

Data supplement

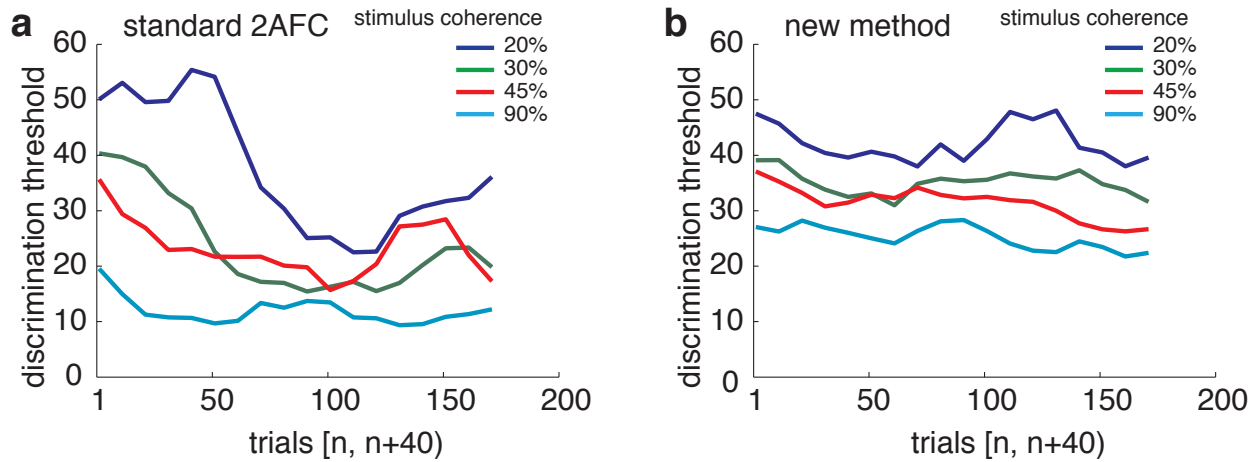


Figure S1: **Effects of learning over the time-course of the visual motion discrimination experiments.** Subjects first performed the standard 2AFC experiment followed by the experiment using the new method. To study effects of learning in both experiments we fitted sequential intervals of data from balanced stimulus conditions using a sliding window of 40 trials in ten-trial increments (e.g., trial intervals [1, 40], [11, 50], ...). **(a)** Noise levels for the standard 2AFC method were significantly higher at the beginning of the experiment compared to the later stages. **(b)** Noise levels for the new method did not decrease significantly during the experiment, suggesting that the rate of learning was lower once the first experiment was completed.

