates of the noise parameter. The estimates for perceptual bias are slightly more robust. Computing the BI for every subject in our pool revealed that, for most subjects, the average value was close to one. This suggests that the chosen reference range in our experiment was too small relative to the perceptual noise levels for most subjects. As a result, we only included subjects that had a BI $<$ 0.9 in our above analysis of perceptual bias over individual motion directions (see Figure 4a). This explains the relatively small number of subjects in that analysis. Figure 5c shows the trial histogram and the BI for a typical subject who was included and one who was not.

Selecting a finite range is necessary because the adaptive method needs to compute the expected entropy decrease online in between trials over a finite set of possible reference pairs $(\theta_{r_1}, \theta_{r_2})$. The maximal size of the psychomatrix is thus constrained by the total available computational resources of the experimental setup. Range and resolution should therefore be carefully traded off against each other in order to minimize the above discussed boundary problem. Note that circular perceptual parameters have a natural maximal limit for the range. In the case of (positive) infinite spaces (e.g., visual speed) compressive mappings, such as mapping to a logarithmic space, can help to reduce the range problem.

Convergence properties

A closer examination of the confidence intervals of the estimated noise parameters (i.e., discrimination thresholds) revealed that they are consistently smaller for the new compared to the standard 2AFC method (Figure 3d). Given that we used an equal number of trials for both methods, this suggests that the new method is more efficient than the standard 2AFC procedure in estimating the noise parameters. This is not unexpected because adding a second reference stimulus will, theoretically at least, make a subject's binary decision in each trial more informative about the underlying free perceptual parameter (assuming that both reference stimuli share the same noise parameter).

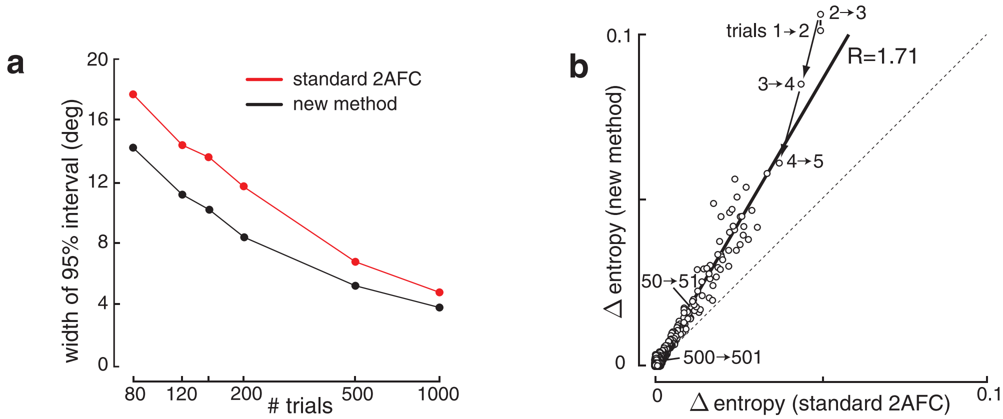

Figure 6. Efficiency and convergence speed. We simulated a discrimination experiment using an optimal allocation of the reference stimuli for both methods. (a) Estimated confidence intervals for the noise parameter σ as a function of trial number for both methods. (b) Expected decrease in estimation uncertainty (entropy) between two consecutive trials. Each dot represents the average entropy decrease between consecutive trial pairs (n, $n+1$) for the standard 2AFC method plotted against the average entropy decrease for the very same trial pair using the new method. The entropy decrease is consistently larger for the new method suggesting that a single decision for a subject is more informative with the new than with the standard 2AFC method.

However, the observed difference could also be due to the fact that we used a simple adaptive staircase algorithm for the standard 2AFC method. In order to perform a fair comparison, we ran a Monte Carlo simulation of a perceptual discrimination experiment using the same optimal Bayesian selection process for assigning the reference values for both methods. The simulation assumed a model observer who made a forced choice based on samples from the test and the reference stimulus distributions. These distributions had zero bias and were characterized by a single parameter σ reflecting the perceptual noise in both test and reference stimulus. The simulated experiment comprised 1,000 trials and was repeated 1,000 times for each of the two methods. We then estimated the noise parameter σ from the psychometric function and the psychomatrix, respectively, at different trial counts for each iteration of the experiment and computed the variance in the estimated values of σ at these counts. As Figure 6a shows, the variance in the estimated σ was consistently lower for the new method compared to using the standard 2AFC task, suggesting that it is indeed more efficient. We also computed the mean entropy of the posterior parameter probabilities at each trial over the iterations of the simulated experiment for both methods and then plotted the expected entropy decrease for all consecutive trial pairs (n, $n+1$) and for both methods against each other. As shown in Figure 6b, the average entropy decrease is consistently higher when using the new method, confirming that the increased efficiency of the new method is due to the increased information content in each of a subject's individual trial decisions. Our result is in line with the rationale of using multiple alternatives in forced choice methods in order to increase efficiency (Blackwell, 1952; Jäkel & Wichmann, 2006).

Discussion

We have shown that the popular standard 2AFC method is not well constrained when the two stimulus choices differ in more than one stimulus parameter: Subjects, when in doubt, might use differences in secondary stimulus parameters in order to make their decisions. We have theoretically and experimentally demonstrated how such a strategy can lead to decision biases in subjects' choice behavior. These biases distort the psychometric function and thus confound a quantitative characterization of perceptual bias and discrimination threshold in a way that cannot be corrected for. We have introduced a new 2AFC method that prevents such decision biases by design. We successfully validated our new method and compared it to the standard 2AFC method using a visual motion-discrimination experiment. We demonstrated that the method allows for a simultaneous estimation of perceptual bias and the noise parameters of both the test and the reference stimuli.

We found that the new method not only prevents subjects from applying decision strategies that lead to decision biases, but is also more efficient than the standard 2AFC procedure. That is, it requires fewer trials in order to achieve a comparable estimation quality of the underlying model parameters. The extracted discrimination thresholds are comparable to yet slightly higher than those measured with the standard 2AFC method. This is likely caused by an increased representational load due to the additional reference stimulus. Similar increases in discrimination threshold have been reported for forced choice methods with more than two alternatives (Jäkel & Wichmann, 2006). Our new method is general and can be used for a wide variety of psychophysical experiments.

