# NS203: Experiments

[NS203: Experiments 1](#_Toc120008400)

[Experiment-1: Working Memory Task 1](#_Toc120008401)

[Task: 1](#_Toc120008402)

[**Methods:** 1](#_Toc120008403)

[Character Set: 1](#_Toc120008404)

[Trial Design: 1](#_Toc120008405)

[Experiment design: 2](#_Toc120008406)

[Biasing the participants. 2](#_Toc120008407)

[**Results and Answers to Questions** 2](#_Toc120008408)

[a. 2](#_Toc120008409)

[b. 3](#_Toc120008410)

[c. 3](#_Toc120008411)

[Experiment-2: To test the ESP in Humans 4](#_Toc120008412)

[Task: 4](#_Toc120008413)

[**Methods**: 4](#_Toc120008414)

[Stimuli: 4](#_Toc120008415)

[Task: 4](#_Toc120008416)

[Results: 4](#_Toc120008417)

All the Code with this experiment is attached as a zip file along with the report, and is also present in this

## Experiment-1: Working Memory Task

Task: To recall a set of characters from an unknown script

### **Methods:**

#### Character Set:

500 CJK (Chines, Japanese and Korean) Characters were considered starting from  **in Unicode.** They were then converted into simplified Chinese characters using [Hanzi Converter](https://github.com/berniey/hanziconv). During each trial, participants would receive a *random selection* of characters in this set.

Reason for selection of Chinese Characters:

This allowed lower set sizes *(less experiment time)* compared to more familiar objects like English letters, words or numbers. *Also, since single characters were being shown, it prevented participants from using hacks like remembering only a single digit from a 3-digit number during recall.*

#### Trial Design:

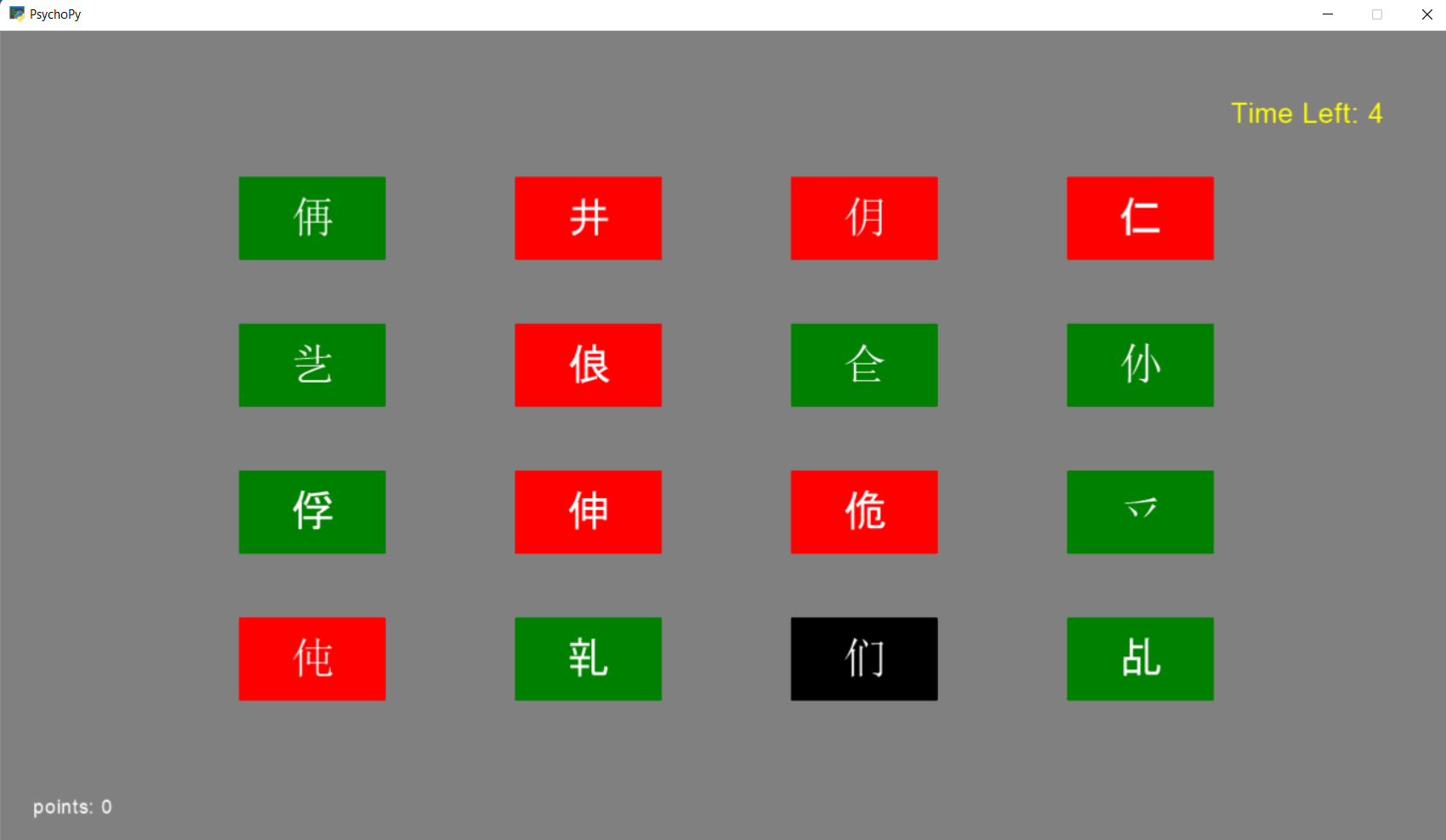
Each Trial consisted of two phases:

**1) Memorisation phase:** where a set number of characters were shown. Participants were given for memorisation. This scaling is done to ensure that effects of memory are being studied, without the effect of comprehension/reading time.



**2) An Interval** of between Memorisation and Recall.

**3) Recall phase:**   
Participants were shown a new set of characters. On most trials, around 50% characters would repeat from the memorisation phase *(this percentage was varied sometimes).* On most of trials, the characters were spatially shuffled up in the recall set.  
Here participants had to left click on stimuli that they had seen before, and right click on the unknown ones. They were given a total of **seconds**, where n is the number of items.



#### Experiment design:

Participants were made to perform the working memory task on set sizes of (3 trials each) after a brief training round.

After this, there were 2 additional trials to study the effect of shuffling the stimuli, and 6 trials to bias the subjects to create a ROC.

The number of stimuli for these was determined using the performance on the first block. We would want the participant to be more confused if we need to be able to bias them. Therefore, the lowest set size for which was chosen.

In addition, the participants were also given points as on each round to reflect their performance. This was to give players feedback as well help bias them in the later rounds.

#### Biasing the participants.

Part of the goal of the task was to vary the participants without varying . This was done as follows:

In Interval between memorisation and recall, the subjects would randomly be informed that:

* Most characters will repeat in the recall phase, so you can be liberal in your responses. The points will be biased to reward Hits much more

OR

* Very few characters will repeat in the recall phase, so please be stringent in your responses. The points will be biased to punish False alarms

In the first case the recall set would have repeated characters, while in the second case only about of the memorisation characters would make it to the recall phase.

