**Supplemental Material for Identifying the Guilty Word: Simultaneous Versus Sequential Lineups for DRM Word Lists**

**Jason R. Finley, John T. Wixted, Henry L. Roediger, III**

Contents:

* FOLDER: data: contains Excel files with data from both experiments, separate files for simultaneous and sequential conditions
* FOLDER: data\_for\_R: contains plain text files formatted for use with the pROC package in R
* FOLDER: materials: contains word lists used in both experiments
* FOLDER: R\_syntax: contains R syntax used for ROC analyses and plots in both experiments. Note when using read.table you will have to replace “FILEPATH” with the actual path to the file on your computer.
* Details on constructing group ROCs
* Expanded explanation of determining cutoffs for pAUC comparison
* Details on excluded data from Experiment 1
* Instructions used in both experiments

**Details on constructing group ROCs**

Here we will describe our process of analysis in more detail than usual, in the hopes that it may help other researchers who similarly find themselves at the perplexing intersection of laboratory recognition methods, eyewitness lineup methods, and signal detection analysis. Our original intent was to construct ROC curves for individual subjects, separately for TP:TA (no critical lures) and TPCL:TACL (critical lures), using the method described by Mickes et al. (2012, pp. 366-367) for lineup research, which we note differs from the ROC construction method traditionally used in basic laboratory recognition research (e.g., Macmillan & Creelman, 2004, pp. 53-57). We would then use familiar statistics (*t*-test, ANOVA) to compare mean partial area under the curve (pAUC) from subjects in the simultaneous condition versus subjects in the sequential condition. However this proved problematic with only eight data points for hit rate and eight for false alarm rate, yielding jagged and noisy individual ROCs. Thus we instead combined data from all subjects to construct group ROCs, using the pROC statistical package for R (Robin et al., 2011), and with help from the tutorial provided in the supplemental materials of Gronlund, Wixted, and Mickes (2014).

In fact, we used an alternative data format to submit our data to the pROC package in R, different from the format illustrated in the tutorial by Gronlund et al. (2014). See the included folder “data\_for\_R”. We formatted our data such that each row represented one subject’s encounter with one lineup. For columns, in addition to columns for subject number, test format, lineup type, and critical lure presence, we also made a column for “TargetPresence” to indicate if the lineup contained a target and coded as 0 or 1 for each row, and another column for “ConfidenceHFA” in which the subject’s confidence was listed *only* when their response was a hit (defined as choosing the target in a TP or TPCL lineup) or a false alarm (defined as choosing any lure in a TA or TACL lineup). For all other response types (false alarms in TP or TPCL lineups, misses, and correct rejections), “ConfidenceHFA” was set to zero. This was a data format compatible with the pROC packaged. Finally, we also made two more columns:

* “ConfidenceHFAcl” for which the confidence was listed only for hits or false alarms to critical lures (this was only used in constructing TP:TACL\_cl ROCs).
* “ConfidenceHFAnoncrit” for which the confidence was listed only for hits or false alarms to non-critical lures (this was only used in constructing TPCL:TACL ROCs).

**Expanded explanation of determining cutoffs for pAUC comparison**

In order to compare the partial area under the curve (pAUC) across test formats, it is necessary to first determine the range of false alarm rates to calculate across. For the left-hand cutoff we chose zero, because our empirical ROC curves were clearly headed toward the origin, and that is where theoretical ROC curves start as a rule. A subject with the most conservative decision criterion possible, who simply says “no” to everything, would indeed achieve a hit rate and false alarm rate of zero, since they are rejecting every lineup. Thus, using zero as the left cutoff makes sense. For calculation of the pAUC, the pROC package extended each empirical ROC curve down to the origin with a straight line (not pictured in Figure 2).

Choosing a right-hand cutoff is less straightforward. For simple detection tasks, such as laboratory old/new recognition tests, the right side of an ROC curve theoretically ends at HR = FAR = 100%. For calculation of the full AUC, the pROC package will extend each empirical ROC curve up to the top-right corner of the plot space with a straight line. However, this is problematic. Lineup identification can be conceived as a type of “detection and localization” task, because it matters not just that the eyewitness says “yes” to a target-present lineup, but also that they correctly say *which* member of the lineup is the target. This is analogous to a radiologist detecting that an abnormal shape is present in a radiograph, and also having to correctly say which area of the radiograph. Even a subject with the most liberal decision criterion possible, who is willing to say “yes” to everything, would not achieve a hit rate of 100% because they would still need to correctly guess which candidate is the target in a target-present lineup for their response to count as a hit. The liberal subject would, however, be able to achieve a false alarm rate of 100% because saying yes to any candidate in target-absent lineups would count as a false alarm. Thus, ROC curves in detection and localization tasks (such as lineup identification) are not expected to extend up to the top-right of the plot area, but rather are more likely to end at FAR = 100% and HR < 100%, which also means that the ROC curve will be at least partly below the diagonal. Examples of such empirical ROC curves from a radiological detection and localization task can be seen in Starr, Metz, Lusted, and Goodenough (1975, Figure 2, dotted lines).