A main goal of this work was to demonstrate and explain why the standard 2AFC method, a very popular method in behavioral neuroscience/psychology, is problematic in experimental conditions in which test and reference stimuli differ in multiple stimulus parameters. These conditions often occur in experiments that are designed to assess perceptual biases, i.e., the effect of a secondary on the percept of a primary stimulus parameter. Studies that have focused on the characterization of perceptual biases have frequently used the standard 2AFC methods, and thus, the reported results might be confounded by decision biases. We would like to emphasize that we do not imply that the results of these studies are necessarily contaminated; many of those measured perceptual biases have been successfully cross-validated with other experiments (e.g., see Hedges, Stocker, & Simoncelli, 2011, comfirming the bias results of Stocker & Simoncelli, 2006 with a categorical task). Also, there are certain instruction techniques that might discourage decision biases. For example, forcing a dissociation of a subject's choice from the identity of the stimulus by turning it into a "yes/no" task might reduce the probability that the subject is applying a mixed decision strategy (e.g., by asking "is the motion in the stimulus to the right more CW than the motion in the stimulus to the left?"; see Figure 1a). However, no matter what instruction is chosen, the standard 2AFC method can never guarantee that subjects do not apply such a strategy. It is also interesting to note that similar concerns have been raised before in the specific context of reported effects of attention on perception (Carrasco et al., 2004; Rahnev et al., 2011; Schneider & Komlos, 2008).

It is important to realize that any standard 2AFC experiment that results in a shift of the psychometric function is necessarily using test and reference stimuli that differ in more than one stimulus parameter and, thus, is potentially prone to decision bias. Stimulus parameters in this context are not limited to immediate features of the stimulus, such as dot coherence, but also include other aspects of a stimulus, such as its location in the visual field. Decision biases in those conditions can reflect the result of an underlying mixed decision strategy that includes a preference for a particular stimulus location. A well-documented example is interval bias, which reflects a subject's preference in the temporal order of test and reference in sequential 2AFC tasks (Klein, 2001; Masin & Fanton, 1989; Nachmias, 2006; Ulrich & Vorberg, 2009; Yeshurun, Carrasco, & Maloney, 2008). These stimulus parameters are often not of interest and thus can be controlled for by randomizing the trials (e.g., randomizing the order in the sequential 2AFC task). This, of course, is not possible when the goal of the experiment in the first place is to characterize the perceptual bias induced by variations in the secondary stimulus parameter!

Psychophysical methods that use more than two stimuli have been suggested for other purposes. Forced choice methods with more than two alternatives were shown to be more efficient than the standard 2AFC method (Blackwell, 1952; DeCarlo, 2012; Gerhard, Wichmann, & Bethge, 2013; Jäkel & Wichmann, 2006; Krauskopf & Gegenfurtner, 1992). However, they are equally problematic in measuring perceptual bias. The triadic judgment bisection task (Falmagne, 1985; Sheu, 2006) was designed to estimate the midway point between two static references; in consumer-preference testing, triads are used to test for a perceived difference between stimuli. Difference scaling (Maloney & Yang, 2003) is similar to our approach in that a subject's choice is based on comparing two perceived distances, but it is not designed to measure perceptual bias caused by secondary stimulus parameters.

The observer model based on SDT represented a good description of the decision process and thus the probability values in the psychomatrix. Some results using magnitude estimation suggest that stimulus judgments based on clearly identifiable stimuli can show significantly larger variability than what is expected from typical discrimination threshold measurements (Laming, 1997). Although we do not exclude the possibility that subjects' decisions are more likely to fluctuate when making suprathreshold comparisons, we did not find any evidence that this is the case in the context of our new method: Subject's decision probabilities were well captured by SDT (see also Devinck & Knoblauch, 2012), and the increases in measured noise levels compared to those extracted with the standard 2AFC method were of similar magnitude as the previously reported increases due to the addition of stimulus alternatives in forced choice methods (Jäkel & Wichmann, 2006). We also found no indication that noncircular stimulus parameters would cause a different decision behavior of subjects performing the new 2AFC task (see Figure S3, *Supplemental Data* for a comparison of characterizing perceived brightness with the new and the standard 2AFC methods). More elaborate signal detection models can be used for modeling decision behavior in our new task. Such models can, for example, include stimulus-dependent noise or nuisance factors, such as lapses and guesses (Wichmann & Hill, 2001). Furthermore, Bayesian observer models represent obvious alternatives that would also allow accounting for prior beliefs of a subject (Stocker & Simoncelli, 2006). In these cases, the entropy method for the efficient allocation of the reference stimuli must be adapted accordingly (see, e.g., Prins, 2013).

Finally, our new method might allow, for the first time, the reliable assessment of perceptual biases in nonhuman primates. This has not been possible using the standard 2AFC method because nonhuman primate subjects, obviously, cannot be verbally instructed to make a choice based only on the perceived difference in the primary stimulus parameter and to ignore differences in secondary parameters. Previous experiments had to rely on "catch trials," i.e., sparsely placed trials in an otherwise balanced experiment. The potential for our new method in studies on nonhuman primates clearly depends on how easy it would be to train the task. This is difficult to judge without actually trying it, but it does not seem infeasible. If successful, it would open the door for a whole new set of interesting studies exploring the neural basis of perceptual biases and its link to behavior.

Methods

Mixed decision strategy model for standard 2AFC tasks

Consider the standard 2AFC experiment shown in Figure 1a in which a subject is presented with a pair of stimuli that differ in their value of a parameter of interest (here, the direction of the overall dot motion θ) and is asked to judge the relative difference between the two stimuli, for example, by indicating which stimulus' motion direction is "*... more counterclockwise than the other*." Let us assume that the two stimuli can also differ in a secondary parameter, the dot coherence c, which takes on the value c_t for the test and c_r for the reference stimulus.

