

Preregistration document for the color-discrimination project using AEPsych for trial placement and Wishart Process model for ad-hoc model fitting

Fangfang Hong¹, Ruby Bouhassira¹, Jason Chow², Craig Sanders², Michael Shvartsman², Alex Williams³, Phillip Guan², David Brainard¹

¹University of Pennsylvania, Department of Psychology

²Reality Lab Research, Meta

³New York University, Center of Neural Science

Preliminary results

As outlined in the preregistration document (Version 2), we collected and analyzed a full dataset from a single subject (Subject 1). The subject completed 12,000 trials, evenly split between AEPsych-based trials and method of constant stimuli (MOCS) trials. Of the 6,000 AEPsych trials, 900 were Sobol trials, and 5,100 were EAVC-based trials (**Fig. 1A**). We fit the Wishart model to the AEPsych trials (**Fig. 1B**) and evaluated the trial placement efficiency of the EAVC-based method. While the efficiency is comparable to our simulation results, it appears slightly lower (**Fig. 1C**). We suspect this discrepancy may be due to how the experiment was implemented: in our real data collection, separate threads were used to track elapsed time and to process AEPsych computations, whereas in the simulations, threading was not split. Thus there may have been less compute time available for AEPsych in the experiment.

For the MOCS trials, we fit a Weibull psychometric function to each condition (**Fig. 2A-B**) and compared the estimated thresholds to those predicted by the Wishart model fits. The results show strong agreement in aggregate with no bias. The deviations are slightly larger than those observed in simulations (**Fig. 2C**).

Minor modifications to apparatus, procedure, design, and analysis

We will now proceed to study 7 additional subjects with the same basic design. We will make a few modifications as described next.

Instead of using the Mac Studio, we will conduct AEPsych calculations on a high-performance PC equipped with an NVIDIA GeForce RTX 3070 GPU and a 12-core Intel i9-10920X processor. This decision was driven by the discovery that AEPsych's trial placement was significantly slowed down when using multithreading on Mac. During pilot testing, we observed that towards the end of each session, a large number of MOCS trials were inserted while waiting for AEPsych computations to finish. This led to suboptimal trial intermixing and longer inter-trial intervals (ITI) when no more MOCS trials were available to bump up. In contrast, our tests showed that multithreading on the PC does not introduce excessive wait times. Additionally, we will implement a cap on how far ahead MOCS trials can be relative to AEPsych trials, limiting the difference to a maximum of 4 trials. Once this limit is reached, Sobol trials pre-generated with subject- and session-specific seeds (**Table 1**) will be inserted instead of MOCS trials. The pre-generated Sobol trials will be scaled by one of three factors (2/8, 3/8, 4/8), with an equal

