Behavior Research Methods
https://doi.org/10.3758/s13428-022-01807-6

fullROC: An R package for generating and analyzing eyewitness-lineup ROC curves

Yueran Yang[1] · **Andrew M. Smith**[2]

Accepted: 27 January 2022

Abstract

A police lineup is a procedure in which a suspect is surrounded by known-innocent persons (fillers) and presented to the witness for an identification attempt. The purpose of a lineup is to test the investigator's hypothesis that the suspect is the culprit, and the investigator uses the witness' identification decision and the associated confidence level to inform this hypothesis. Whereas suspect identifications provide evidence of guilt, filler identifications and rejections provide evidence of innocence. Despite the capacity of lineups to provide exculpatory information, past research has focused, almost exclusively, on inculpatory behaviors. We recently developed a method for incorporating all lineup outcomes in a single receiver operator characteristic (ROC) curve. The area under the full lineup ROC curve reflects the total capacity of a lineup procedure to discriminate guilty suspects from innocent suspects. Here, we introduce a Comprehensive R Archive Network (CRAN) package, *fullROC*, to support eyewitness researchers in using the full ROC approach to analyze lineup data. The *fullROC* package provides functions for adjusting identification rates, generating full ROC curves for lineup data, computing the area under the ROC curves (AUC), and statistically comparing the AUCs of different lineups. Using both simulated and empirical data, we illustrate the functionality of the *fullROC* CRAN package. In brief, the *fullROC* package provides a useful tool for eyewitness researchers to analyze lineup data using the full ROC method, which incorporates both the inculpatory and exculpatory information of eyewitness behaviors.

Keywords Eyewitness identification · ROC curves · Investigator discriminability · R package

When police suspect that someone committed a crime, they often construct a lineup and present it to a witness for an identification attempt. A lineup is a procedure in which police surround the person that they suspect committed the crime (i.e., the suspect) with some number of known-innocent persons called "fillers" (usually five fillers). The rationale for surrounding the suspect with fillers is that the suspect might be innocent and if so, fillers offer some protection from mistaken identifications as many witnesses will identify a known-innocent filler rather than the innocent suspect. While a filler identification is still a mistaken identification on the part of the witness, it does not carry the same forensic implications of an innocent-suspect identification. Indeed, when a witness identifies a filler, the investigator knows the witness has made an error and the filler is not at risk of arrest and conviction. Conversely, when a witness identifies an innocent suspect, it provides support for the investigator's erroneous hypothesis that the suspect is guilty and puts the innocent suspect at risk of arrest and conviction. Because of their protective features, best-practice recommendations encourage investigators to use lineups rather than presenting a lone suspect to a witness for an identification attempt (Wells et al., 2020).

Implicit in the above description is a lineup's purpose. Contrary to intuition, the purpose of a lineup is not to test a witness' memory. Rather, the purpose of a lineup is to test an investigator's hypothesis that the suspect is the culprit (Smith et al., 2020; Wells & Luus, 1990). Because the investigator does not know the suspect's guilt status, she uses a lineup to obtain evidence—eyewitness outcomes—to test her hypothesis that the suspect is guilty. Whereas some eyewitness outcomes support the hypothesis that the suspect is guilty (e.g., a suspect identification), some support the

---

✉ Yueran Yang
yuerany@unr.edu

[1] Department of Psychology, University of Nevada, Reno, 1664 N. Virginia St, Reno, NV 89557, USA

[2] Department of Psychology, Iowa State University, Ames, IA, USA

alternative hypothesis that the suspect is innocent (e.g., a rejection). Hence, investigators can use lineups both to "rule in" or inculpate guilty suspects and to "rule out" or exculpate innocent suspects. The capacity of an investigator to discriminate guilty suspects from innocent suspects—or the total informational value of a lineup—depends on a lineup's capacity both to "rule in" the guilty and to "rule out" the innocent (Smith et al., 2020; Smith & Ayala, 2021; Starns et al., 2021; Wells et al., 2015; Wells & Lindsay, 1980).

In a recently published paper, we introduced a novel means for measuring the total informational value of a lineup procedure (or what we have called investigator discriminability) with receiver operating characteristic (ROC) curves that include all lineup outcomes: suspect identifications, filler identifications, and rejections (Smith et al., 2020). Our method can be viewed as a way to extend the "partial" suspect-only ROC curves. Whereas the "partial" ROC curves only measure a lineup's capacity to "rule in" the guilty, the full ROC curves which incorporate all lineup outcomes measure a lineup's total informational value. In the present work, we introduce a novel R package, *fullROC*, for comparing the total informational value (or investigator discriminability) of two lineup procedures. We accompany the introduction of *fullROC* with a tutorial demonstrating various features and functions of this package. We first elaborate how full ROC curves measure lineup procedures' capacities to both "rule in" and "rule out" suspects. We then use both simulated and empirical data to show readers step by step how to create full ROC curves for lineup data.

Signal detection and ROC analysis

According to signal detection theory, two factors influence how a diagnostic system makes a classification decision: discriminability, which reflects the system's capacity to distinguish signal from noise; and response criterion, which reflects the amount of evidence the system requires to make an affirmative response (Green & Swets, 1966; Swets, 1988). Scientists typically place a premium on maximizing discriminability. This is because, for any given false-positive rate, a system with better discriminability will result in a higher true-positive rate than will a system with worse discriminability. Conversely, variations in response criteria will result in mere trade-offs, increasing (or decreasing) both the true-positive rate and the false-positive rate without influencing discriminability.

In practice, it is not always easy to determine which diagnostic system leads to better discriminability. It is often the case that one system leads to both a higher true-positive rate and a higher false-positive rate than does a comparison system (e.g., Clark, 2012). ROC curves were specifically designed to deal with these trade-offs. ROC curves measure discriminability independent of response criterion. This is accomplished by plotting the true-positive rates against the false-positive rates for all potential response criteria. Connecting these pairs of true- and false-positive rates (also called operation points) creates the diagnostic system's ROC curve. Because the ROC curve runs through each operating point, it places equal weight on the operating points. Thus, the area under the ROC curve (AUC) provides a measure of discriminability independent of response criterion (Macmillan & Creelman, 2005).

ROC curves have become a popular method for eyewitness researchers to handle the trade-offs that are commonly involved in lineup comparisons (Clark, 2012; Wixted & Mickes, 2012). Whereas some researchers have separately considered the incriminating properties of suspect identifications and the exculpating properties of rejections and filler picks (e.g., Wells & Lindsay, 1980; Wells & Olson, 2002; Wells & Turtle, 1986), the field has focused mainly on suspect identifications and the capacity of a lineup to incriminate. This has typically been examined with *"partial" suspect-only* ROC curves that only include suspect identifications and thus cover a fraction of the ROC space (e.g., Mickes et al., 2012; Wetmore et al., 2015). The problem with "partial" suspect-only ROC analysis, as with any analysis that focuses only on suspect identifications, is that it does not consider all information that is relevant to determining which of two procedures is superior.

The value of a lineup is not limited to its capacity to "rule in" or incriminate guilty persons. Lineups also have the capacity to "rule out" or exculpate the innocent. The total informational value of a lineup procedure is based both on that procedure's capacity to "rule in" the guilty and to "rule out" the innocent (e.g., Smith et al., 2020; Smith & Ayala, 2021; Starns et al., 2021). Just as being able to demonstrate the absence of disease is an important property of many diagnostic tests, being able to demonstrate innocence is an important property of a lineup. Failure to clear innocent persons from police suspicion can completely stymy police investigations, let alone the consequences uncleared innocent persons are forced to shoulder. We cannot speak sensibly about which of two lineups is superior unless we consider the total informational value of those procedures, which is based both on their capacities to "rule in" the guilty and to "rule out" the innocent. Critically, the lineup procedure that does a better at "ruling in" guilty persons is not necessarily the procedure that has greater total informational value. In fact, it is often the case that the procedure that does a better job "ruling in" does a worse job "ruling out" (Smith et al., 2020; Wells et al., 2015) and often times the procedure that is worse at "ruling in" is so much better at "ruling out" that it has greater total informational value (Smith & Ayala, 2021; Starns et al., 2021).

Below we illustrate how full ROC curves can summarize a lineup's capacity to both "rule in" guilty suspects and to "rule out" innocent suspects. We first propose a model that describes how investigators use eyewitness evidence to classify a suspect's guilt. We then show the connection between investigators' signal detection model and full ROC curves and discuss how full ROC curves can reflect the total informational value of a lineup procedure.

Investigator signal detection model

When conducting a lineup, investigators obtain eyewitness evidence to inform their hypothesis that the suspect is guilty (or the alternative hypothesis that the suspect is innocent). In other words, investigators rely on eyewitness evidence to decide whether to arrest the suspect or not (Smith et al., 2020). Combined with the two possible states of ground truth, investigators' responses conform to a 2 (culprit present vs. culprit absent) × 2 (arrest vs. not arrest) confusion matrix. Figure 1 shows a signal detection model of the investigators' detection task.