Let $C \in \{r, t\}$ represent a subject's choice. Then $p(C = r|\theta_r, \theta_t, c_r, c_t)$ describes the psychometric function, i.e., the probability with which the subject chooses the reference stimulus given the values of all stimulus parameters. Ideally, this psychometric function should only depend on θ as defined by the task instruction. However, because the two stimuli differ in more than one stimulus parameter, a subject's choice in each trial can be based either correctly on the primary stimulus parameter θ (decision strategy $S = S_\theta$) or incorrectly based on the secondary stimulus parameter c (decision strategy $S = S_c$). We model this mixed decision process by a mixture of choice probabilities under decision strategies S_θ and S_c, respectively (DeCarlo, 2002):

$$
\begin{aligned}
p(C = r|\theta_r, \theta_t, c_r, c_t) &= \sum_S p(C = r, S|\theta_r, \theta_t, c_r, c_t) \\
&= \sum_S p(C = r|S, \theta_r, \theta_t, c_r, c_t) p(S|\theta_r, \theta_t, c_r, c_t). \quad (1)
\end{aligned}
$$

When adopting the correct strategy $S = S_\theta$, subjects choose the reference if they perceive the motion direction of the reference θ_r to be more clockwise than the test θ_t, thus overall with probability $p(\hat{\theta}_r > \hat{\theta}_t)$.[2] Adopting an incorrect strategy $S = S_c$, however, subjects could use any arbitrary criteria based on the secondary parameter. One reasonable assumption is that they pick the stimulus that appears more reliable; in our example, the stimulus with higher perceived dot coherence and thus with probability $p(\hat{c}_r > \hat{c}_t)$. Assuming that the subject's perception of the secondary perceptual parameter is independent of the primary parameter, we can write the choice probabilities under each strategy as

$$
\begin{aligned}
&p(C = r|S = S_\theta, \theta_r, \theta_t, c_r, c_t) \\
&\quad = p(\hat{\theta}_r > \hat{\theta}_t|\theta_r, \theta_t, c_r, c_t)
\end{aligned}
$$

and

$$p(C = r|S = S_c, \theta_r, \theta_t, c_r, c_t) = p(\hat{c}_r > \hat{c}_t|c_r, c_t),$$

respectively. Figure 7a and b shows the two choice probabilities according to SDT. For reasons of simplicity, we assume constant, zero-mean Gaussian noise in the representations of both the test and the reference stimulus in both stimulus dimensions ($\sigma_\theta = 0.3$, $\sigma_c = 2$ [a.u.]). We further assume that the reference has higher coherence than the test stimulus ($c_r > c_t$, $\Delta_c = 0.6$ standard deviations).

We model the probability of the subject adopting one or the other decision strategy $p(S|\theta_r, \theta_t, c_r, c_t)$ as the ratio of the relative uncertainties in making the decision based on the primary or the secondary parameter, respectively. We express these uncertainties as the entropies of the corresponding criterions for each decision strategy, namely

$$
\begin{aligned}
H_\theta &= H(p(\hat{\theta}_r > \hat{\theta}_t)) \\
H_c &= H(p(\hat{c}_r > \hat{c}_t)),
\end{aligned}
$$

where $H(p) = p \log_2(p) + (1 - p) \log_2(1 - p)$. As the two events $S = S_\theta$ and $S = S_c$ are mutually exclusive, we can define the probability of the subject adopting a decision strategy S as

$$
\begin{aligned}
p(S = S_\theta|\theta_r, \theta_t, c_r, c_t) &= \frac{1 - H_\theta}{(1 - H_\theta) + \alpha(1 - H_c)} \\
p(S = S_c|\theta_r, \theta_t, c_r, c_t) &= 1 - p(S = S_\theta|\theta_r, \theta_t, c_r, c_t),
\end{aligned}
$$

where α modulates a general tendency of the subject adopting the correct strategy. As can be seen in Figure 7c, $p(S = S_\theta)$ approaches a minimum when the reference and the test stimulus are close in θ, increasing thus the

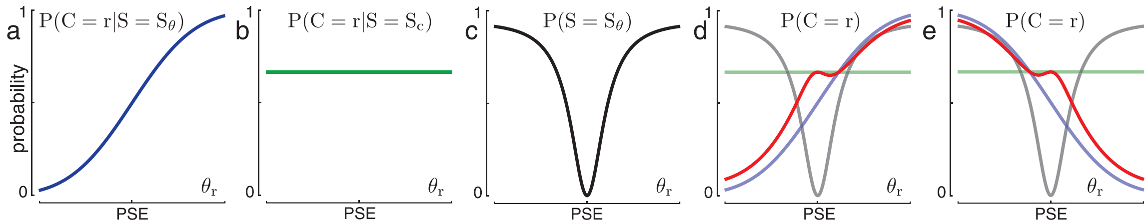

Figure 7. Mixed decision strategy model for standard 2AFC. (a) Probability of a subject choosing the reference under correct strategy S_θ if the subject is asked to indicate which stimulus' θ was "*… more counterclockwise than the other.*" (b) Probability of a subject choosing the reference under incorrect strategy S_c, assuming the subject prefers the stimulus with larger coherence c (here, $c_r > c_t$). (c) Probability of a subject using the correct strategy $S = S_\theta$ ($\alpha = 1$). (d) Choice probability for mixed strategy (red curve). The psychometric function is distorted and shifted to the left. (e) Same as in (d) but with the subject instructed to indicate which stimulus' θ was "*… more clockwise than the other.*"

probability of a subject adopting the incorrect strategy S_c.