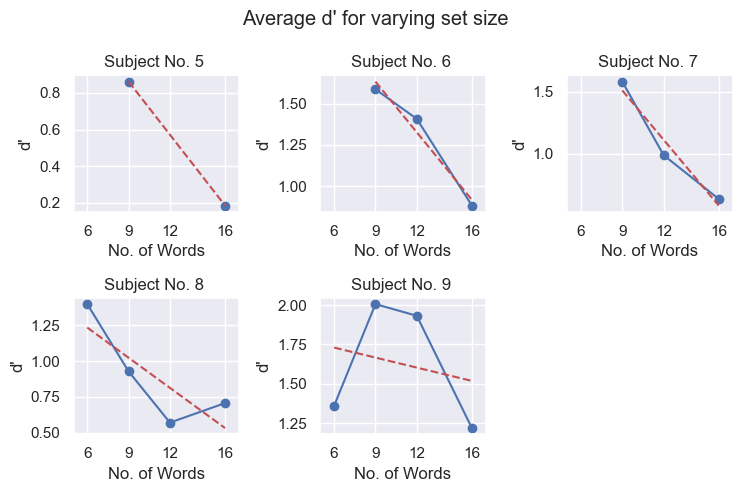
This should not change itself because the task still remains the same amount of difficult. But it will get people to vary their criterion. For example, in the second case, to get better scores, participants will be very “stringent” in their decision making. This would further be reinforced by the fact that most characters in the recall set are new.

### **Results and Answers to Questions**

*6 subjects participated in the test (in exchange for many kit-kats). One Result was filtered manually out because of the high rate of “unclassified characters” (none of hits, misses, false alarms or correct rejections)*

a. Does d’ vary with set size? Plot d’ vs. set size – this is called a psychophysical function

Across almost all participants, d’ falls with set size.



*(Some d’ values are infinite because of zero false-alarms or misses)*

For some subjects, like S-9, the subject seems to be trained during the test which gives the curve its appearance.

b. Does bias vary with set size? Interpret your findings

Bias doesn’t seem to have a clear trend with varying set size. This means that bias is not scaling appropriately with .

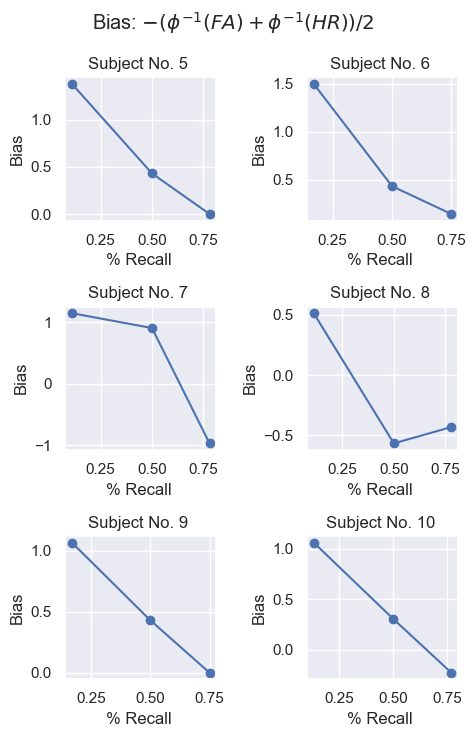
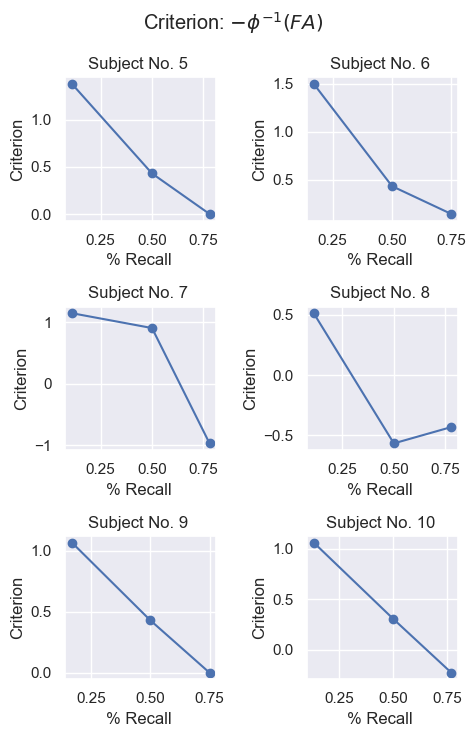
Although for most participants, positive bias seems to be increasing i.e. the criterion for classification of words as “seen” is increasing. One reason for this could be that once the participant is confused(higher set size), they prefer to only mark only the few words that they are confident of as “seen”. (i.e with increasing difficulty the participant is being more stringent in reporting something as “seen”.



c. Construct an ROC and show how it relates to d’.