Note that the distributions in Fig. 1 are discrete rather than continuous. This is because the decision variable on which investigators' classification decisions rely—eyewitness evidence—is discrete rather than continuous (Yang & Moody, 2021). For example, Fig. 1 shows nine possible discrete eyewitness outcomes, combining both witness' responses (suspect identifications, filler identifications, or rejections) and associated confidence levels (high, medium, or low). Because eyewitness evidence can come from either culprit-present or culprit-absent lineups, the model contains two distributions, which represent the probability mass (i.e., response rate) of the eyewitness outcomes from either lineup. In Fig. 1, the *white* and *gray* bars present the probability mass distributions of eyewitness outcomes in culprit-absent and culprit-present lineups, respectively. The *vertical dashed lines* present the decision criteria investigators can use.

In completing the classification task, investigators use the witness' response and associated confidence to infer whether the suspect is guilty or not. As shown in Fig. 1, investigators have control over how much evidence is needed to make an arrest decision (or some other investigative decision). For example, some investigators may decide to arrest a suspect only when witnesses identify the suspect with high confidence (i.e., when obtaining evidence to the right of the decision criterion $c_{IDS_{high}}$ in Fig. 1) and not to arrest the suspect when witnesses make other responses (i.e., when obtaining any evidence to the left of $c_{IDS_{high}}$). But others may decide to arrest a suspect even after a low-confidence identification (i.e., when obtaining any evidence to the right of $c_{IDS_{low}}$ in Fig. 1) and not to arrest when witnesses identify a filler or reject the lineup (i.e., when obtaining any evidence to the left of $c_{IDS_{low}}$). Investigators could be even more liberal, deciding to arrest a suspect as long as witnesses do not reject the lineup with high confidence (i.e., when obtaining any evidence to the right of $c_{IDF_{high}}$ in Fig. 1). Therefore, different combinations of witness responses and associated confidence reflect different decision criteria investigators can use to make investigative decisions.

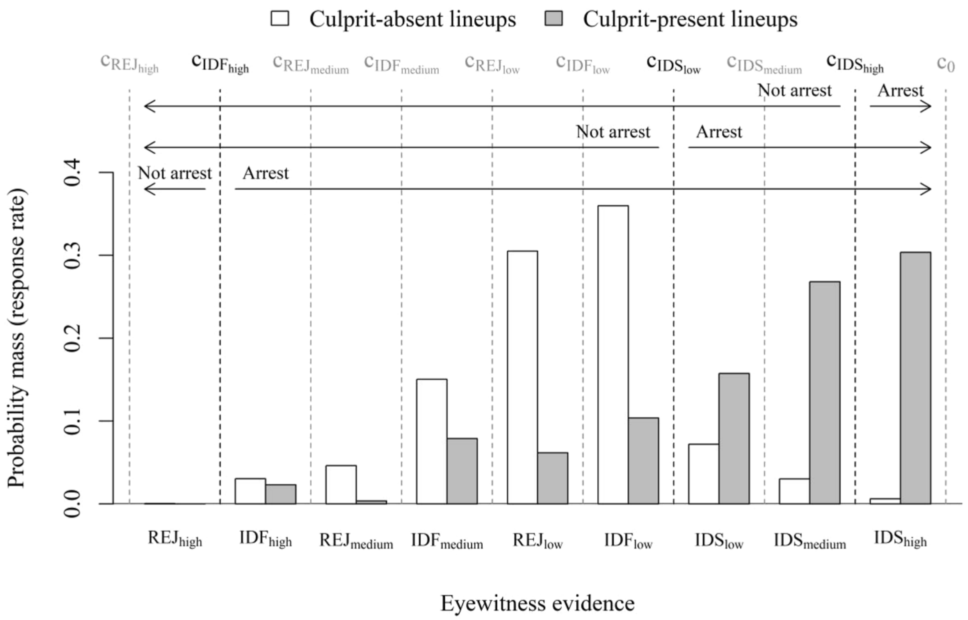


Fig. 1 Investigator signal detection model

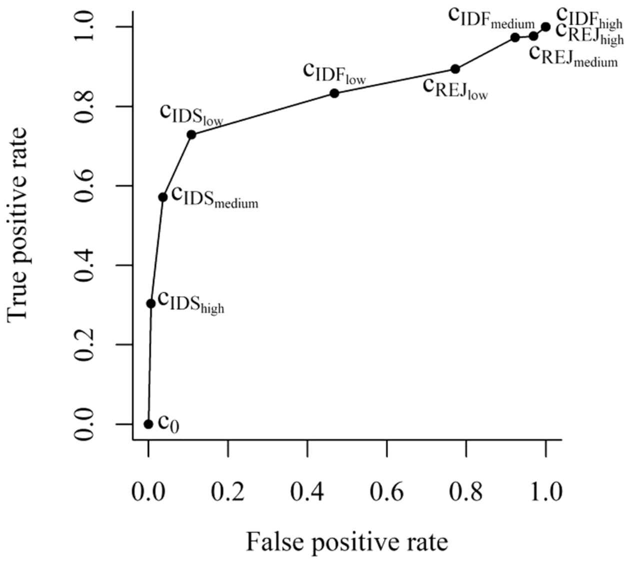


Fig. 2 Investigator ROC curves

Investigator ROC curves

Based on the investigator signal detection model, a full ROC curve that spans the entire ROC space can be produced to include all eyewitness outcomes (i.e., all possible investigator decision criteria). A full ROC curve is simply a collection of investigators' true- and false-positive rates at each level of the decision criteria, moving from the most conservative to the most liberal criterion (i.e., moving from right to left on the decision variable coordinate). Because of the discrete nature of investigators' decision variable (eyewitness evidence), investigator ROC curves are discrete.

Figure 2 displays the ROC curve from the investigator signal detection model in Fig. 1. The operating points on ROC curve present investigators' true- and false-positive rates when investigators make investigative decisions under different decision criteria. For example, the origin (0, 0) presents the situation in which investigators use the most conservative criterion, deciding not to arrest the suspect no matter what responses witnesses make (i.e., using the criterion c_0 in Fig. 1). The termination point (1, 1) presents the situation in which investigators use the most liberal criterion, deciding to arrest the suspect no matter what response witnesses make (i.e., using the criterion $c_{REJ_{high}}$ in Fig. 1). The points in-between present investigators' true- and false-positive rates when investigators use the criteria in-between these two limits.

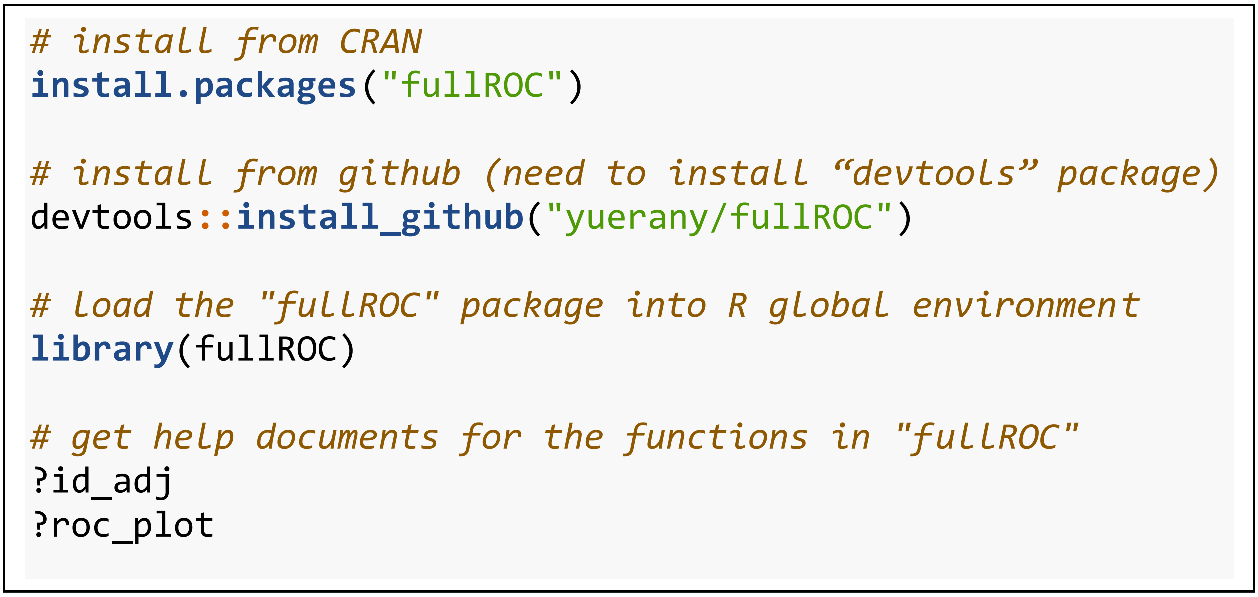
Because the full ROC curves include all possible criteria investigators can use for making investigative decisions, they provide a useful tool for comparing lineup procedures' total *capacity* to help investigators distinguish between guilty and innocent suspects (or total informational value). A lineup's capacity to "rule out" innocent suspects (i.e., to disconfirm the investigator's hypothesis) is just as important as its capacity to "rule in" guilty suspects (i.e., to confirm the investigator's hypothesis). Hence, ROC curves should include all eyewitness outcomes that can inform on suspects' likely guilt or innocence. When including all eyewitness outcomes, a lineup's total informational value is reflected by the area under the full ROC curves (Smith et al., 2020). When ROC curves do not cross over, whichever lineup procedure that produces a larger area under the full ROC curve is the procedure with better informational value for both "rule in" guilty suspects and "rule out" innocent suspects. Such a procedure can help investigators do a better job distinguishing between guilty and innocent suspects.