We obtain the total choice probability $p(C = r)$ by applying Equation 1:

$$\begin{aligned}
p(C = r|\theta_r, \theta_t, c_r, c_t)& \\
= p(C = r|S = S_\theta, \theta_r, \theta_t, c_r, c_t)& \\
p(S = S_\theta|\theta_r, \theta_t, c_r, c_t)& \\
+ p(C = r|S = S_c, \theta_r, \theta_t, c_r, c_t)& \\
p(S = S_c|\theta_r, \theta_t, c_r, c_t)& \\
= p(\hat{\theta}_r > \hat{\theta}_t)\left[\frac{1 - H_\theta}{(1 - H_\theta) + \alpha(1 - H_c)}\right]& \\
+ p(\hat{c}_r > \hat{c}_t)\left[1 - \frac{1 - H_\theta}{(1 - H_\theta) + \alpha(1 - H_c)}\right]&.
\end{aligned}$$

The choice probability is shown in panel 7d (red curves) and exhibits the expected distortion due to a mixed decision strategy. Fitting these distorted psychometric functions with standard sigmoidal functions would result in a biased estimate of the PSE, which we refer to as *decision bias* (see Figure 1b). The direction of the decision bias depends both on the criterion that the subjects use under strategy $S = S_c$ and on the sign of the difference $\Delta_c = c_r - c_t$. Furthermore, it also depends on the instruction the subject was given; if the subject was instructed to indicate which stimulus' motion direction was "*… more clockwise than the other,*" the choice probability for the reference under the correct decision strategy S_θ is given by $p(\hat{\theta}_r < \hat{\theta}_t)$, leading to the total choice probability shown in Figure 7e.

Observer model for the new 2AFC method

We use SDT (Green & Sweets, 1966; Macmillan & Creelman, 2005) to derive an observer model for our 2AFC method. We assume that an observer's percept of the primary stimulus parameter for the two reference and the test stimuli $(\hat{\theta}_{r_1}, \hat{\theta}_{r_2}, \hat{\theta}_t)$ is corrupted by Gaussian noise with standard deviations σ_r, σ_t, respectively. Also, because we measure perceptual bias relative to the reference stimuli, we assume the expected percept of the reference stimuli to reflect their true values while the expected percept of the test $\langle\hat{\theta}_t\rangle$ is potentially biased and unknown. Under these assumptions, we have a full description of the perceptual distributions for each stimulus as $p(\hat{\theta}_{r_1}|\theta_{r_1}) \equiv \mathcal{N}(\theta_{r_1}, \sigma_r)$, $p(\hat{\theta}_{r_2}|\theta_{r_2}) \equiv \mathcal{N}(\theta_{r_2}, \sigma_r)$ and $p(\hat{\theta}_t|\theta_t) \equiv \mathcal{N}(\langle\hat{\theta}_t\rangle, \sigma_t)$ (see Figure 2c).

We then can express the probability that the subject chooses the second reference stimulus as being closer to the test than the first reference, as

$$\begin{aligned}
&p(|\Delta(\hat{\theta}_{r_1}, \hat{\theta}_t)| > |\Delta(\hat{\theta}_{r_2}, \hat{\theta}_t)|) \\
&= \int_0^{2\pi} p(\hat{\theta}_t|\theta_t) \int_{\hat{\theta}_t}^{2\pi} p(\hat{\theta}_{r_2}|\theta_{r_2}) \int_0^{2\hat{\theta}_t - \hat{\theta}_{r_2}} p(\hat{\theta}_{r_1}|\theta_{r_1}) d\theta_{r_1} d\theta_{r_2} d\theta_t \qquad (2) \\
&+ \int_0^{2\pi} p(\hat{\theta}_t|\theta_t) \int_{\hat{\theta}_t}^{2\pi} p(\hat{\theta}_{r_2}|\theta_{r_2}) \int_{\hat{\theta}_{r_2}}^{2\pi} p(\hat{\theta}_{r_1}|\theta_{r_1}) d\theta_{r_1} d\theta_{r_2} d\theta_t \\
&+ \int_0^{2\pi} p(\hat{\theta}_t|\theta_t) \int_0^{\hat{\theta}_t} p(\hat{\theta}_{r_2}|\theta_{r_2}) \int_0^{\hat{\theta}_{r_2}} p(\hat{\theta}_{r_1}|\theta_{r_1}) d\theta_{r_1} d\theta_{r_2} d\theta_t \\
&+ \int_0^{2\pi} p(\hat{\theta}_t|\theta_t) \int_0^{\hat{\theta}_t} p(\hat{\theta}_{r_2}|\theta_{r_2}) \int_{2\hat{\theta}_t - \hat{\theta}_{r_2}}^{2\pi} p(\hat{\theta}_{r_1}|\theta_{r_1}) d\theta_{r_1} d\theta_{r_2} d\theta_t.
\end{aligned}$$

Each of the summands in the above equation represents one of the four possible orders within which a subject can perceive the second reference stimulus as being closer than the first reference, i.e., $(\hat{\theta}_{r_1} > \hat{\theta}_t > \hat{\theta}_{r_2})$, $(\hat{\theta}_t > \hat{\theta}_{r_2} > \hat{\theta}_{r_1})$, $(\hat{\theta}_{r_1} > \hat{\theta}_{r_2} > \hat{\theta}_t)$, and $(\hat{\theta}_{r_2} > \hat{\theta}_t > \hat{\theta}_{r_1})$, respectively. For a given θ_t, Equation 2 describes the decision probability as represented by the *psychomatrix* as a function of the three parameters σ_r, σ_t, and $\langle\hat{\theta}_t\rangle$ (Figure 2d). In practice, we estimate these parameter values using a maximum-likelihood fit of the above observer model to the psychomatrix, i.e., the empirical choice probabilities over repeated trials at discrete value pairs $(\theta_{r_1}, \theta_{r_2})$ assuming binomial choice distributions.