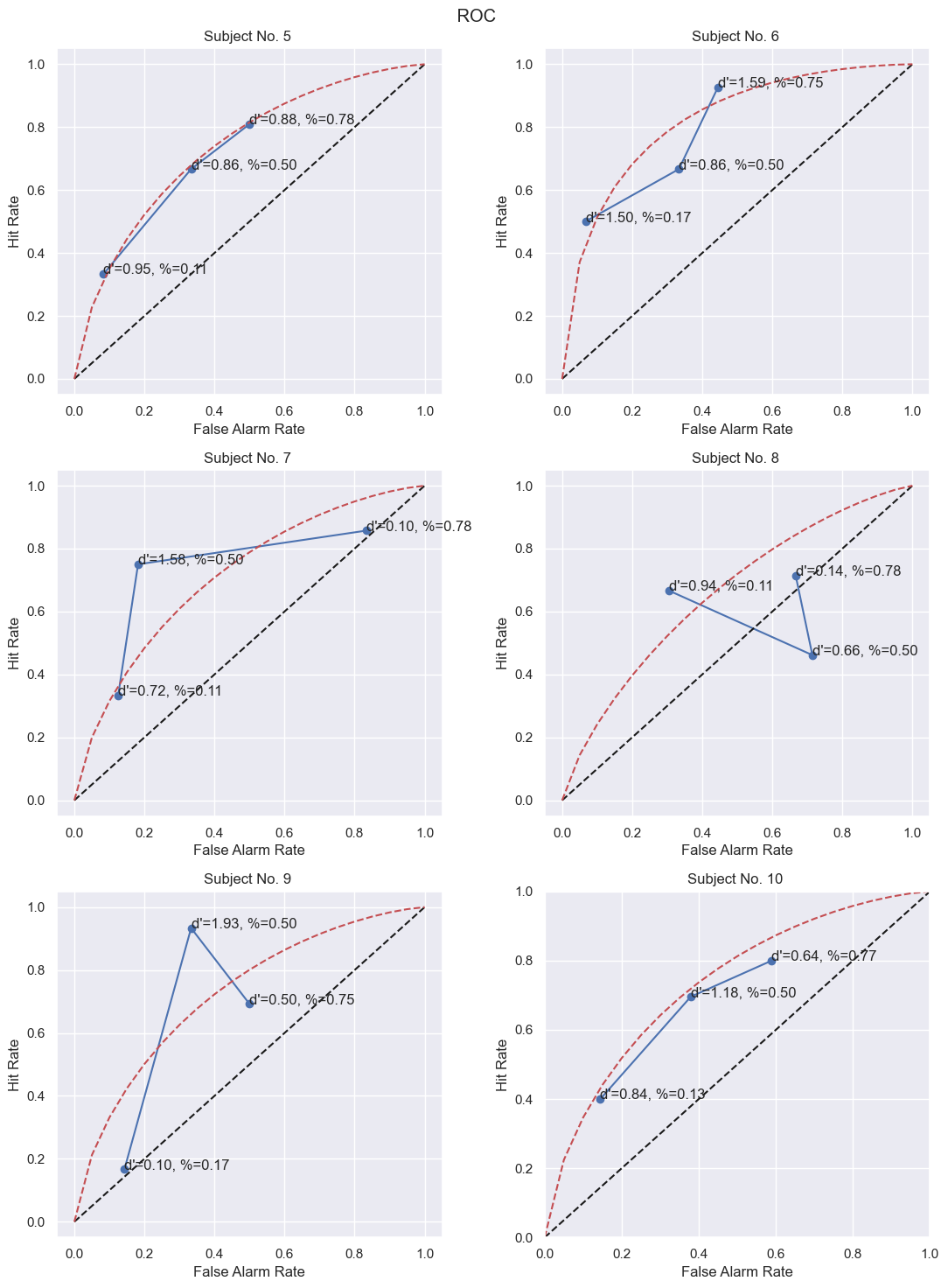
As mentioned in the methods section, was varied by varying the percentage of new vs repeated words and the rewards for Hits vs penalisation for False alarms.