Although multiple software packages are available for creating ROC curves and calculating AUCs, two unique features of eyewitness data limit the ability of extant packages for analyzing eyewitness data. First, eyewitness evidence is discrete rather than continuous. Therefore, researchers need to order eyewitness outcomes for creating ROC curves. However, most extant packages deal with continuous decision variables, which have a natural order, and thus do not enable researchers to freely change the order of the decision variable. Second, when an experiment does not designate an innocent suspect, researchers need to adjust filler identifications to estimate innocent suspect identifications. This adjustment creates fractional data in culprit-absent lineups. The extant packages cannot bootstrap samples using fractional data and thus cannot inferentially compare ROC curves. To overcome these difficulties, we introduce a package specifically designed for researchers to analyze eyewitness data and to plot and inferentially compare full ROC curves of lineups.

Analyze eyewitness data and create full ROC curves

We introduce *fullROC*, a Comprehensive R Archive Network (CRAN) package for generating and statistically comparing full ROC curves for eyewitness lineups. R is an open-source programming language which supports data visualization and statistical analysis (R core team, 2020). In addition to R, we recommend readers to use RStudio as a convenient interface to R (RStudio, 2020). We illustrate how to use the *fullROC* package to generate ROC figures and to statistically compare lineup ROC curves.

We display all codes with `Consolas` font. In addition to using R base functions, we created functions to calculate adjusted identification rates, plot full ROC curves, and statistically compare the area under the ROC curves (AUC). To use these functions, users can download and install the



```
# install from CRAN
install.packages("fullROC")

# install from github (need to install “devtools” package)
devtools::install_github("yuerany/fullROC")

# load the "fullROC" package into R global environment
library(fullROC)

# get help documents for the functions in "fullROC"
?id_adj
?roc_plot
```

Box 1 Install and load the fullROC package

fullROC package from CRAN (for the stable version) or from GitHub (for the developer version) (see Box 1). Once installed, users can run the library command to load the functions into the R environment. We explain how to use these pre-packed functions in the examples below. Users can also run the `?function` command to retrieve the help documents for the functions.

We include two examples in this paper. The first example uses simulated data to show the data cleaning and organizing process; the second example uses empirical data to compare performance across various experimental conditions.

Simulation example

We simulated witness responses according to the signal detection model shown in Fig. 1. In the simulation, each lineup contained six members (i.e., one guilty suspect and five fillers in the culprit present lineup; six fillers in the culprit absent lineup). We simulated n = 20,000 responses (n = 10,000 for each lineup condition) and saved the simulated responses into an R data frames, simu_data. Box 2 contains a snapshot of the simulated data.

The data frame `simu_data` saved simulated responses in both culprit present and absent conditions. The data frame contained four variables: lineup indicates the lineup condition (culprit present or absent), ID indicates the witnesses' response, confidence indicates the witness' confidence, and IDconfidence combines the witness' response and associated confidence. Because we did not designate an innocent suspect, there were no suspect identifications in culprit absent lineups

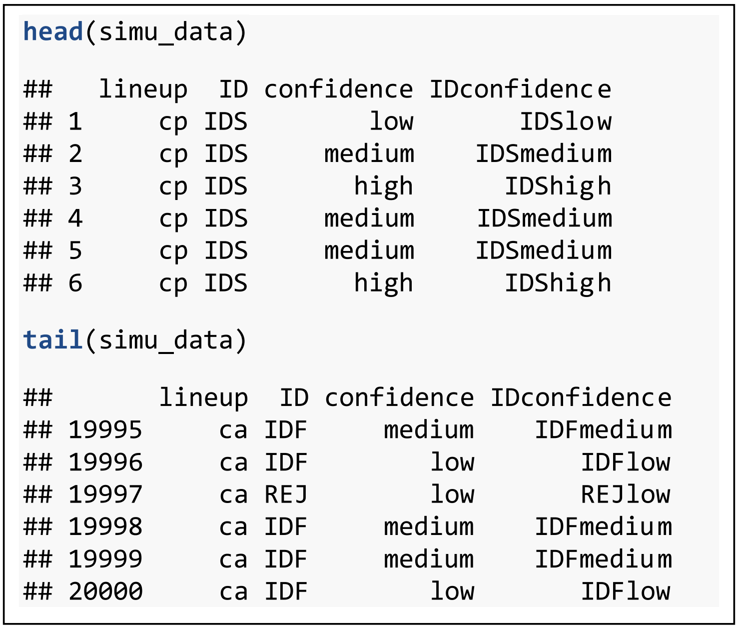
Calculate response frequencies and rates

Box 3 contains the R code to calculate the response frequencies and rates in both lineup conditions. In the R code, we first re-ordered the factor levels of the variables to match the order of those in the experimental design. We then calculated a frequency table for witness responses using the `table` command. We next converted the frequencies into percentages and saved the both the response frequencies and rates into a data frame, `simu`. Table 1 displays the calculated response rates in both culprit present and absent lineups.

Adjust response rates in culprit absent lineups

When researchers do not designate an innocent suspect in eyewitness experiments, the suspect identification rate is zero in culprit absent lineups (see Table 1). In such situations, one common method to estimate the suspect identification rate is to use the 1/(lineup size) method (e.g., Quigley-McBride & Wells, 2021). The 1/(lineup size) method adjusts the suspect identification rate to be the 1/(lineup size) × original filler identification rate and the filler identification rate to be (lineup size – 1)/(lineup size) × original filler identification rate for culprit absent lineups (also see Table 1).

Mathematically, let us use m to present the lineup size and $p(IDF| CA)$ to present the original filler identification rate in



```
head(simu_data)

##   lineup  ID confidence IDconfidence
## 1     cp IDS        low       IDSlow
## 2     cp IDS     medium    IDSmedium
## 3     cp IDS       high      IDShigh
## 4     cp IDS     medium    IDSmedium
## 5     cp IDS     medium    IDSmedium
## 6     cp IDS       high      IDShigh

tail(simu_data)

##       lineup  ID confidence IDconfidence
## 19995     ca IDF     medium    IDFmedium
## 19996     ca IDF        low       IDFlow
## 19997     ca REJ        low       REJlow
## 19998     ca IDF     medium    IDFmedium
## 19999     ca IDF     medium    IDFmedium
## 20000     ca IDF        low       IDFlow
```

Box 2 Snapshot of simulated responses in both culprit present and culprit absent lineups

Table 1 Witness response rates based on n = 10,000 simulations in each lineup condition

Response	Confidence	Lineup condition			Diagnosticity ratio
		Culprit present (n = 10,000)	Culprit absent (n = 10,000)		
			Original	Adjusted	
Suspect identification	High	0.303	0	0.006 = 0.036/6	50.011
	Medium	0.268	0	0.030 = 0.180/6	8.910
	Low	0.158	0	0.072 = 0.432/6	2.189
Filler identification	High	0.023	0.036	0.030 = 0.036×5/6	0.765
	Medium	0.079	0.180	0.150 = 0.180×5/6	0.525
	Low	0.104	0.432	0.360 = 0.432×5/6	0.289
Rejection	High	0	0.0005	0.0005	0
	Medium	0.004	0.046	0.046	0.076
	Low	0.062	0.305	0.305	0.202

Diagnosticity ratio = Culprit present response rate / Adjusted culprit absent response rate

culprit absent lineups. Then the adjusted suspect identification rate will be,

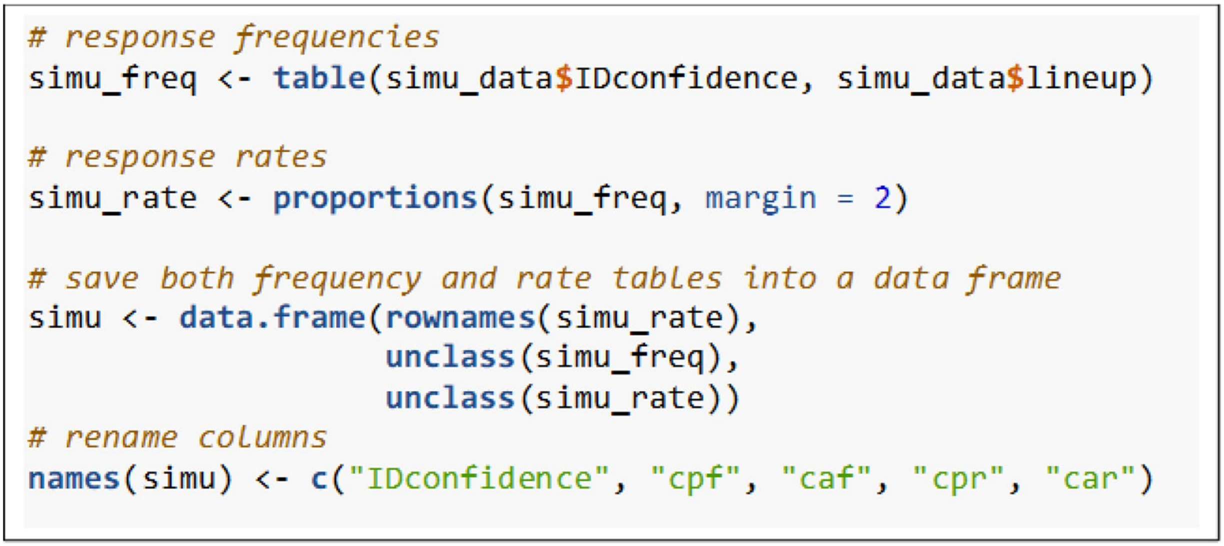
$$p_{adj}(IDS|CA) = \frac{1}{m}p(IDF|CA)$$

The adjusted filler identification rate will be,

$$p_{adj}(IDF|CA) = \frac{m-1}{m}p(IDF|CA)$$

Box 4 contains the R code to adjust the response rates. Specifically, we created a projection matrix `proj` to calculate the suspect and filler identification rates using the 1/(lineup size) method. Note that the order of the response rates in the to-be-adjusted variable `simu$car` must match the order of the multipliers in the projection matrix.