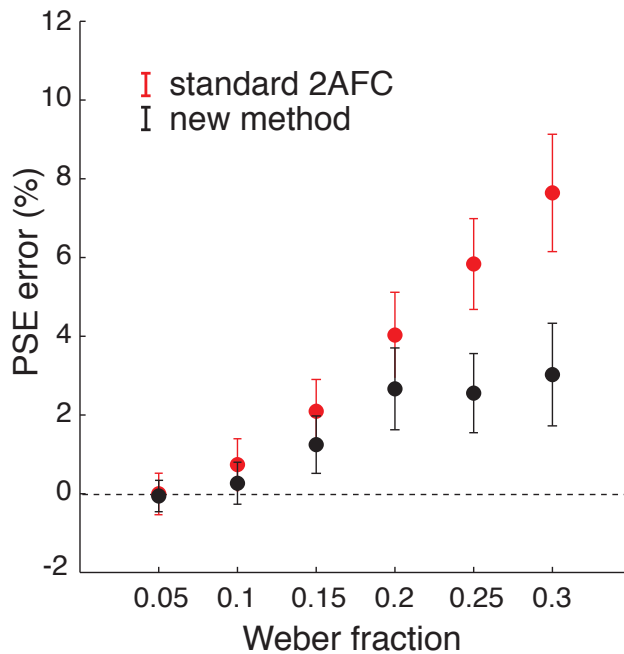


Figure S2: **Violation of the constant noise assumption.** We tested to what degree assuming signal independent, constant noise can lead to errors in the bias estimation if the observer's noise is actually stimulus dependent. For that purpose, we simulated an observer performing a discrimination experiment using both the standard and our new 2AFC method. We assumed that the observer's sensory noise was stimulus dependent according to different Weber fractions. We then fit the simulated psychometric functions and the psychomatrices, respectively, with a SDT model that assumed constant noise, and estimated the error in determining the PSE values from that fit. As expected, the error increases with increasing Weber fraction, thus with an increasing violation of the constant noise assumption. However, the errors are consistently smaller for our new method compared to the standard 2AFC method. This advantage is due to the 2D geometry of the psychomatrix: as the PSE lies on a diagonal, the Weber fraction has to be larger by a factor of $\sqrt{2}$ for an equivalent change in the PSE. Simulations for both methods used identical parameters. In particular, the total number of simulated trials was the same (psychomatrix of size 17x17 with 10 trials per reference pair compared to a psychometric curve with 17 reference values and 170 trials per reference value). Errorbars represent the standard deviations over 100 repetitions of the simulated experiment.

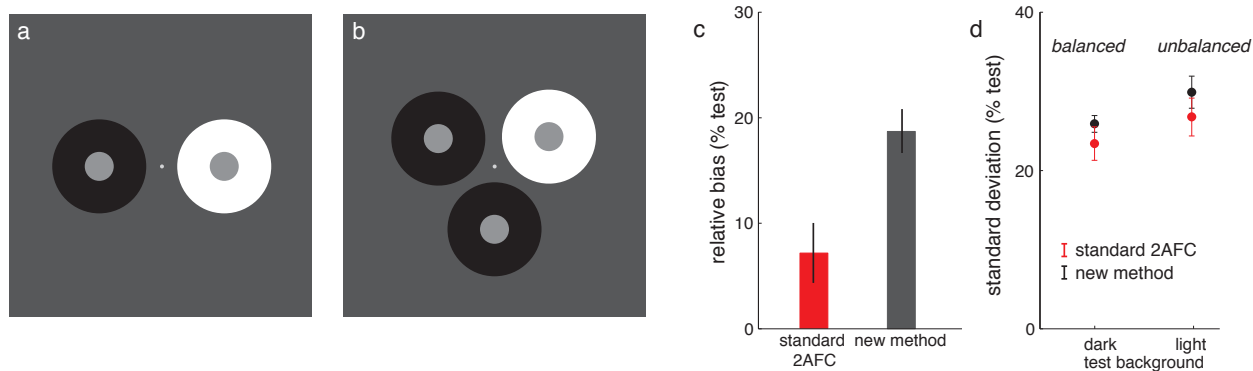


Figure S3: **Comparison between the standard and the new 2AFC method in measuring perceived luminance.** **(a)** Standard 2AFC task: A trained subject discriminated the brightness between two circular stimuli (diameter 2 deg, 4 deg eccentricity) each surrounded by a circular background (diameter 6 deg). In the balanced condition (not shown), both the test (fixed luminance 50 cd/m²) and the reference (luminance 5-95 cd/m², regulated by an adaptive staircase) stimulus were presented on identical, dark backgrounds (luminance 2.5 cd/m²). In the *unbalanced* condition, the background of the reference stimulus was as in the balanced condition, yet the test stimulus was presented on a light background (100 cd/m²). **(b)** The same two conditions were tested with the new method where in the unbalanced condition, both reference stimuli had identical, dark backgrounds. Subject had to choose which one of the two reference stimuli appeared more similar to the test in terms of their brightness. **(c)** With both methods we found a bias in perceived luminance that matches the known illusion that the test is perceived brighter when surrounded with a dark rather than a light background. However, the bias estimated with the new method was significantly larger (error bars represent the standard deviation over 100 bootstrapped samples of the data) **(d)** The discrimination thresholds (measured as the standard deviation of the test distribution from the SDT model fit) was lower for the standard 2AFC method for both dark and light backgrounds, but not by much. This is similar to what we found for the motion direction experiment in the main text, and thus suggests that the decision behavior is independent of whether the stimulus variable is circular (motion direction) or not (luminance). Crucially, the magnitude of these threshold increases is comparable to the magnitudes found in previous studies that were engaging subjects in a fine discrimination task with multiple choices, suggesting that the comparison task in our new method does not involve decision strategies that are beyond signal detection theory.