In our data, the right end of our empirical ROCs end at far less than FAR = 100%, and there is no simple and accurate way to extend these ROC curves all the way to the right of the plot area. Functions could be fit to the data to extend the curve, but that seems hardly worth the trouble when our only goal is to statistically test whether one ROC curve is higher than the other. Fortunately, the maximum false alarm rate (i.e., right edge of the ROC curve) did not differ drastically between the two test formats, so choosing either of those as the right-hand cutoff for pAUC would only involve a small amount of truncation or extension. The pROC package, when asked to calculate pAUC beyond the point where the empirical data end, still uses the extension method of drawing a straight line toward the top-right of the plot area; but because this line segment covers only a very small span, it is not nearly as problematic as it would have been for full AUC. That still leaves the dilemma of choosing *which* of the two maximum FARs to use for the right-hand cutoff: the smaller (usually yielded by the simultaneous test format), or the larger (usually yielded by the sequential test format). Having no principled way to choose, we used both. That is, we performed comparison of pAUCs for simultaneous versus sequential test formats once using the smaller right-hand cutoff, and once using the larger right-hand cutoff. The pattern of results ended up being consistent across these two approaches.

We are not yet familiar with the following approach, but we note here that Smith et al. (2019) have proposed an alternative to comparing pAUCs aimed at avoiding the problem of differing maximum FARs, a measure called deviation from perfect performance (DPP).

Smith, A. M., Lampinen, J. M., Wells, G. L., Smalarz, L., & Mackovichova, S. (2019). Deviation from Perfect Performance measures the diagnostic utility of eyewitness lineups but partial Area Under the ROC Curve does not. *Journal of Applied Research in Memory and Cognition*, *8*(1), 50-59.

Starr, S. J., Metz, C. E., Lusted, L. B., & Goodenough, D. J. (1975). Visual detection and localization of radiographic images. *Radiology*, *116*(3), 533-538.

**Details on excluded data from Experiment 1**

The following data were excluded from analysis:

* Due to a programming error, target items for two particular DRM lists were not properly presented in TP and TPCL lineups. Data were excluded from analyses as follows:
  + data from the “DOCTOR” DRM list in both simultaneous and sequential conditions
  + data from the “GIRL” DRM list in sequential condition
* Simultaneous condition: any trials in which subject didn’t respond within 8 s
* Sequential condition: any lineups with any trials in which subject didn’t respond within 8 s

**Instructions Used in Both Experiments**

**Experiment 1: simultaneous**

Initial Instructions:

In this task, you will study and be tested on a total of 32 lists consisting of 10 words each. However, they will be divided into smaller sets to make the task easier. There will be four blocks of studying the lists, with a test after each block. That is, you will study 8 lists and then you will be tested on those eight lists. This same kind of cycle will be repeated three more times until you have been studied and tested on all the lists.

During the test, you will be shown a test list of words as in a multiple-choice test. Your task is to decide which word is a studied word. You should select only one word as being studied. However, unlike standard multiple choice tests, in some of these test lists there is no word that was actually from the list, so in this case you should select the "Not Present" option. After you make your decision, you will be asked to rate how confident you are in your response using a 0-100 scale.

Again, keep in mind that in some test sequences, none of the words shown will be words you studied. In such a case, please select the "Not Present" option. You will also be asked to rate how confident you are in this decision just as you are when you select a word.

Do you have any questions?

Test instructions:

You will now be tested on the last 8 lists you studied. You will be tested in the order you studied the lists, so that the first test sequence tests you on the first list you studied and so on.

You will be shown all the potential answer choices for each list simultaneously. Please select the word you studied from the choices. You will have 8 seconds to make a response.

Please be aware that in some test sequences, the words you studied will not be present. In such a case, please select "Not Present".

After your decision, you will rate how confident you are in your response using the 0-100 scale where 0 indicates “absolutely unsure” and 100 indicates “absolutely sure”.

Before each list test:

You are about to be tested on a list you studied. Please remember you are being tested in the order you studied the lists.

**Experiment 1: sequential**

Initial Instructions:

In this task, you will study and be tested on a total of 32 lists consisting of 10 words each. However, they will be divided into smaller sets to make the task easier. There will be four blocks of studying the lists, with a test after each block. That is, you will study 8 lists and then you will be tested on those eight lists. This same kind of cycle will be repeated three more times until you have been studied and tested on the lists.