Alternatively, the *fullROC* package provides three different functions to adjust the response rates (see Box 4). The `id_adj()` function uses the projection matrix described above, and requires that all responses must have the same number of confidence levels and be pre-ordered by response type ("IDS", "IDF", and "REJ") and confidence (same confidence order for all response types). When using the `id_adj()` function, users can easily change the lineup size



```
# response frequencies
simu_freq <- table(simu_data$IDconfidence, simu_data$lineup)

# response rates
simu_rate <- proportions(simu_freq, margin = 2)

# save both frequency and rate tables into a data frame
simu <- data.frame(rownames(simu_rate),
                   unclass(simu_freq),
                   unclass(simu_rate))
# rename columns
names(simu) <- c("IDconfidence", "cpf", "caf", "cpr", "car")
```

Box 3 R Code for Calculating Response Rates

by specifying the `lsize` argument (default to 6) and the number of confidence levels by specifying the `csize` argument (default to 3).

The `id_adj_pos()` and `id_adj_name()` functions are more flexible than `id_adj()`. Users can specify the positions of filler and suspect identifications in the `id_adj_pos()` function or the confidence levels for filler and suspect identifications in the `id_adj_name()` function. Box 4 contains the R code using these three functions to adjust response rates for culprit absent lineups.

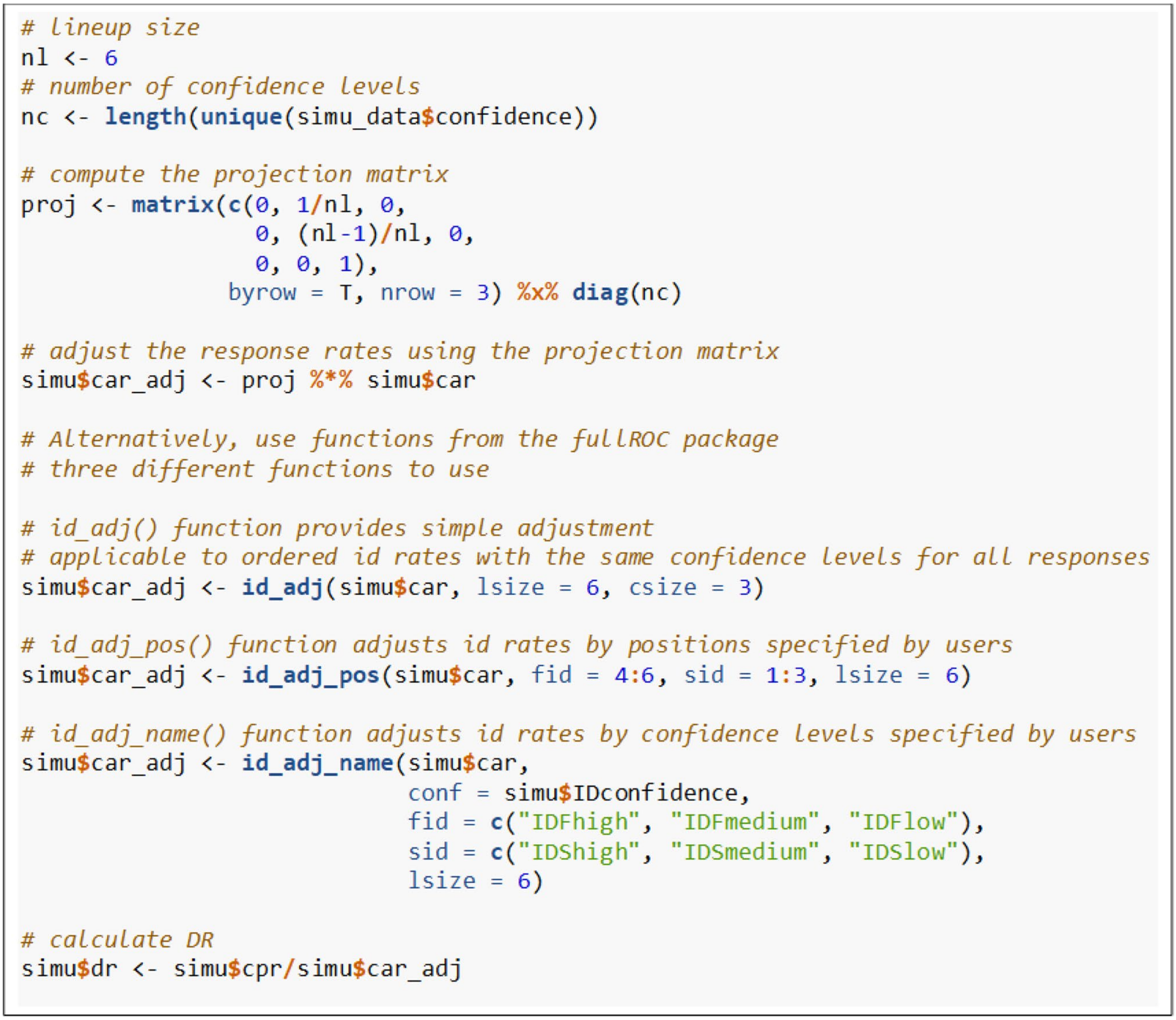
As shown in Box 4, we saved the adjusted response rates for culprit absent lineups into a new variable `simu$car_adj`. Using the adjusted response rates, we also calculated the diagnosticity ratios, which are ratios between guilty suspect identification rates and innocent suspect identification rates. Table 1 displays the original and adjusted responses rates in the culprit absent condition as well as the diagnosticity ratios.

Create ROC curves and calculate AUCs

From the investigator signal detection model, the order of eyewitness outcomes reflects how diagnostic investigators consider the outcomes are in their decision process. This brings about the question of whether eyewitness outcomes should be ordered a priori or based on the diagnosticity of the outcomes. Ordering by diagnosticity ratio results in an upper-bound estimate of the AUC and reflects the best possible practice. However, this method has one potential pitfall: The order of diagnosticity ratios may differ across samples or lineup methods due to sampling errors. Sorting by diagnosticity ratios obtained from sample data then implies that investigators change how they sort eyewitness outcomes across samples or lineup methods. To avoid this problem, an alternative method is to decide the order of eyewitness outcomes *a priori* based on theoretical estimates or empirical meta-analyses (Smith et al., 2020). Such estimates provide better and more stable information on the diagnostic values of different eyewitness outcomes.

In this example, we illustrate how to create full ROC curve using both methods—sorting eyewitness outcomes by a pre-assigned order or by diagnosticity ratios. When pre-assigning the order, we use the order of "suspect identifications high to low confidence", and then "filler identifications and rejections alternately low to high confidence." We chose this order because theoretically any decision with low confidence should have less informational value than decisions made with higher confidence (Wells et al., 2015). However, other orders are possible. For example, the best-above model predicts a dissociation between confidence and the diagnostic values of filler identifications (Smith & Ayala, 2021). According to this model, researcher may collapse confidence levels of filler identifications and thus consider the order of "suspect identifications high to low confidence", "all filler identifications", and then "rejections low to high confidence."