The notation of the observer model (Equation 2) can be simplified by noting that, in order for $\hat{\theta}_{r_2}$ to be perceptually closer to $\hat{\theta}_t$ than $\hat{\theta}_{r_1}$, $\hat{\theta}_t$ has to lie above the mean of the perceived values of the two references when $\hat{\theta}_{r_2} > \hat{\theta}_{r_1}$ and, conversely, below this point when $\hat{\theta}_{r_2} < \hat{\theta}_{r_1}$. These constraints can be expressed as

$$\begin{aligned} &p(|\Delta(\hat{\theta}_{r_1}, \hat{\theta}_t)| > |\Delta(\hat{\theta}_{r_2}, \hat{\theta}_t)|) \\ &\quad = p(\hat{\theta}_{r_1} - \hat{\theta}_{r_2} > 0|\theta_{r_1}, \theta_{r_2})\, p\left(\frac{\hat{\theta}_{r_1} + \hat{\theta}_{r_2}}{2} > \hat{\theta}_t|\theta_{r_1}, \theta_{r_2}, \theta_t\right) \\ &\quad + p(\hat{\theta}_{r_1} - \hat{\theta}_{r_2} < 0|\theta_{r_1}, \theta_{r_2})\, p\left(\frac{\hat{\theta}_{r_1} + \hat{\theta}_{r_2}}{2} < \hat{\theta}_t|\theta_{r_1}, \theta_{r_2}, \theta_t\right), \end{aligned}$$

which is equivalent to

$$\begin{aligned} &p(|\Delta(\hat{\theta}_{r_1}, \hat{\theta}_t)| > |\Delta(\hat{\theta}_{r_2}, \hat{\theta}_t)|) \\ &\quad = p(\hat{\theta}_{r_1} - \hat{\theta}_{r_2} > 0|\theta_{r_1}, \theta_{r_2}) \\ &\qquad p(\hat{\theta}_{r_1} + \hat{\theta}_{r_2} - 2\hat{\theta}_t > 0|\theta_{r_1}, \theta_{r_2}, \theta_t) \\ &\qquad + p(\hat{\theta}_{r_1} - \hat{\theta}_{r_2} < 0|\theta_{r_1}, \theta_{r_2}) \\ &\qquad\quad p(\hat{\theta}_{r_1} + \hat{\theta}_{r_2} - 2\hat{\theta}_t < 0|\theta_{r_1}, \theta_{r_2}, \theta_t). \end{aligned}$$

With substitutions $u = \theta_{r_1} - \theta_{r_2}$ and $v = \theta_{r_1} + \theta_{r_2} - 2\theta_t$, we get

$$\begin{aligned} &p(|\Delta(\hat{\theta}_{r_1}, \hat{\theta}_t)| > |\Delta(\hat{\theta}_{r_2}, \hat{\theta}_t)|) \\ &\quad = p(\hat{u} > 0|u)p(\hat{v} > 0|v) + p(\hat{u} < 0|u)p(\hat{v} < 0|v). \end{aligned}$$

The above substitutions rotate the axes of the decision probability space so that u now represents the direction orthogonal to the positive diagonal in the psychomatrix or the difference between the two references, and v represents the direction orthogonal to the negative diagonal or the mean of the two references. With our initial assumptions about perceptual distributions, $p(\hat{u} > 0 \mid u)$, $p(\hat{v} > 0|v)$, $p(\hat{u} < 0 \mid u)$, and $p(\hat{v} < 0|v)$ are cumulative probabilities of the normal distributions

$$p(\hat{u}|u) \equiv \mathcal{N}(\mu_u, \sigma_u)\,;\, \mu_u = 0,\ \sigma_u = \sqrt{2\sigma_r^2}$$

$$\begin{aligned} p(\hat{v}|v) \equiv \mathcal{N}(\mu_v, \sigma_v)\,;\, \mu_v &= 2(\langle\hat{\theta}_t\rangle - \theta_t), \\ \sigma_v &= \sqrt{2\sigma_r^2 + 4\sigma_t^2}\,. \end{aligned}$$

Efficient allocation of reference stimuli

Having two reference stimuli requires an efficient procedure to selectively sample the psychomatrix. We resort to an adaptive technique that optimally selects the values of the two reference stimuli r_1 and r_2 in each trial such that the expected information gain by the subject's choice is maximal. We modified the procedure proposed by Kontsevich and Tyler (1999), originally designed for one-dimensional psychometric functions, for our new 2AFC method. For each trial, the procedure selects the optimal values $(\theta_{r_1}, \theta_{r_2})$ of the two references that will lead to the biggest expected reduction in entropy of $p(\lambda)$ given the two possible choices $C \in \{r_1, r_2\}$ of the subject, where $\lambda = (\sigma_r, \sigma_t, \langle\hat{\theta}_t\rangle)$ represents the unknown parameters of the observer model. The procedure is as follows:

- Precompute $p(C|\lambda, \theta_{r_1}, \theta_{r_2})$ for all discretized values of λ and $\theta_{r_{1,2}}$ (using SDT, Equation 2).
- Initialize p(λ) as a uniform prior
- For each trial,
  1. Compute probabilities of response
  $$p(C|\theta_{r_1}, \theta_{r_2}) = \sum_\lambda p(C|\lambda, \theta_{r_1}, \theta_{r_2})p(\lambda)$$
  2. Compute the posterior probability by Bayesian inference
  $$p(\lambda|\theta_{r_1}, \theta_{r_2}, C) = \frac{p(C|\lambda, \theta_{r_1}, \theta_{r_2})p(\lambda)}{p(C|\lambda, \theta_{r_1}, \theta_{r_2})}$$
  3. Estimate the entropy of the probability density over the parameter space given a trial at $\theta_{r_1}, \theta_{r_2}$ and response C
  $$\begin{aligned} &H(\theta_{r_1}, \theta_{r_2}, C) \\ &\quad = -\sum_{\lambda,C} p(\lambda|\theta_{r_1}, \theta_{r_2}, C)\log(p(\lambda|\theta_{r_1}, \theta_{r_2}, C)) \end{aligned}$$
  4. Compute the expected entropy
  $$E[H(\theta_{r_1}, \theta_{r_2})] = \sum_C H(\theta_{r_1}, \theta_{r_2}, C)p(C|\theta_{r_1}, \theta_{r_2})$$
  5. Run a trial at $\theta^*_{r_1}$ and $\theta^*_{r_2}$ that minimizes the expected entropy in step 4
  6. Use the subject's response C^* to update the prior
  $$p(\lambda) = p(\lambda|\theta_{r_1} = \theta^*_{r_1}, \theta_{r_2} = \theta^*_{r_2}, C^*)$$

Figure 8a shows an example of the expected entropy pattern in one actual trial of the motion-discrimination experiment. It illustrates that the areas with lowest expected entropy are on either side of the positive diagonal at both ends and in the upper right (and lower left) corner. To prevent excessive repeats of placing the references at the same locations, a minimal perturbation step was introduced by randomly choosing r_1 and r_2 from the coordinate pairs with the four smallest