During the test for each of the lists, you will be shown a series of words presented individually. For each word, you must decide whether or not you studied that word or not. If it was on one of the lists you studied, you should respond, "YES". If it was not on any list, you should respond, "NO". After each decision you make, you will be asked to rate how confident you are in your response using a 0-100 scale. You will slide a cursor from 0 to 100 to indicate how sure you are that the word was studied (YES) or was not studied (NO).

Please be aware that in some test sequences, none of the words shown will be words you studied. In such a case, you will (ideally) have answered that you did not study any of the words.

Do you have any questions?

Test instructions:

You will now be tested on the lists you studied. You will be tested in the order you studied the lists, so that the first test sequence tests you on the first list you studied and so on.

Each word will be presented individually and you must make a decision on whether you studied each word or not. You will have 8 seconds to respond to each word.

If you did study the word, select, "Yes".

If you did not study the word, select, "No".

Please be aware that in some test sequences, none of the words you studied will be present. In such a case, you will have responded, "No", to all words. That is okay, because in some lists no word from the study list is present.

After each decision you make, you will rate how confident you are in your response using the 0-100 scale where 0 indicates "absolutely unsure" and 100 indicates "absolutely sure".

Before each list test:

You will now be tested on the lists you studied. Please remember you will be tested in the order you studied the lists.

You will be shown each word one at a time. You must decide whether each word was a word you studied earlier. Your first "Yes" response for each list will be the only one counted.

**Experiment 2: simultaneous**

Initial Instructions:

In this task, you will study and be tested on a total of 32 lists consisting of 10 words each. However, they will be divided into smaller sets to make the task easier. There will be four blocks of studying the lists, with a test after each block. That is, you will study 8 lists and then you will be tested on those eight lists. This same kind of cycle will be repeated three more times until you have been studied and tested on all the lists.

During the test, you will be shown a test list of words as in a multiple-choice test. Your task is to decide which word is a studied word. You should select only one word as being studied. However, unlike standard multiple choice tests, in some of these test lists there is no word that was actually from the list, so in this case you should select the "Not Present" option. After you make your decision, you will be asked to rate how confident you are in your response using a 0-100 scale.

Again, keep in mind that in some test sequences, none of the words shown will be words you studied. In such a case, please select the "Not Present" option. You will also be asked to rate how confident you are in this decision just as you are when you select a word.

Do you have any questions?

Test instructions:

You will now be tested on the last 8 lists you studied. You will be tested in the order you studied the lists, so that the first test sequence tests you on the first list you studied and so on.

You will be shown all the potential answer choices for each list simultaneously. Please select the word you studied from the choices. You will have 8 seconds to make a response.

Please be aware that in some test sequences, the words you studied will not be present. In such a case, please select "Not Present".

After your decision, you will rate how confident you are in your response using the 0-100 scale where 0 indicates “absolutely unsure” and 100 indicates “absolutely sure”.

Before each list test:

You are about to be tested on a list you studied. Please remember you are being tested in the order you studied the lists.

**Experiment 2: sequential**

Initial Instructions:

In this task, you will study and be tested on a total of 32 lists consisting of 10 words each. However, they will be divided into smaller sets to make the task easier. There will be four blocks of studying the lists, with a test after each block. That is, you will study 8 lists and then you will be tested on those eight lists. This same kind of cycle will be repeated three more times until you have been studied and tested on the lists.

During the test for each of the lists, you will be shown a series of words presented individually. For each word, you must decide whether or not you studied that word or not. If it was on one of the lists you studied, you should respond, "YES". If it was not on any list, you should respond, "NO". After each decision you make, you will be asked to rate how confident you are in your response using a 0-100 scale. You will slide a cursor from 0 to 100 to indicate how sure you are that the word was studied (YES) or was not studied (NO).

Please be aware that in some test sequences, none of the words shown will be words you studied. In such a case, you will (ideally) have answered that you did not study any of the words.

Do you have any questions?

Test instructions:

You will now be tested on the lists you studied. You will be tested in the order you studied the lists, so that the first test sequence tests you on the first list you studied and so on.

Each word will be presented individually and you must make a decision on whether you studied each word or not.

If you did study the word, select, "Yes".

If you did not study the word, select, "No".

Please be aware that in some test sequences, none of the words you studied will be present. In such a case, you will have responded, "No", to all words. That is okay, because in some lists no word from the study list is present.

After each decision you make, you will rate how confident you are in your response using the 0-100 scale where 0 indicates "absolutely unsure" and 100 indicates "absolutely sure".

Before each list test:

You will now be tested on the lists you studied. Please remember you will be tested in the order you studied the lists.

You will be shown each word one at a time. You must decide whether each word was a word you studied earlier. Your first "Yes" response for each list will be the only one counted.