To create ROC curves and calculate AUCs, users can use the `roc_plot()` function in the *fullROC* package. Box 5 shows how to generate an ROC curve and its AUC from the response rates saved in `simu`. As discussed above, users could sort response rates either by a pre-assigned order of confidence levels or by the order of diagnosticity ratios. We created ROC curves for the simulated lineup data with both mechanisms (see Box 5). To sort response rates by confidence levels, we first re-ordered the levels of the `IDconfidence` variable and sorted the data accordingly. We then



```r
# lineup size
nl <- 6
# number of confidence levels
nc <- length(unique(simu_data$confidence))

# compute the projection matrix
proj <- matrix(c(0, 1/nl, 0,
                 0, (nl-1)/nl, 0,
                 0, 0, 1),
               byrow = T, nrow = 3) %x% diag(nc)

# adjust the response rates using the projection matrix
simu$car_adj <- proj %*% simu$car

# Alternatively, use functions from the fullROC package
# three different functions to use

# id_adj() function provides simple adjustment
# applicable to ordered id rates with the same confidence levels for all responses
simu$car_adj <- id_adj(simu$car, lsize = 6, csize = 3)

# id_adj_pos() function adjusts id rates by positions specified by users
simu$car_adj <- id_adj_pos(simu$car, fid = 4:6, sid = 1:3, lsize = 6)

# id_adj_name() function adjusts id rates by confidence levels specified by users
simu$car_adj <- id_adj_name(simu$car,
                            conf = simu$IDconfidence,
                            fid = c("IDFhigh", "IDFmedium", "IDFlow"),
                            sid = c("IDShigh", "IDSmedium", "IDSlow"),
                            lsize = 6)

# calculate DR
simu$dr <- simu$cpr/simu$car_adj
```

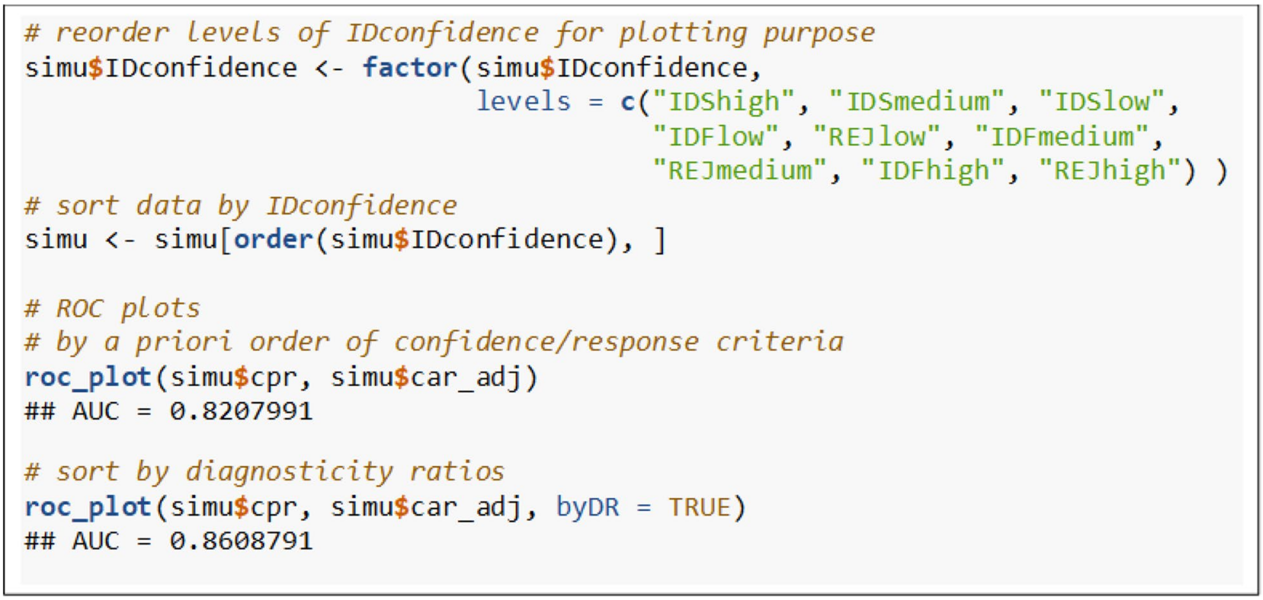
Box 4 R code for adjusting suspect and filler identification rates in culprit absent lineups

used the `roc_plot()` function to plot the ROC curve and calculate the corresponding AUC. To sort response rates by diagnosticity ratios, we used the argument `byDR=TRUE` (default to `FALSE`) in the `roc_plot()` function, which sorts data automatically by diagnosticity ratios before creating ROC curves. Figure 3 displays the ROC curves generated by both methods.

In addition to plotting ROC curves, the *fullROC* package also provides a function, `auc_ci()`, to calculate bootstrap inferential confidence intervals for AUCs. Because `auc_ci()` only takes response frequencies as inputs, we first adjusted the response frequencies for the culprit absent lineups and then calculated the bootstrapping confidence interval for the AUC (see Box 6).

The bootstrap method used in `auc_ci()` differs slightly from the traditional sampling with replacement method. As mentioned above, researchers need to estimate and adjust the innocent-suspect identifications and filler identifications when they do not assign a designated innocent suspect in culprit-absent lineups. Such adjustments sometimes create fractions in the estimated identification frequencies. The traditional sampling with replacement method requires to sample from integer data (or discrete categories), and thus cannot work with fractional data.

The `auc_ci()` function, instead, draws random samples from a multinomial distribution with the probability parameters calculated from the frequencies, which is equivalent to sampling from the original data with replacement (Efron & Tibshirani, 1993). This approach is not influenced



```
# reorder levels of IDconfidence for plotting purpose
simu$IDconfidence <- factor(simu$IDconfidence,
                            levels = c("IDShigh", "IDSmedium", "IDSlow",
                                       "IDFlow", "REJlow", "IDFmedium",
                                       "REJmedium", "IDFhigh", "REJhigh") )
# sort data by IDconfidence
simu <- simu[order(simu$IDconfidence), ]

# ROC plots
# by a priori order of confidence/response criteria
roc_plot(simu$cpr, simu$car_adj)
## AUC = 0.8207991

# sort by diagnosticity ratios
roc_plot(simu$cpr, simu$car_adj, byDR = TRUE)
## AUC = 0.8608791
```

Box 5 R code for creating ROC curves and calculating AUCs

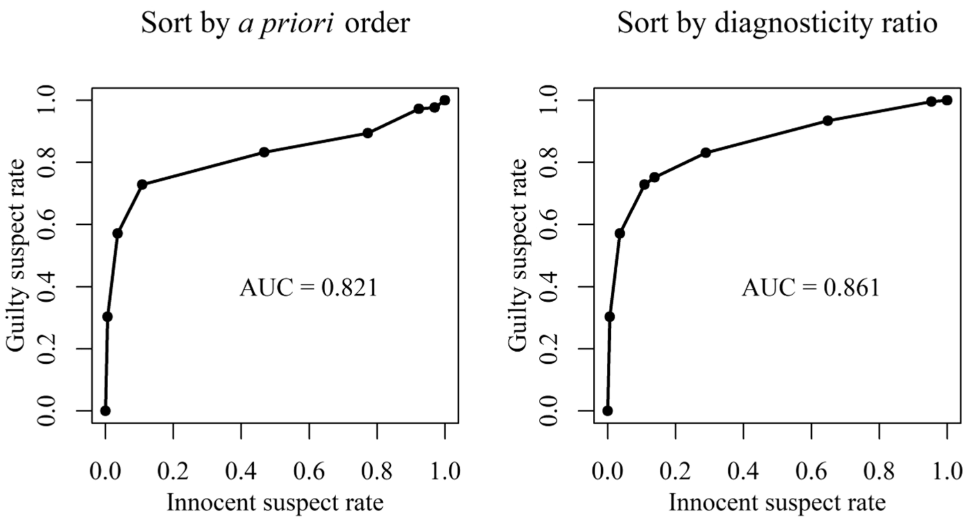
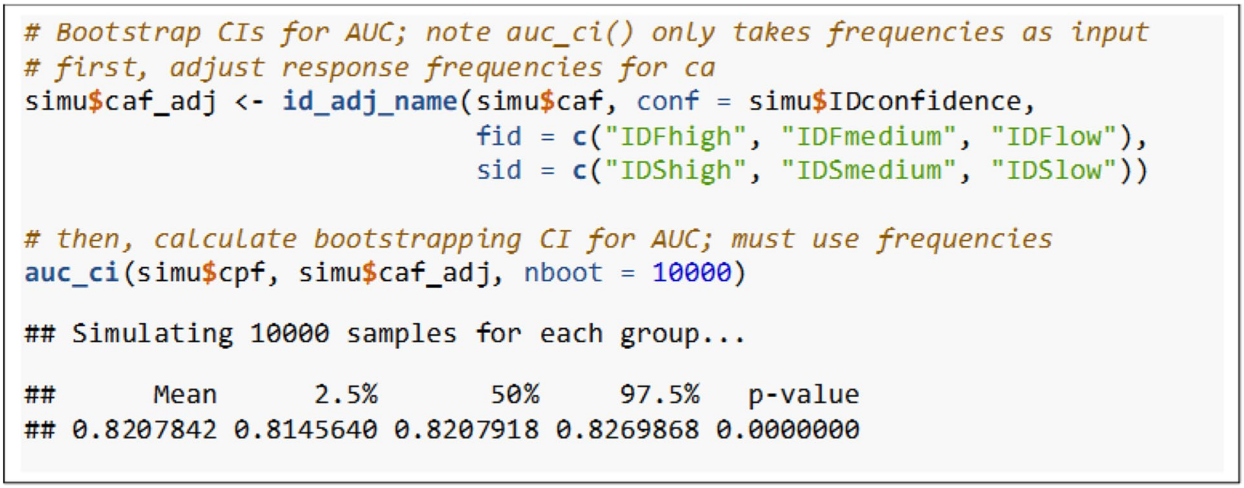


Fig. 3 Investigator ROC curves for simulated data

by whether the frequencies involve fractions, therefore can work with the adjusted frequencies. With these random samples, `auc_ci()` calculates their AUCs and then generates the mean, the percentile confidence interval (CI), and the *p* value for the bootstrap AUCs.