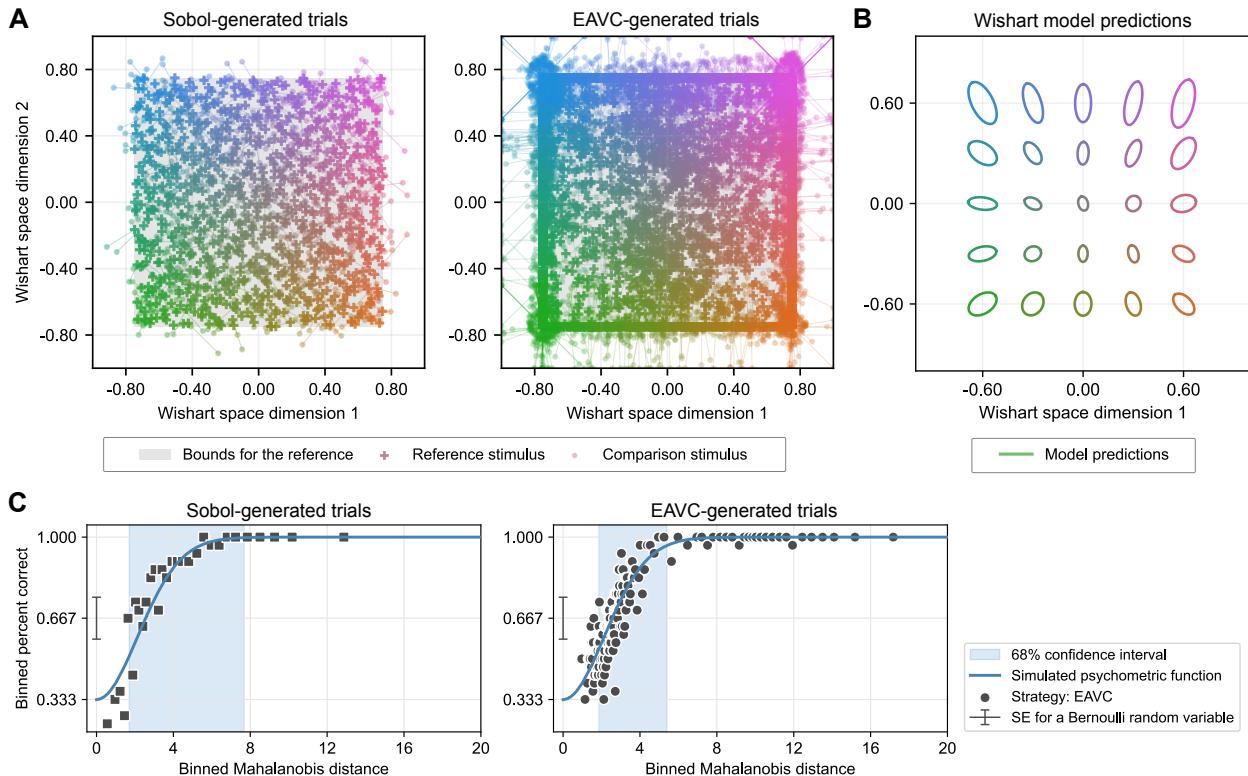


Figure 1. Trial placement guided by AEPsych and Wishart model predictions. (A) Sobol-generated trials ($N = 900$) and EAVC-generated trials ($N = 5,100$). The reference stimulus is sampled within the range $[-0.75, 0.75]$, while the delta values are constrained within $[-0.25, 0.25]$, ensuring that both reference and comparison stimuli remain within the bounds of $[-1, 1]$. (B) Wishart model predictions. Note that the axis scale in B is different from A.