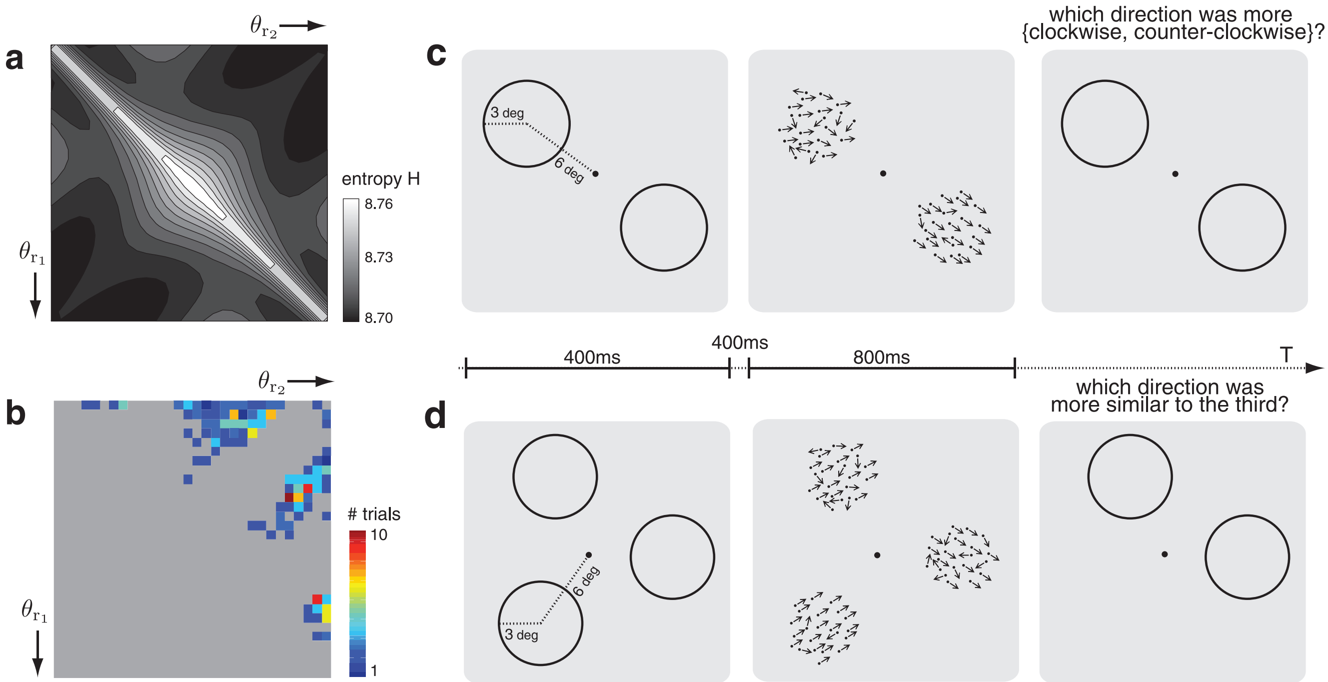

Figure 8. Adaptive trial allocation. (a) Expected entropy $H(\lambda|\theta_{r_1}, \theta_{r_2})$ of the model parameters displayed for any reference pair $(\theta_{r_1}, \theta_{r_2})$ for a randomly picked trial. Gray scale reflects the entropy level from high (white) to low values (black). (b) A distribution of trials for one of the subjects who completed 216 trials with the new method. The adaptive allocation procedure leads to an optimal sampling of the psychomatrix. (c) Sequence of a single trial using the standard 2AFC method: Subjects are first cued where the two stimuli will appear with circular apertures. After the random-dot motion stimuli were presented for 800 ms, the two apertures reappear to indicate the choice alternatives, and the subject has to select which stimulus showed an overall motion direction that was more CW (members of the CW group) or CCW (members of the CCW group). (d) Sequence for the new 2AFC method: Again, subjects are first cued where the three stimuli will appear (relative positions are fixed, but overall position randomly rotated across trials). After stimulus presentation, *two* apertures reappear indicating the two reference stimuli. Subjects have to select the aperture that represents the reference that appeared more similar to the test.

entropies. Figure 8b shows the final distribution after 216 trials of a real experiment also exploiting the point symmetry across the positive diagonal. It reveals the very selective sampling of the psychomatrix as a result of the adaptive procedure, which nicely matches the locations of lowest entropy shown in Figure 8a. The algorithm is computationally demanding, which can impose constraints on how finely and over what range λ and θ can be discretized, depending on experimental requirements and hardware resources. If needed, sampling methods could help to reduce the computational demands (Kujala & Lukka, 2006).

Experimental procedures

Thirteen human subjects took part in the experiment (nine females, four males). All subjects except one were naïve to the purpose of the experiment. Participants had normal or corrected-to-normal vision, and all gave informed consent prior to the experiment. The study was approved by the University of Pennsylvania Institutional Review Board (protocol #813601). Subjects were sitting in a darkened room, and their head position was controlled with a chin rest. Stimuli were displayed at a distance of 60 cm on a Samsung Dell P992 CRT 17-in. computer display with a 120-Hz refresh rate and 1024 × 768 pixel resolution. Gamma was corrected. The experiment was programmed in MATLAB (Mathworks, Inc.) with display routines from the MGL toolbox (http://justingardner.net/mgl) and was running on an Apple Mac Pro computer with a 2.93-GHz quad-core Intel Xeon processor.