As shown in Box 6, we input the frequency data for both culprit present and absent lineups. Note that the order of the frequency data determines the order of eyewitness outcomes when calculating AUCs, and therefore should be sorted beforehand. Using the frequency data, the `auc_ci()` function simulates 10,000 bootstrap samples and calculates the AUCs. Based on the bootstrap AUCs, `auc_ci()` generates their mean, the 95% bootstrap confidence interval, and the *p* value for the null hypothesis test ($H_0: AUC = 0$ for a single AUC or $H_0: \Delta AUC = 0$ for the difference between AUCs).

The simulated data present a simple example how to analyze eyewitness data and create ROC curves using functions from the *fullROC* package. Next, we show a second example using experimental data that compare eyewitness performance across various experimental conditions.



```
# Bootstrap CIs for AUC; note auc_ci() only takes frequencies as input
# first, adjust response frequencies for ca
simu$caf_adj <- id_adj_name(simu$caf, conf = simu$IDconfidence,
                            fid = c("IDFhigh", "IDFmedium", "IDFlow"),
                            sid = c("IDShigh", "IDSmedium", "IDSlow"))

# then, calculate bootstrapping CI for AUC; must use frequencies
auc_ci(simu$cpf, simu$caf_adj, nboot = 10000)

## Simulating 10000 samples for each group...

##      Mean      2.5%       50%     97.5%   p-value
## 0.8207842 0.8145640 0.8207918 0.8269868 0.0000000
```

Box 6 R code for calculating bootstrap confidence intervals for AUCs

Table 2 Response frequencies in all experimental conditions from Colloff et al. (2016)

	Culprit present			Culprit absent		
	IDS	IDF	REJ	IDS	IDF	REJ
Replication						
0–20	21	45	40	9.5	47.5	40
30–40	36	52	45	16.5	82.5	68
50–60	96	127	124	34.5	172.5	158
70–80	106	93	102	28	140	120
90–100	88	65	85	16	80	127
Pixelation						
0–20	29	52	30	14.83	74.17	41
30–40	38	56	53	19.83	99.17	56
50–60	92	151	130	32.83	164.17	163
70–80	84	105	106	21.83	109.17	134
90–100	77	47	95	13	65	116
Block						
0–20	27	62	41	14.83	74.17	59
30–40	51	61	63	17.5	87.5	62
50–60	101	137	132	37	185	158
70–80	71	93	97	22.17	110.83	143
90–100	73	37	81	9	45	112
Do-nothing						
0–20	17	32	22	18	29	43
30–40	35	36	36	37	50	50
50–60	156	70	88	113	69	122
70–80	155	44	66	74	49	107
90–100	266	24	63	122	22	112

Experimental data example

Our second example uses the experimental data from Colloff et al. (2016); also see Table 2). In addition to culprit presence/absence, Colloff et al. (2016) manipulated the similarity of fillers to a suspect who had a distinctive facial feature via four conditions. The *replication* condition replicated the suspect's distinctive feature on fillers; the *block* condition blocked the suspect's distinctive feature and fillers' corresponding facial areas; the *pixelation* condition pixelated the suspect's distinctive feature and fillers' corresponding facial areas; and the *do-nothing* condition did nothing to fillers but included an innocent suspect with the same distinctive feature in culprit absent lineups. Table 2 listed response frequencies in each confidence bin in all experimental conditions from Colloff et al. (2016). The replication, pixelation, and block conditions did not have a designated innocent suspect in culprit absent lineups. Thus, the suspect and filler identification rates were adjusted using the 1/(lineup size) method for these three conditions in Colloff et al. (2016). The do-nothing condition used a designated innocent suspect in culprit absent lineups, and the identification rates did not get adjusted.

Before creating the ROC curves and calculating AUCs, we re-organized the data so that all responses in culprit present lineups were saved in one column (variable `cpf` in the newly generated data frame `d2`) and all responses in culprit absent lineups in another column (variable `caf` in `d2`). We also re-arranged the order of confidence levels to reflect the pre-assigned criterion order (suspect identifications high to low confidence, and then filler identifications and rejections alternately low to high confidence). To show readers how to process raw lineup data into frequency tables, we also converted Colloff et al. (2016) frequency data (unadjusted) into raw data in which each row corresponds to a datum from one participant. We then started from the raw data to conduct all analyses. The R script for processing raw eyewitness data is available at osf.io/daq7x.

After re-organizing data, we used the `roc_plot()` function to create ROC curves (see Box 7). The `group` argument in `roc_plot()` takes information about the grouping variable and enables the function to create separate ROC curves for different groups. Users can also change the ROC curves to grayscale by setting the argument `grayscale=TRUE` (default to `FALSE`). Figure 4 displays the ROC curves for all four filler similarity conditions.

The ROC curves in Fig. 4 indicate that the replication and block conditions performed similarly, as their ROC curves are overlapping to a large extent. The pixelation condition performed slightly worse than the replication and block conditions, as the pixelation ROC curve is dominated by replication and block ROC curves in the middle range of innocent suspect rates.

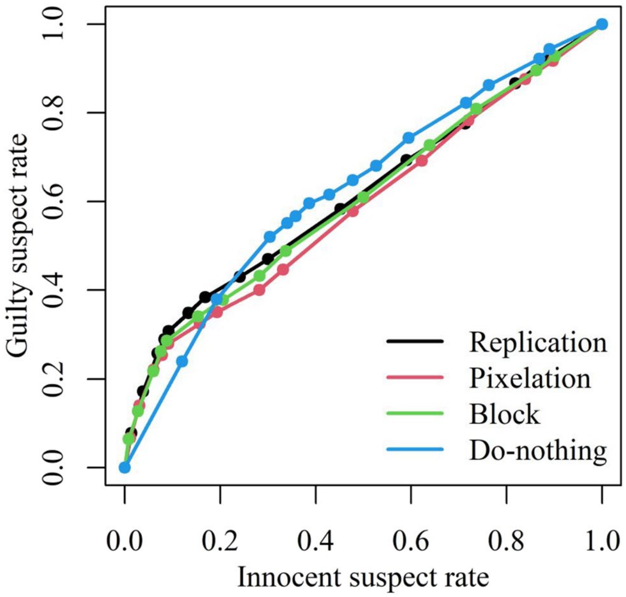


Fig. 4 Investigator ROC curves for Colloff et al. (2016) experimental data

The do-nothing condition, however, performed differently from the other three conditions. The do-nothing ROC curve intersected with the other three ROC curves. The do-nothing ROC curve was dominated by the other three ROC curves before the intersection, but dominated the other three ROC curves after the intersection. To compare their overall performance, we calculated the bootstrap CIs and *p* values for the differences in AUCs among the four similarity conditions. Box 8 displays the R code and outputs. From the bootstrap CIs and *p* values ($H_0: AUC = 0$ for a single AUC or $H_0: \Delta AUC = 0$ for the difference between AUCs), the do-nothing condition had a slightly higher AUC than the pixelation condition (95% CI = [0.003, 0.068], $p = 0.035$, and the AUCs did not differ for other comparisons ($ps \geq 0.150$). To account for multiple comparisons, we also used the Bonferroni adjustment to adjust the significance level, $\alpha^* = \frac{\alpha}{6} = 0.00833$. With the adjusted significance level, none of the AUC differences were statistically significant, $ps \geq 0.035 > 0.00833$.

Based on the comparison of ROC curves and AUCs, the difference between the do-nothing lineups and the unbiased lineups (replication, pixelation, and block) reflects the difference in their capacities to "rule in" guilty suspects and "rule out" innocent suspects. The do-nothing lineups were more liberal than the other three types of lineups. Indeed, because the fillers did not match the suspect on a critical diagnostic facial feature, the suspect tended to stand out from the other lineup members and was picked more frequently in the do-nothing lineups compared to the

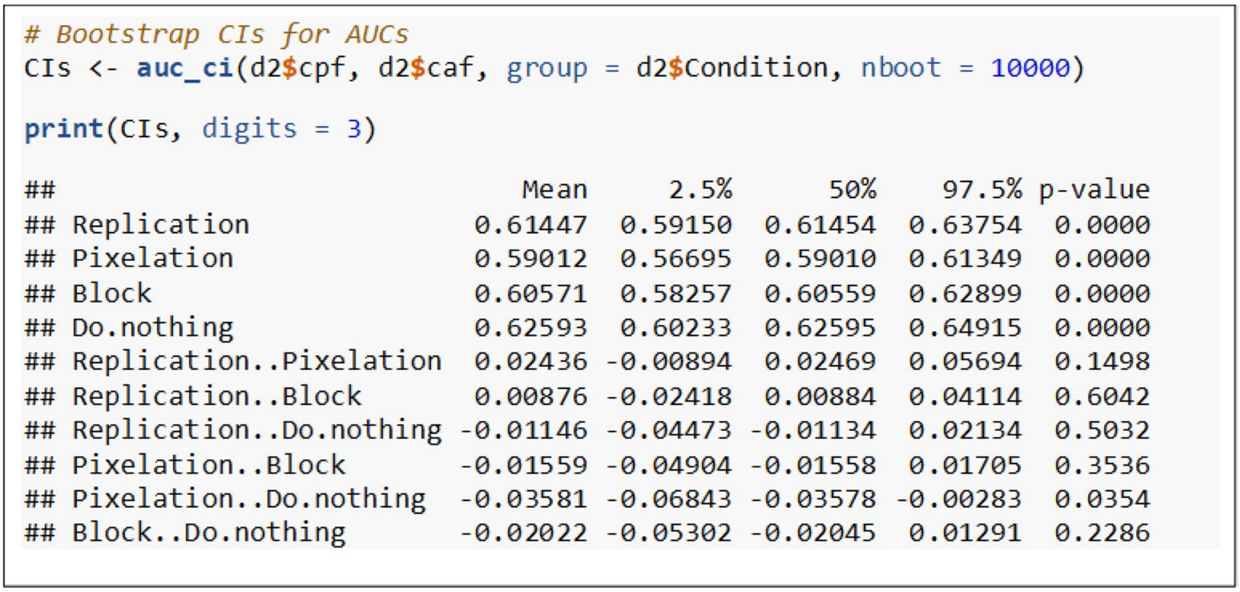
```
# order data by IDconfidence
d2$IDconfidence <- paste0(d2$ID, d2$Confidence)
# reorder factor levels
d2$IDconfidence <- factor(d2$IDconfidence,
                          levels = c("IDS90-100", "IDS70-80", "IDS50-60",
                                     "IDS30-40", "IDS0-20", "IDF0-20",
                                     "REJ0-20", "IDF30-40", "REJ30-40",
                                     "IDF50-60", "REJ50-60", "IDF70-80",
                                     "REJ70-80", "IDF90-100", "REJ90-100"))

# order data
d2 <- d2[order(d2$Condition, d2$IDconfidence), ]


# create ROC curves
roc_plot(d2$cpf, d2$caf, group = d2$Condition)

## Replication
##   AUC = 0.6143125
## Pixelation
##   AUC = 0.590092
## Block
##   AUC = 0.6056836
## Do-nothing
##   AUC = 0.6258768
```