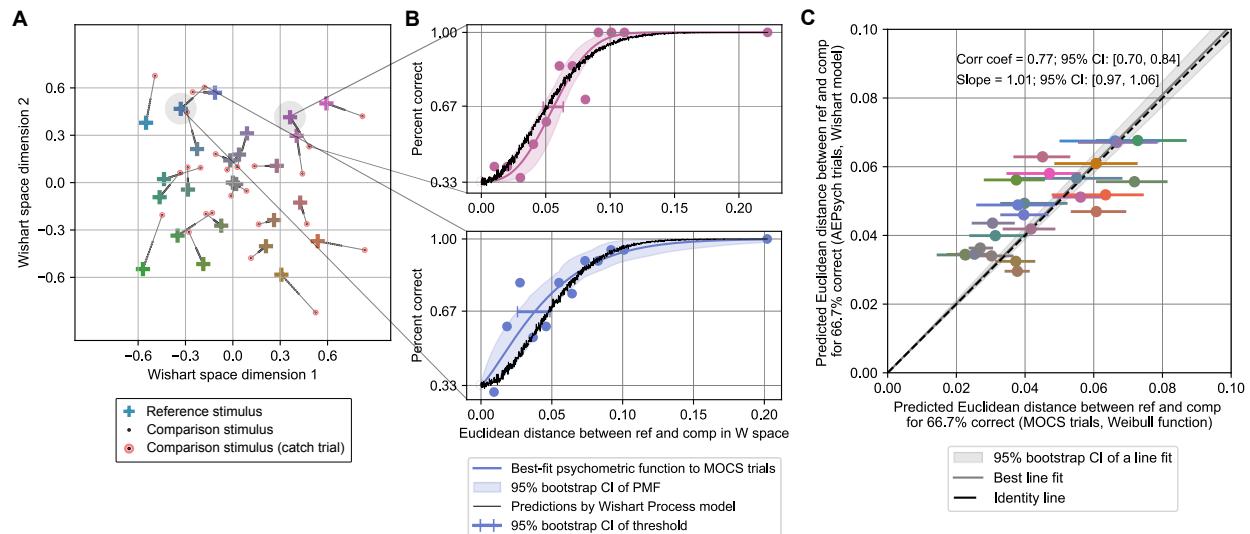


Figure 2. Selection of MOCS conditions and comparison of thresholds predicted by the Wishart model and Weibull psychometric functions. (A) Distribution of 25 MOCS conditions, sampled by a Sobol-generated reference stimulus and a corresponding chromatic direction along which comparison stimuli are varied. One of the reference stimuli is forced to be at the achromatic color. (B) Best-fit psychometric functions for two selected MOCS conditions. Horizontal axis: the Euclidean distance between the reference and its corresponding comparison stimuli in the Wishart space; vertical axis: the average percent correct after combining 20 trials per level. Top panel: An example where the threshold predicted by the Wishart model closely matches the threshold estimated from the Weibull psychometric function fit to MOCS trial data. Bottom panel: A case where the two threshold estimates diverge, with the 95% confidence interval of the Weibull-predicted threshold falling outside the Wishart-predicted threshold. (C) Threshold comparison across all 25 MOCS conditions. The solid line represents a linear fit to the data.

	Source	Practice	Experiment
Pre-generated MOCS trials	Python	sub# x 10,000	sub# x 100
Pre-generated Sobol trials	Python	N/A	sub# x 100 + session#
Sobol scalers	Python	sub# x 10,000	sub# x 100
Interleaving AEPsych and MOCS trials	Python	sub# x 10,000 + session#	sub# x 100 + session#
Location of the odd stimulus	C#	sub# x 10,000 + session#	sub# x 100 + session#

Table 1. Scheme for selecting seeds for different subjects from random-number generators. Five aspects of the experiment require shuffling: (1) the presentation order of the MOCS trials, which consist of 25 conditions, 12 levels, and 20 repetitions per level; (2) the sequence of pre-generated Sobol trials, which are inserted when AEPsych computations take too long and MOCS trials can no longer be used due to reaching the cap or running out, (3) the order of Sobol scalers (1/4, 2/4, 3/4) applied to Sobol-generated trials to balance task difficulty; (4) the shuffling sequence between AEPsych and MOCS trials; and (5) the location of the odd stimulus, which can appear at the top, bottom left, or bottom right. The first two aspects are determined only once during the first session, after which subsequent sessions reference the pre-determined indices. In contrast, the shuffling order between AEPsych and MOCS trials, as well as the odd stimulus location, are generated separately for each session.

distribution across these scalers. MOCS trials will resume once AEPsych trials have fully caught up.

Subjects will complete 12 sessions, with 1,000 trials per session. They will take a break after every 100 AEPsych trials, resulting in 4 breaks per session. For analysis, we will combine the pre-generated Sobol trials, which may vary in number slightly across subjects, with the 6,000 AEPsych-based trials and jointly fit the Wishart model to this dataset.

The analysis of MOCS trial data will be postponed until we finalize our Wishart model fitting approach, including potential adjustments to the prior over the Wishart process. That is, we will not analyze the MOCS validation data until we have more fully examined the data from Subject 1. We may also decide to analyze the Wishart model fits from additional subjects before finalizing the Wishardt fitting procedure. We will post an additional document describing the Wishart fitting and comparison with MOCS that we will perform. It is also possible we will decide to collect additional data from some subjects before examining the MOCS data, if we are not satisfied with the convergence of the Wishart model fits. This would include collecting additional interleaved MOCS data. We will also post an additional document if we make this decision, before collecting any additional data.

Until we post the document describing how we will compare with the MOCS data, the MOCS data will not be examined, so that our comparisons to that data are made in a planned rather than post-hoc manner.