

# Task Free Recall

Computational Memory Laboratory  
University of Pennsylvania

January 13, 2010

## 1 Procedure

38 participants from the University of Pennsylvania community received payment in accordance with the University’s IRB guidelines. Stimuli were presented with a computer running PyEPL (Python Experiment Programming Library: <http://pyepl.sourceforge.net>, Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007). Verbal responses were recorded with a microphone and parsed with the pyParse package.

Participants ran in 4 sessions. Each session consisted of 12 free recall lists, a final free recall period, and a recognition period.

### 1.1 Stimuli

Words were chosen from a subset of the Toronto Noun Pool. All the words used have associated Word Association Space (WAS) scores. Words were excluded if they were ambiguous in meaning or were ill-defined with respect to the size and animacy judgments used in the experiment. For example, “METAL” was excluded since its size is ill-defined. The final word pool contained 1655 words.

### 1.2 Free Recall

On each trial, a list of 24 words was presented; each word was presented with a task cue above it, indicating the judgment that the participant should make for that word (“Size” or “Living/Nonliving”). Each word was presented

for 3 seconds. The two tasks were a size judgment (Will this item fit into a shoebox?) and an animacy judgment (Does this word refer to something living or non-living?). Participants made their response to each word using the index and middle fingers of their right hand to press one of four keys labeled “Big”, “Small”, “Living”, and “Nonliving”. If participants failed to respond during the time the item was onscreen, a beep was presented along with a message asking them to respond more quickly. If participants pressed one of the two keys that was inappropriate for the current task (e.g. “Living” if asked to make a size judgment), a beep was presented with a message asking them to press one of the appropriate keys. Half of each participant’s performance-based bonus pay (\$5 per session) was based on how many slow or inappropriate responses they made (the other half was based on the number of times they blinked during word presentations, as measured by EOG recordings).

The screen was blank for a  $1000 \pm 200$ ms inter-stimulus interval (ISI) between each word. ISI was randomly varied in order to de-correlate brain activity associated with each word presentation from activity during subsequent word presentations. Immediately following the list, a row of asterisks appeared, along with a beep, indicating the start of the recall period. Participants were given 90 seconds to recall as many words as they could remember from the most recent list, in any order.

There were two trial conditions, control and task shift. On control lists, every word was judged with the same encoding task. On the task-shift lists, participants shifted back and forth between the two tasks. Words were presented in groups, or “trains”, where a number of contiguous items were studied with the same encoding task. The length of each train was random, with the constraints that each train was between 2 and 6 items long (inclusive), and that the total number of items studied with each task was equal within the list. We counterbalanced (across lists) the task used to start the list and the number of trains in the list (6 or 7). This was done to balance whether lists started and ended with the same task, or started and ended with different tasks. The number of control and shift lists was balanced within each session. Lists were ordered in groups of four, where each group contained two control (one of each task) and two shift lists, and the order of lists was randomized within each group.

Using pilot data where participants were asked to rate each word using both encoding tasks, the words on a given list were chosen such that in total there would be a roughly equivalent number of items associated with

Ratio	# Lures	# Targets
0.125	41	288
0.250	96	288
0.375	158	264
0.500	192	192

Table 1: The four conditions for the recognition period. “Ratio” indicates number of lures divided by total number of stimuli.

each response (big, small, living, and nonliving). Many words are ambiguous with regard to the correct judgment (e.g., given the word dog, an image of a chihuahua might elicit a small judgment, while an image of a Great Dane might elicit a big judgment). Items for each list were chosen so that no two items were above a certain threshold of WAS similarity.

### 1.3 Final Free Recall

Once all the free recall trials were completed, participants were asked to recall words from the entire session, in any order. Participants were given six minutes to recall as many words as they could.

### 1.4 Recognition

Following the final free recall period, participants were presented with words that either had been presented during the session (target) or had not been presented during the session (lure). Complete lists from the session were randomly chosen for inclusion in the pool of targets. (NWM: for taskFR2, this was changed to choosing individual words at random. The choice to grab complete lists was deliberate, but no one seems to remember why we did it.)

The (Number of lures)/(Number of Targets and Lures) ratio was varied across sessions within each participant. The four conditions are reported in Table 1. The order of conditions was counterbalanced across participants according to a latin square.

Order of presentation of items was randomized. For each word that was presented, participants were asked to verbally indicate whether the item was a target. Participants were instructed to answer by saying either “pess” for

yes, or “po” for no. This was done to facilitate subsequent scoring of response onset times.

After responding yes or no, participants rated their confidence in their response on a scale from 1 (not at all sure) to 5 (completely confident). Participants were instructed to make their yes or no response as soon as they were sure which way they were leaning, then take extra time to decide their confidence if necessary. The first ten participants (LTP001-LTP010) made their confidence rating verbally; after speaking their response and confidence rating, they pressed a key to move on to the next word. Participants LTP011-LTP038 made their confidence rating by pressing 1-5 on a numeric keypad, which caused the program to advance to the next word.

The screen was blank for a  $1000 \pm 200$  ms inter-stimulus interval (ISI) between each word. ISI was randomly varied in order to de-correlate brain activity associated with each trial from activity during subsequent trials. Participants were given a chance to rest after each block of 20 items.

## 2 EEG Recordings

EEG measurements were recorded using commercially available equipment (Net Amps 200 amplifier, Net Station 4.1 acquisition software, Electrical Geodesics, Inc.). Recordings were made from 129 electrodes, and digitized at a sampling rate of 500 Hz. Recordings were originally referenced to Cz, and later converted to an average reference. Electrodes that had poor contact with the scalp were identified through manual inspection of raw EEG traces, and were excluded from the re-referencing.

In order to identify epochs contaminated with eye motion artifacts, recordings were made from electrooculogram (EOG) channels. Epochs were excluded from analyses if the weighted running average for either the left or right EOG pair exceeded a  $100 \mu\text{V}$  threshold.