Box 7 R code for creating ROC curves for different conditions



```
# Bootstrap CIs for AUCs
CIs <- auc_ci(d2$cpf, d2$caf, group = d2$Condition, nboot = 10000)

print(CIs, digits = 3)

##                             Mean     2.5%      50%    97.5% p-value
## Replication              0.61447  0.59150  0.61454  0.63754  0.0000
## Pixelation               0.59012  0.56695  0.59010  0.61349  0.0000
## Block                    0.60571  0.58257  0.60559  0.62899  0.0000
## Do.nothing               0.62593  0.60233  0.62595  0.64915  0.0000
## Replication..Pixelation  0.02436 -0.00894  0.02469  0.05694  0.1498
## Replication..Block       0.00876 -0.02418  0.00884  0.04114  0.6042
## Replication..Do.nothing -0.01146 -0.04473 -0.01134  0.02134  0.5032
## Pixelation..Block       -0.01559 -0.04904 -0.01558  0.01705  0.3536
## Pixelation..Do.nothing  -0.03581 -0.06843 -0.03578 -0.00283  0.0354
## Block..Do.nothing       -0.02022 -0.05302 -0.02045  0.01291  0.2286
```

Box 8 R code for calculating bootstrap CIs for AUCs and difference in AUCs other three types of lineups (in which the fillers matched the suspect on the critical diagnostic facial feature). Holding discriminability constant, suspect identifications become increasingly diagnostic of guilt as procedures become increasingly conservative (Macmillan & Creelman, 2005; Wixted & Mickes, 2014). Because the fair lineup procedures were more conservative than the do-nothing procedure, suspect identifications were more diagnostic of guilt from the fair lineups than from the do-nothing (biased) lineup. Conversely, holding discriminability constant, filler identifications and rejections become increasingly diagnostic of innocence as procedures become increasingly liberal. Because the do-nothing (biased) lineup is more liberal than the fair lineups, filler identifications and rejections from the do-nothing lineup were more diagnostic of innocence than were filler identifications and rejections from the three fair lineups. Because all lineups had similar capacities to discriminate guilty suspects and innocent suspects, and the do-nothing lineup was more liberal than the three fair lineups, the ROC curves crossed over.

How do investigator practices influence the area under the ROC curve?

In addition to comparing investigator classification performance across different lineup methods, full ROC curves can also be used to compare investigator classification performance for the same lineup method under different investigative practices. For example, investigators may collapse witness confidence and use fewer confidence bins (e.g., use low, medium, and high confidence for each witness response) when making arrest decisions. Similarly, investigator may use all witness confidence for suspect identifications but collapse confidence for filler identifications and rejections. In other words, investigators may take the same action for a filler identification (or a rejection) regardless of witness confidence. To examine the influence of different practices on investigator classification performance, we compared the ROC curves and the AUCs for the replication lineups under three different practices: investigators use all confidence bins, use reduced confidence bins, or use all confidence bins for suspect identifications but collapse confidence bins for filler identifications and rejections.

From Fig. 5, the ROC curves under the three different practices are overlapping to a large extent, indicating investigator performance would be similar no matter how they use witness confidence to make arrest decisions. To statistically compare the AUCs, we calculated the bootstrap CIs and *p* values for the difference in AUCs among the three practices. Box 9 displays the R code and outputs. From the bootstrap CIs and *p* values, the AUCs did not differ significantly among the three practices (*p*s ≥ 0.623).

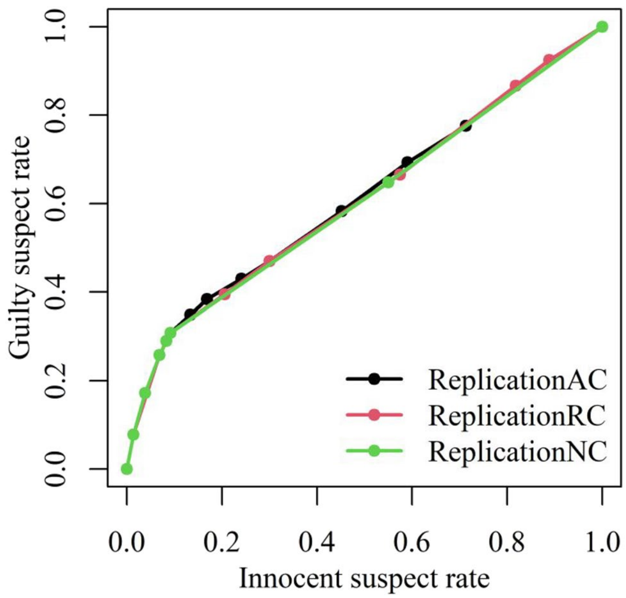


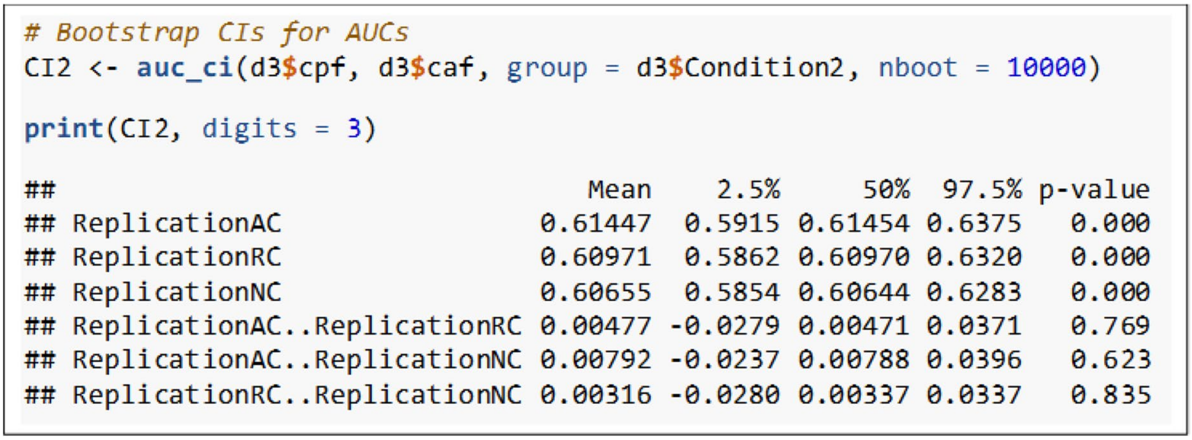
Fig. 5 Investigator ROC curves under different practices. *AC* stands for All Confidence, *RC* stands for Reduced Confidence, and *NC* stands for No Confidence (collapse all confidence for filler identifications and rejections)

Functions in *fullROC*

As shown in the two examples above, the *fullROC* package provides useful functions to create full ROC curves and statistically compare AUCs for lineup data. Other than the functions introduced above, the *fullROC* package provides additional functions to simulate witness responses, calculate AUCs (without generating ROC curves), etc. Table 3 summarizes the functions provided by *fullROC*. Readers can read the help documents for more detail about these functions and their usage.