Each trial consisted of a sequence of cue (400 ms), blank (400 ms), stimulus (800 ms), and response displays (Figure 8c, d). A fixation mark (luminance 40 cd/m^2) on a uniform background (10 cd/m^2) was displayed throughout the trial, and subjects were instructed to maintain fixation for the whole duration of the trial. The cue display contained circular shapes (3° in diameter, 40 cd/m^2) that indicated where the test and the reference stimuli would appear yet without disclosing their identities. Locations of the test were sampled from a uniform distribution around the circle (radius 6°) at 100 discrete values, and locations of the references were assigned relative to the test at the same eccentricity (a) opposite the test location for the standard 2AFC method (Figure 8c) and (b) at two locations that formed an equilateral triangle for our new method (Figure 8d). Random dot stimuli were composed of three interleaved frames, each with its own independent set of moving dots (but with the same direction of coherent motion). Each frame had 20 white dots (90 cd/m^2) moving at 4°/s behind a 3° aperture. Each dot had a lifetime of 500 ms, after which it reappeared at a random location. A dot that moved out of the aperture reappeared at the opposite boundary. A percentage of dots moved coherently in the stimulus direction, and the remaining dots had random directions. Test stimulus direction θ_t was sampled from a uniform distribution at 70 equally spaced discrete values covering 360° and excluding the cardinal directions. The direction of the reference (relative to the test) in the standard 2AFC task was governed by two one-up/one-down staircases. Directions of the two references in the new method were governed by the adaptive procedure described in the previous subsection. For the standard 2AFC method, after simultaneous exposure of reference and test (800 ms), both circular cues reappeared, and subjects had to choose between the two with a mouse click. Five subjects had to indicate "which of the two stimuli had a direction of motion that is more counterclockwise," and eight other subjects had to indicate "which of the two stimuli had a direction of motion that is more clockwise." For the new method, after simultaneous exposure of two references and a test (800 ms), only two circular cues reappeared, indicating the positions of the two references, and the subjects had to select the reference that answered the question, "Which one of the two cued stimuli had a motion direction that was most similar to the motion direction of the noncued stimulus?" A noninformative auditory feedback was given after the mouse click except in training when a lower pitch tone was played for wrong answers.

There were three hour-long sessions devoted to the standard 2AFC method, followed by three sessions devoted to the new method. During each session, subjects completed two blocks with 252 trials. All blocks had randomly interleaved trials from seven conditions: four conditions that were balanced in the secondary dimension with $(c_r, c_t) = (20,20)$, (30,30), (45,45), (90,90)% and three unbalanced conditions with $(c_r, c_t) = (30,20)$, (30,45), (30,90)%. Each subject completed 216 trials for each condition, totaling 1512 trials per method. Subjects were trained at the beginning of using each of the two methods, by providing auditory feedback on easy trials (<100 trials). Subsequent sessions started with only a small number (≈10) of training trials to refresh the subject's memory of the task. All training trials were separated and were not included in the analysis.

Data analysis and fits

For our analysis, we excluded data sets (entire psychometric functions) of individual subjects for conditions for which the fitted noise levels σ_θ exceeded 90° (nine sets out of a total of 91, seven of the nine sets were of the noisy conditions [30, 20] or [30, 30], and all excluded sets were from using the standard 2AFC method). The average subject was obtained by pooling the remaining sets across subjects. After recruiting the first 10 subjects (five for each group), we found that most of the excluded data belonged to the "clockwise" group. We therefore added three more subjects to this group for a total of 13 subjects.

Biases in Figure 3a were extracted via a maximum likelihood fit of the individual psychometric functions with a cumulative Gaussian. Biases for the new method were extracted using the fit $\hat{\theta}_t$ values of the observer model (Equation 2) for all unbalanced conditions. Noise parameters in Figure 3d were extracted from balanced conditions only. For the standard 2AFC method, they represent $\sigma/\sqrt{2}$ of a cumulative Gaussian fit to the psychometric functions assuming zero bias, and noise parameters for the new method were obtained by fitting the observer model with one free σ parameter and zero bias. Significance value ($p = 10^{-3}$) for the different group behavior in Figure 3a was calculated as multiple nonparametric one-tailed tests, testing the null hypothesis that all biases of the CCW group are smaller than biases for the CW group when $\Delta_c > 0$ and vice versa when $\Delta_c < 0$. Error bars in Figure 3 represent the 95% confidence interval over the fits to 1,000 bootstrapped samples of the data (Efron, 1979).

The analysis shown in Figure 4 was performed on the combined data of subjects for which the average fraction of trials placed on the boundary of the psychomatrix (BI score) did not exceed 0.9. Relative biases for the (0°, 90°) quadrant in Figure 4a were extracted at 5° intervals. Each interval pooled data from four principal directions mirrored across the horizontal and vertical cardinals, and for each direction, data was pooled across four neighboring test directions (a 15° range). After collapsing, the psychomatrix for each interval included 48 trials per subject for a total of 144 trials for all subjects used in the analysis. Noise parameters in Figure 4b were estimated as averages and standard deviations of the fitted σ_r and σ_t across all intervals. Data from the same subjects and the same data pooling was used in the analysis of the standard 2AFC method; thus, individual values for bias and noise level were extracted from psychometric functions reflecting also 144 trials.

Simulations

Simulations of both methods (Figures 5 and 6) are based on a signal detection model observer who makes a decision based on values of reference and test stimuli that are sampled from Gaussian distributions with widths σ_r and σ_t, respectively. We fit the resulting psychometric functions (for standard 2AFC) with a cumulative Gaussian and the psychomatrix (for the new method) with the observer model (Equation 2). Simulations of the new method for different ranges of the psychomatrix (Figure 5) used the following parameter values: $\sigma_r = 25°$, $\sigma_t = 20°$, and zero perceptual bias. For all ranges, the size of the psychomatrix was 30 × 30 with equal sampled reference intervals. Each condition was repeated 100 times with 216 trials in each repetition. Simulations for the convergence analysis (Figure 6a) had the following model parameters: $\sigma_{r,t} = 30°$ and zero perceptual bias. We fitted the data and extracted parameters after 80, 100, 150, 200, 500, and 1,000 trials 1,000 times and plotted the width of the 95% confidence interval based on 1,000 samples of $\sigma_{r,t}$.

Keywords: psychophysics, perceptual bias, discriminability, two-alternative forced choice (2AFC), methods

Acknowledgments

This work was supported by the Office of Naval Research (grant N000141110744). The authors would like to thank the members of the Computational Perception and Cognition lab, Toni Saarela, Xuexin Wei, and Alex Tank for helpful comments and discussions as well as Saul Sternberg and Josh Gold for their feedback on earlier versions of the manuscript. Some of the results were presented at the Annual Symposium of the Visual Science Society in May 2012.

