**Study Information**

1. **Title**

Feature-Binding Errors in Associated Objects (Expt 3)

1. **Authorship**

Paul Scotti, Yoolim Hong, Andrew Leber, & Julie Golomb

1. **Research Questions**

How are features bound to object identities given capacity constraints in visual working memory?

1. **Hypotheses**

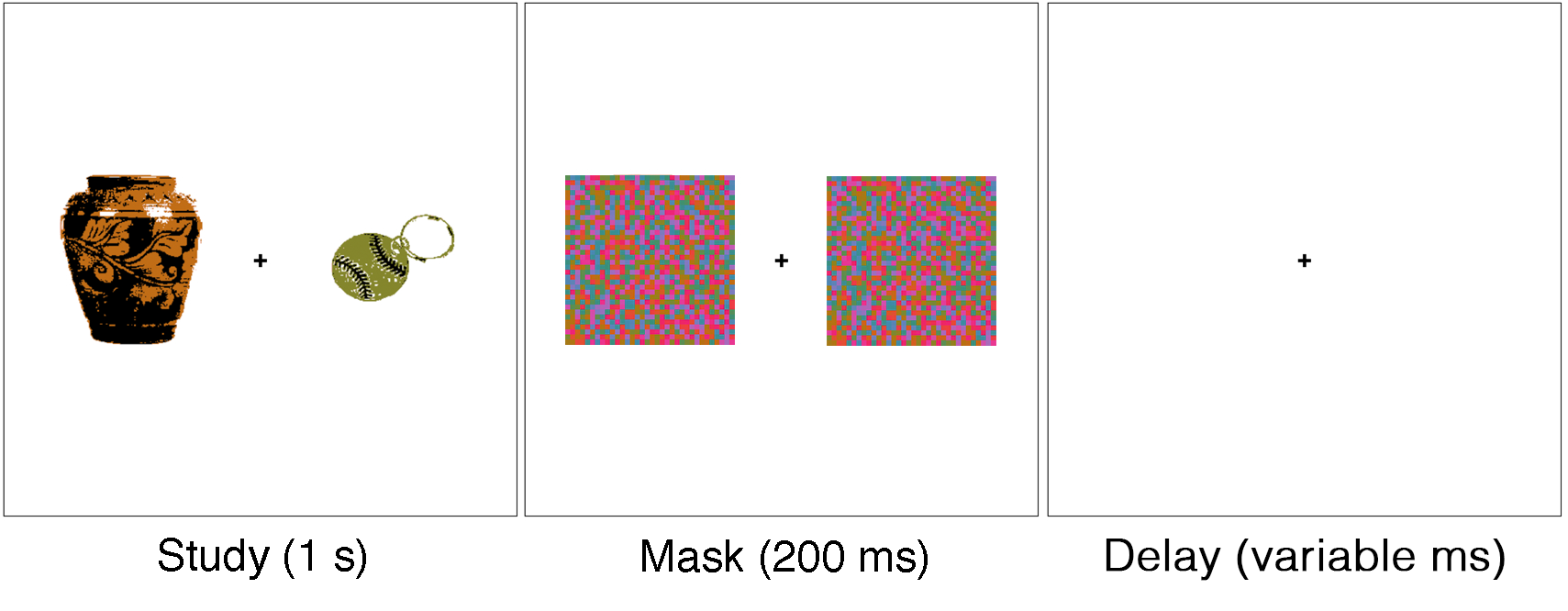
In Experiment 3, we aim to merge Experiment 1 (1 sec delay) and Experiment 2 (3 sec delay) into a within-subject design for increased statistical power. We hypothesize that trials with a 3-second maintenance delay will demonstrate larger shift and/or swap errors than trials with a 1-second maintenance delay.

**Design Plan**

1. **Study Type**
   1. Psychophysics experiment
2. **Blinding**
   1. All participants that pass exclusionary criteria will be included.
   2. MTurk: Participants’ worker IDs will be discarded prior to analysis to prohibit the ability to trace an individual to their Amazon Mechanical Turk account.

**Design**

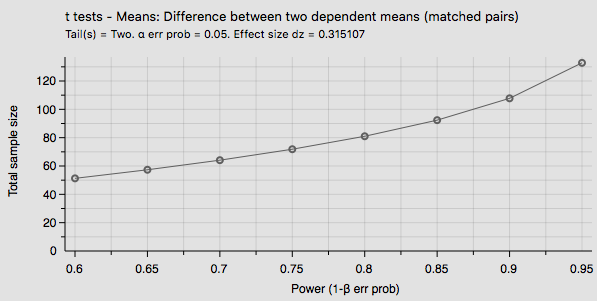
1. Below is an example trial procedure. There will be 280 trials total, split into 14 blocks of 20 trials each. Every trial, two unique real-world objects with a single associated color (360 deg. circle in CIE L\*a\*b\* space with coordinates L\*=70, a\*=20, b\*=38, radius 60; Zhang & Luck, 2008) will be displayed. Unknown to participants, the color distance between objects will be either 45 deg. or 90 deg., and the maintenance delay will be either 1 second or 3 seconds. Each of these four critical conditions (1sec/45deg, 1sec/90deg, 3sec/45deg, 3sec/90deg) will occur for 70 trials each, and their presentation order will be randomized within each subject. Participants will be probed with a grayscale representation of one of the two objects, and have to select the original color of the object and specify a confidence range (the smallest range of colors they believe contains the correct color; see Chen, Leber, & Golomb, under revision). As the mouse moves around the color wheel, the initially grayscale object will dynamically change to the color closest the mouse pointer. Feedback and bonus are then presented.
2. The bonus is composed of two parts, with each part awarding a max of 1 cent. The first part is degrees of error (distance from report to the original color). If x is absolute degrees of error, then cents awarded equals *(1 – x/45)*, such that more fractions of a penny are awarded for less deg. of error but nothing is awarded if x>=45. The 2nd part is based on the confidence report, specifically, if y is the confidence range (360 being a highlight of the entire color wheel), then cents awarded equals *(360 - y)/359*, such that smaller intervals award more money (359 because the smallest highlight possible encompasses 1 degree). Except, this only occurs if the highlighted region contains the true original color, if it doesn't, then no bonus is awarded for this part. A negative bonus is never awarded.