Discussion

The purpose of a lineup procedure is to provide evidence for investigators to test their hypothesis that the police suspect is the culprit. When the investigator presents the lineup to a witness, in effect, the investigator is using the witness' recognition memory to shed light on the hypothesis that the suspect is the culprit. If the witness identifies the suspect from the lineup, the investigator has confirmatory evidence towards the hypothesis that the suspect is the culprit (i.e., "rule in" guilty suspects). If the witness identifies a known innocent-filler or rejects the lineup, the investigator has disconfirmatory evidence towards the hypothesis that the suspect is the culprit (i.e., "rule out" innocent suspects). In other words, suspect identifications have inculpatory value and filler identifications and rejections have exculpatory



```
# Bootstrap CIs for AUCs
CI2 <- auc_ci(d3$cpf, d3$caf, group = d3$Condition2, nboot = 10000)

print(CI2, digits = 3)

##                                 Mean    2.5%     50%  97.5% p-value
## ReplicationAC                0.61447  0.5915 0.61454 0.6375   0.000
## ReplicationRC                0.60971  0.5862 0.60970 0.6320   0.000
## ReplicationNC                0.60655  0.5854 0.60644 0.6283   0.000
## ReplicationAC..ReplicationRC 0.00477 -0.0279 0.00471 0.0371   0.769
## ReplicationAC..ReplicationNC 0.00792 -0.0237 0.00788 0.0396   0.623
## ReplicationRC..ReplicationNC 0.00316 -0.0280 0.00337 0.0337   0.835
```

Box 9 R code for comparing AUCs under different investigative practices

Table 3 Functions in *fullROC*

Function	Usage
auc_boot	Bootstrap witness responses from input frequencies and calculate the AUCs.
auc_ci	Calculate the means, percentile confidence intervals, and *p* values for the bootstrap AUCs and their differences.
id_adj	Adjust identification rates for culprit absent lineups using the 1/(lineup size) method. All witness responses must have the same confidence levels and be sorted by confidence levels.
id_adj_name	Adjust identification rates for culprit absent lineups using the 1/(lineup size) method. Match suspect and filler ids by input confidence names.
id_adj_pos	Adjust identification rates for culprit absent lineups using the 1/(lineup size) method. Match suspect and filler ids by input positions.
response_simu	Simulate witness responses from Gaussian distributions (assuming equal variance) using the BEST decision rule.
roc_auc	Calculate the area under an ROC curve.
roc_plot	Plot an ROC curve for lineup data and calculate the area under the curve.

value. In order to determine which of two lineup procedures is superior, one must consider not only the relative capacities of those procedures to incriminate – as has often been done in the eyewitness identification literature – but must also consider the relative capacities of those procedures to exculpate (as has rarely been done in the eyewitness identification literature).

Smith et al. (2020) recently introduced a method for creating ROC curves of lineups that include all three lineup outcomes and inform on the total informational value of a lineup to facilitate investigators to sort between guilty suspects and innocent suspects. Only by considering all outcomes of lineups and the total capacity of a lineup to sort between culprits and innocent suspects can we make informed decisions about which of two lineup procedures is superior. In the present work, we introduced a CRAN package for implementing the full ROC approach to comparing eyewitness lineups. Among the many functionalities of this package are the capacities to both plot and inferentially compare full lineup ROC curves.

This is the first and only package for implementing full ROC curves of eyewitness lineup procedures. While there are several packages available for implementing standard ROC analysis in R, including *pROC* (Robin et al., 2011) and *ROCR* (Sing et al., 2005), none of these packages are equipped for plotting and inferentially comparing full ROC curves of eyewitness lineup procedures. As shown in the above examples, investigator classification tasks and eyewitness data have several unique features that are not well accommodated by the extant ROC packages.

First of all, the extant packages are intended for traditional binary classification paradigms in which the decision variable distributions are continuous (e.g., signal plus noise versus noise-only distributions). For investigator classification tasks, however, the decision variable distributions are discrete rather than continuous because investigators use discrete eyewitness responses (suspect identifications, filler identifications, and rejections) to make classification decisions. The discrete nature of eyewitness responses enables researchers to compare investigator performance when assuming investigators would use and sort eyewitness evidence from the same procedure differently. As shown in our analysis, the *fullROC* package allows researchers to freely order or collapse investigator operating points on ROC curves and thus to easily compare investigator performance under different assumptions.

In addition, a unique problem in lineup identification research is that suspect identifications are often estimated by dividing the total false-positive rate from the culprit-absent lineup by the total number of lineup members (usually six). The result is that the estimated number of innocent-suspect identifications and filler identifications from the culprit-absent lineup are not round numbers. For example, if a culprit-absent lineup resulted in 34 false-positive identifications, the estimated innocent-suspect identification rate would be 5.67 (34 false positives / 6 lineup members), and the estimated culprit-absent filler-identification rate would be 28.33 (34 total false positives – 5.67 estimated innocent-suspect identifications). This is particularly problematic for inferential comparison of lineup ROC curves. The bootstrap methods available in extant ROC packages are designed to sample whole-number frequencies but cannot deal with fractions. Hence, in order to use existing packages to inferentially compare lineup ROC curves, one would be forced to round their innocent-suspect and filler identification rates to whole numbers. In fact, the rounding could be even more complicated, because the rounding must be done at every single level of an eyewitness' expressed level of confidence. This is both less than optimal and not feasible without extremely large numbers of observations. By comparison, *fullROC* introduces a method for bootstrapping samples using fractional frequency data generated from adjustments. This is one prime example of how *fullROC* is uniquely situated for the analysis of eyewitness lineup data.

For decades, those researching eyewitness lineup procedures have made inferences about which of two procedures is superior based only on their relative capacities to incriminate (cf. Wells et al., 2015; Wells & Lindsay, 1980; Wells & Turtle, 1986). But the diagnostic value of a lineup is not limited to its capacity to incriminate. Lineups also have the capacity to exculpate. Exculpating the innocent is every bit as important as inculpating the guilty. The goal of a lineup is to determine if the suspect is guilty or innocent. When a witness picks out a filler or rejects a lineup, such a behavior is informative because it suggests that the suspect might be innocent. All behaviors that inform on the likely guilt or innocence of the culprit must be factored into any inference about which of two lineup procedures is superior. Indeed, the procedure that is inferior at inculpating is sometimes superior at exculpating and vice versa (e.g., Smith et al., 2020). If we want to determine which of two procedures better helps investigators sort between the presence and absence of the culprit in a lineup, our analyses and inferences must take all lineup outcomes into account.

The full ROC approach to analyzing eyewitness lineup data creates a new standard for those researching lineup procedures and making applied inferences. As both inculpatory and exculpatory behaviors inform on whether a suspect is guilty or innocent, applied inferences about which of two procedures is superior must be informed by both sets of behaviors. *fullROC* provides researchers with a desperately needed tool for analyzing eyewitness lineup data.

Acknowledgments We appreciate the support of the National Science Foundation, SES 2017046, for this research project. We have no conflicts of interest to disclose. The data and materials can be accessed at osf.io/daq7x.

References

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, *7*(3), 238–259. https://doi.org/10.1177/1745691612439584

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, *27*(9), 1227–1239. https://doi.org/10.1177/0956797616655789

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*(4), 361–376. https://doi.org/10.1037/a0030609

Quigley-McBride, A. & Wells, G. L. (2021). Methodological considerations in eyewitness identification experiments. In A. M. Smith, M. P. Toglia, & J. M. Lampinen (Eds.), Methods, measures, and theories in eyewitness recognition research. Taylor & Francis.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/. Accessed 11 Dec 2020

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. https://doi.org/10.1186/1471-2105-12-77

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio. http://www.rstudio.com/

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics*, *21*(20), 3940–3941. https://doi.org/10.1093/bioinformatics/bti623

Smith, A. M. & Ayala, N. T. (2021). Do traditional lineups undermine the capacity for eyewitness memory to rule out innocent suspects? *Journal of Applied Research in Memory and Cognition, 10*(2), 215-220. https://doi.org/10.1016/j.jarmac.2021.03.003

Smith, A. M., Yang, Y., & Wells, G. L. (2020). Distinguishing Between investigator discriminability and eyewitness discriminability: A method for creating full receiver operating characteristic curves of lineup identification performance. *Perspectives on Psychological Science*, *15*(3), 589–607. https://doi.org/10.1177/1745691620902426

Starns, J. J., Cohen, A. L., & Rotello, C. M. (2021). A complete method for assessing the effectiveness of eyewitness identification procedures: Expected information gain. *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000332

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285-1293. https://doi.org/10.1126/science.3287615

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior, 44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, *88*(3), 776–784. https://doi.org/10.1037/0033-2909.88.3.776

Wells, G. L., & Luus, C. A. E. (1990). Police lineups as experiments: Social methodology as a framework for properly conducted lineups. *Personality and Social Psychology Bulletin*, *16*(1), 106–117. https://doi.org/10.1177/0146167290161008

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, *8*(3), 155–167. https://doi.org/10.1037/1076-898X.8.3.155

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin*, *99*(3), 320–329. https://doi.org/10.1037/0033-2909.99.3.320

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior*, *39*(2), 99–122. https://doi.org/10.1037/lhb0000125

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, *4*(1), 8–14. https://doi.org/10.1016/j.jarmac.2014.07.003

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science*, *7*(3), 275–278. https://doi.org/10.1177/1745691612442906

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262–276. https://doi.org/10.1037/a0035940

Yang, Y. & Moody, S. A. (2021). Use and misuse of receiver operating characteristic analysis in eyewitness identification. *Journal of Applied Research in Memory and Cognition. 10*(3), 335–340. https://doi.org/10.1016/j.jarmac.2021.06.003

Open sciences statement The data and codes used in the manuscript can be accessed at osf.io/daq7x. The source code of the *fullROC* package can be downloaded from the Comprehensive R Archive Network (CRAN) and the first author's GitHub page (github.com/yuerany/fullROC).