First, let us make sure that and are being varied as we expected.



The bias and criterion fall with recall percentage. As described in [Biasing the participants.](#_Biasing_the_participants.), this is expected. *(when “seen” words are more common, and the reward for hits is increased, the participant decreased their criterion)*

Finally, the ROC is plotted along with a the ideal ROC ( based of the average d’)



We can see that some curves resemble ROC curves with constant .

But for **subjects other than subjects 5 and 10**, varies a bit. Therefore, I was not very in varying the criterion alone without varying . A better method would have probably been to ask the participants to rate their familiarity with the words on scale of 1-5. Then the criterion could be varied by changing the cut-off familiarity for a Hit and a Miss.

## **Experiment-2: To test the ESP in Humans**

Task: To detect characters written on a piece of paper hidden inside a book.

### **Methods**:

Stimuli:

The stimuli was a 40 character long sequence of heads and tails. There were 6 such sets of sequences. Each sequence was generated using a numpy’s random package

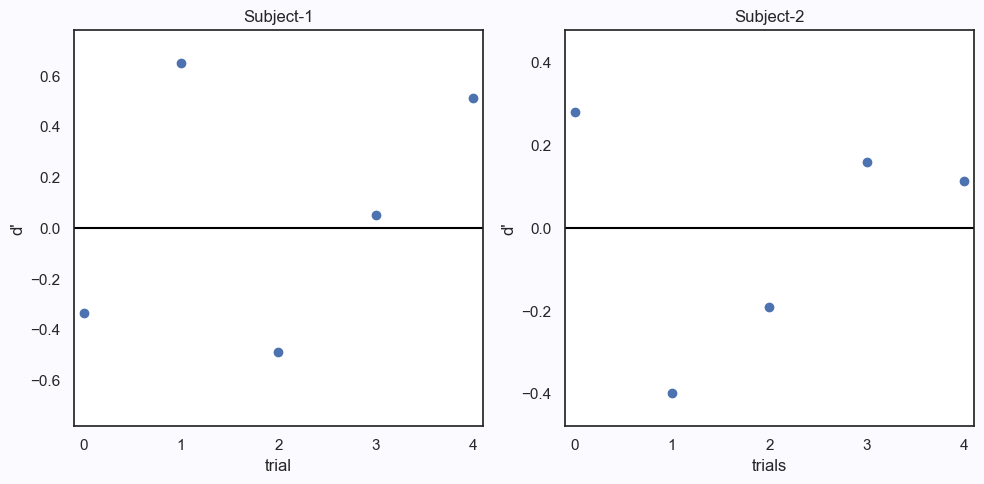
#### Task:

The sequence of stimuli was printed out, folded and placed in the middle of a very fat book, such that no part of the sheet was sticking out. The book was placed out of reach of the participant, and it was made sure that the participant had not seen the sheet of paper before.

The participant was asked to predict each of the sequences, by writing it on a piece of paper, and by reciting it verbally, when I would note it down. Both of the papers were collected and placed securely.

## **Results**:

Below are the values for each subject, for each trial.

As seen above, the changes drastically with each trial.

The average values are:  
subject- 1   
subject- 2

The averages are around 0 with high standard deviation. In fact, it changes signs. If the participants actually had ESP, the would have a constant sign.

Therefore, it’s reasonable to assume that both participants don’t have any ESP. ( Although scientifically one should use a rigorous statistical test to back this claim, I don’t see any obvious test that can be applied to this scenario)