Commercial relationships: none.
Corresponding author: Matjaž Jogan.
Email: mjogan@sas.upenn.edu.
Address: Computational Perception and Cognition Laboratory, University of Pennsylvania, Philadelphia, PA, USA.

Footnotes

[1]Throughout this paper, we use $\hat{\theta}$ to denote a subject's perceived value of a stimulus parameter θ.

[2]Here, > means "more clockwise."

References

Blackwell, H. R. (1952). Studies of psychophysical methods for measuring visual thresholds. *Journal of the Optical Society of America, 42*(9), 606–614.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A*, *14*(7), 1393–1411.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*(3), 308–313.

Dakin, S. C., & Alais, D. (2010). Anisotropies in judging the direction of moving natural scenes. *Journal of Vision*, *10*(11):5, 1–19, http://www.journalofvision.org/content/10/11/5, doi:10.1167/10.11.5. [PubMed] [Article]

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review*, *109*(4), 710–721.

DeCarlo, L. T. (2012). On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*, *56*(3), 196–207.

Devinck, F., & Knoblauch, K. (2012). A common signal detection model accounts for both perception and discrimination of the watercolor effect. *Journal of Vision, 12*(3):19, 1–14, http://www.journalofvision.org/content/12/3/19, doi:10.1167/12.3.19. [PubMed] [Article]

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*(1), 1–26.

Falmagne, J. C. (1985). *Elements of psychophysical theory. Number 6 in Oxford Psychology Series*. New York: Oxford University Press.

Fechner, G. (1860/1966). *Elements of psychophysics.* New York: Holt, Rinehart and Winston. (Original work published 1860)

García-Pérez, M. A., & Alcala-Quintana, R. (2011). Improving the estimation of psychometric functions in 2AFC discrimination tasks. *Frontiers in Psychology, 2*, 1–9.

Gerhard, H. E., Wichmann, F. A., & Bethge, M. (2013). How sensitive is the human visual system to the local statistics of natural images? *PLoS Computational Biology, 9*(1), e1002873.

Girshick, A., Landy, M., & Simoncelli, E. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14*(5), 926–932.

Green, D., & Sweets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hedges, J., Stocker, A., & Simoncelli, E. (2011). Optimal inference explains the perceptual coherence of visual motion stimuli. *Journal of Vision, 11*(6):14, 1–16, http://www.journalofvision.org/content/11/6/14, doi:10.1167/11.6.14. [PubMed] [Article]

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):13, 1307–1322, http://www.journalofvision.org/content/6/11/13, doi:10.1167/6.11.13. [PubMed] [Article]

Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics, 63,* 1377–1388.

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63*(8), 1421–1455.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. New York: Cambridge University Press.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39*(16), 2729–2737.

Krauskopf, J., & Gegenfurtner, K. R. (1992). Color discrimination and adaptation. *Vision Research*, *11*(32), 2165–2175.

Kujala, J., & Lukka, T. (2006). Bayesian adaptive estimation: The next dimension. *Journal of Mathematical Psychology*, *50*(4), 369–389.

Laming, D. (1997). *The measurement of sensation*. New York: Oxford University Press.

Lesmes, L. A., Jeon, S.-T., Lu, Z.-L., & Dosher, B. A. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method. *Vision Research, 46*(19), 3160–3176.

Loffler, G., & Orbach, H. S. (2001). Anisotropy in judging the absolute direction of motion. *Vision Research, 41,* 3677–3692.

Macmillan, N., & Creelman, C. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision, 3*(8): 5, 573–585, http://www.journalofvision.org/content/3/8/5, doi:10.1167/3.8.5. [PubMed] [Article]

Masin, S., & Fanton, V. (1989). An explanation for the presentation-order effect in the method of constant stimuli. *Attention, Perception & Psychophysics, 46,* 483–486, doi:10.3758/BF03210864.

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.

Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Research*, *46*(15), 2456–2464.

Pelli, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *9*(2), 1508–1532.

Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, *13*(7): 3, 1–17, http://www.journalofvision.org/content/13/7/3, doi:10.1167/13.7.3. [PubMed] [Article]

Putzeys, T., Bethge, M., Wichmann, F., Wagemans, J., & Goris, R. (2012). A new perceptual bias reveals suboptimal population decoding of sensory responses. *PLoS Computational Biology*, *8*(4), e1002453.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience, 14,* 1513–1515.

Rauber, H.-J., & Treue, S. (1998). Reference repulsion when judging the direction of visual motion. *Perception, 27,* 393–402.

Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of Vision, 8*(15):3, 1–10, http://www.journalofvision.org/content/8/15/3, doi:10.1167/8.15.3. [PubMed] [Article]

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, *8*(7), 522–535.

Sheu, C.-F. (2006). Triadic judgment models and Weber's law. *Journal of Mathematical Psychology*, *50*(3), 302–308.

Sims, J. A., & Pelli, D. (1987). The ideal psychometric procedure. *Investigative Ophthalmology and Visual Science (Supplement), 28,* 366.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680.

Stocker, A., & Simoncelli, E. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 4*(9), 578–585.

Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre-surround antagonism in visual motion processing. *Nature*, *424*(6946), 312–315.

Tomassini, A., Morgan, M. J., & Solomon, J. A. (2010). Orientation uncertainty reduces perceived obliquity. *Vision Research, 50*(5), 541–547.

Ulrich, R., & Vorberg, D. (2009). Estimating the difference limen in 2AFC tasks: Pitfalls and improved estimators. *Attention, Perception, & Psychophysics, 6*(71), 1219–1227.

Watson, C., Kellogg, S., Kawanishi, D., & Lucas, P. (1973). The *uncertain* response in detection-oriented psychophysics. *Journal of Experimental Psychology,* (99), 180–185.

Webb, B. S., Ledgeway, T., & McGraw, P. V. (2010). Relating spatial and temporal orientation pooling to population decoding solutions in human vision. *Vision Research*, *50*(22), 2274–2283.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293–1313.

Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research*, *17*(48), 1837–1851.