**Sampling Plan**

1. **Existing Data**
   1. As of the time of submission of this research plan for preregistration, no data have been created, collected, or analyzed.
2. **Data Collection Procedures.**
   1. *Timeline.* The study will take place during the Spring 2019, with the aim to finish the proposed experiment by Summer 2019.
   2. *Participants.* Subjects will be recruited through Amazon Mechanical Turk and will receive $6.00 an hour as compensation, plus a bonus of up to $5.60 depending on performance.
3. **Sample Size**
   1. Expt. 3: Amazon Mechanical Turk: 81 participants.
4. **Sample Size Rationale**
   1. The overall idea behind our power analysis was to estimate the necessary sample size to conduct within-subject statistics between maintenance delays, as opposed to the between-subject statistics available for experiments 1 and 2. To do this, we estimated within-subject variability for the decreased number of trials in our critical condition (i.e., mu parameter differs dependent on delay) and incorporated the expected change in effect size by taking the difference in averaged mu parameter estimates for experiments 1 and 2.
   2. For Expt 1 (1 sec delay, 50 subjects), the average mu parameter estimate across subjects was -0.35, with a standard deviation of 5.29. For Expt 2 (3 sec delay, 50 subjects), the average mu parameter estimate was -1.82, with a standard deviation of 2.72. We took the experiment with the larger standard deviation, Expt 1, and ran every subject through both a model that only included their odd-numbered trials and a model that only included their even-numbered trials. This yielded 50 even-models and 50 odd-models. We removed the mean difference in the two groups (group 1: odd models, group 2: even models) so that each group had the same average mu parameter estimates across subjects. We then subtracted 1.47 from the mu parameter estimate for each of the 50 subjects for the odd models (1.47 was obtained by taking the difference between -1.82 and -0.35). G\*Power was then utilized for a two-tailed paired t-test, suggesting a sample size of 81 subjects for a power of .80 (see figure below).
   3. Note that this power analysis was fairly conservative because we chose to estimate within-subject variability based on the experiment with larger variance in mu parameter estimates across subjects.



**Exclusion Criteria**

1. **Subject exclusions**
   1. Exclusion will depend on individual subject fits to a model that includes all memory reports. We plan to exclude subjects if their proportion of random guessing exceeds .50. These exclusion criteria are based off the exclusion criteria in Golomb et al. (2014) and Golomb (2015).

**Analysis Plan**

1. **Probabilistic Mixture Modeling**
   1. Using probabilistic mixture modeling, memory responses will be characterized according to one of four underlying distributions: a target distribution (responses around the correct color), a swap distribution (*S*: responses around the distractor object’s color), a swap-comparison distribution (*C*: responses around the color in the opposite direction to the distractor object’s color), and a random guessing distribution (uniform responding across all colors). Errors made towards the distractor object’s color will be assigned a positive sign, and errors made away from the distractor object’s color will be assigned a negative sign. In this way, we may observe a mean shift in the target distribution wherein responses are either towards or away from the distractor object’s color. We can therefore characterize shift errors (attraction or repulsion from the distractor color; 0) as well as swap errors (misremembering the original color as the distractor color; *S* > *C*).
   2. The probability distribution can be expressed as

where is the difference between the reported and correct color values, is the proportion of trials on which the participant responded at random, is a von Mises distribution with mean , , or , and concentration or (standard deviation ), is the proportion of trials on which the participant responded around the distractor object’s color (von Mises distribution with mean , the distance from the original color to the distractor color, and concentration ), and comparably is the proportion of trials on which the participant responded around the color in the opposite direction to the distractor object’s color.

* 1. For each participant, we will separately model responses for trials containing objects that were 45 deg. apart in color space, and trials containing objects that were 90 deg. apart in color space. We will also model responses for all trials regardless of distance in color space. We will use standard t-tests and ANOVAs to compare maximum a posteriori estimates between conditions.
  2. In addition to the above model fits, we will separately model memory errors across subjects after splitting each subject’s data into their most and least confident memory responses, comparing model parameters between confidence levels. For example, for the 45-degrees model, we will measure each subject’s median confidence range across trials that had objects 45 deg. apart in color space. Each subject’s trials will be divided into more and less confident halves, and a model will be separately fit to each half.
  3. We may, in addition, use a Bayesian hierarchical model to fit memory responses. We can then attain both group-level and subject-level parameter estimates and use 95% highest posterior density intervals (HDIs) to determine significance (Kruschke, 2011).

1. **Follow-up Analyses**

We may conduct supplementary analyses not listed in this document, for instance, to explore subjects with poor fits, the importance of color category, and alternative explorations into how confidence reports relate to different types of memory distortions. We will report any additional analyses as exploratory.