Supporting Information

The Reliability of Eyewitness Identifications from Police Lineups

John T. Wixted, Laura Mickes, John C. Dunn, Steven E. Clark, & William Wells

Inventory of Supplemental Information:

Methods

The experiment was conducted in the Houston Police Department (HPD) Robbery Division because the largest volume of photo spread and lineup procedures are conducted by robbery investigators.

Experimental Conditions

The experiment was designed to test the outcomes of four different methods of showing photos and individuals to eyewitnesses:

- blind sequential
- blind simultaneous
- blinded sequential
- blinded simultaneous

Each of these procedures was consistent with HPD eyewitness identification procedures at the time of the study and thus did not introduce new procedures.

With the simultaneous presentation method the eyewitness viewed all photos or live lineup members at the same time. In the sequential presentation method, photos and live lineup members were viewed one at a time. In the blind procedure, the primary investigator who knew the identity of the suspect prepared the photo spread, but an investigator with no knowledge of the suspect's identity administered the viewing with the witness. In the blinded procedure, the primary investigator conducted the viewing, but a mechanism was used to prevent the investigator from 1) knowing the suspect's position in the photo spread or lineup and 2) knowing which photo or individual the suspect was viewing.

For the blinded procedures, the primary investigator selected the suspect's photo and the 5 filler photos. A different investigator then randomly ordered the photos and placed them into a folder (or folders) for the investigating officer to administer. For the sequential procedure, one photo was placed into each of 6 folders. The stack of folders was then provided to the investigating officer, who then administered the photos one at a time to the witness. During the procedure, the investigator was positioned so it was not possible to view the photos while the witness viewed the photos. The blinded simultaneous procedure was similar except that the investigator who randomly ordered the photos placed the simultaneous photospread into a single folder before providing it to the investigating officer (who then administered the photos simultaneously to the witness).

In late 2012 all HPD criminal investigators were trained on these procedures. Investigators were provided with a user-friendly instruction booklet they could reference throughout the study period. This provided investigators with information about the procedures in the event that questions arose when they were preparing to conduct an identification procedure. A complete description of the blind and blinded procedures used in the experiment can be found in Appendix F of W. Wells (24).

Informed Consent

All of the investigators who participated in the study signed an informed consent document (SHSU protocol # 2012-08-202) and witnesses were provided with a cover letter that explained risks and their rights. In addition, at the conclusion of the ID procedure, a survey was provided to each witness. The survey asked questions about how the photos were shown to them (all at once or one at a time); could the detective see which photos they were viewing; did they pick someone from the photos; etc. If they completed and returned the survey to the detective then they were agreeing to participate.

Case Assignment

All robbery reports made to HPD are routed to the Robbery Division. When a case is referred to the Robbery Division it is placed into one of the following categories: *work, contact, pending*, *monitor, office*, and *transfer*. “Work” and “office” cases are assigned to robbery investigators because these cases have investigative leads, including information about suspects, video footage, or license plate numbers of involved vehicles. “Office” cases are those in which the individuals involved are non-strangers, whereas "work" cases involve strangers. “Contact” cases do not contain leads so they are not assigned to an investigator. “Pending” cases do not contain strong investigative leads but are assigned to investigators because leads may develop. “Monitor” robbery cases are assigned to be investigated by another unit, such as the Special Victims Division or the Juvenile Unit. “Transfer” cases have been referred to the Robbery Division but should be assigned to another investigative division.

All “work,” “office,” and “pending” reports received by the Robbery Division were assigned to receive one of the four experimental procedures for showing photo spreads and lineups to eyewitnesses. These cases were assigned to a treatment condition because they generate investigative activities, such as lineups. The characteristics of each determined whether a live lineup or a video lineup would be conducted (e.g., prosecutors may request a live viewing when a suspect is in custody). Only a small percentage of cases involved live or video lineups, and they were not included in any of our analyses, which focused solely on the more commonly used photo spread lineup procedure.

When a robbery report is routed to the Robbery Division, it is automatically assigned an HPD case number (referred to as an “incident number” by HPD personnel) by the computerized information management system. The case number determined which experimental procedure was used in a case. The third digit and the fourth digit from the last digit represent sequential numbers that are automatically generated by HPD data systems. The third from the last digit determined whether a simultaneous or a sequential method was used (simultaneous when even, sequential when odd). The fourth from the last digit determined whether a blinded or a blind method was used (blinded when even, blind when odd). Thus, assignment to condition was a non-systematic, pseudo-random process.

Note that the study randomly assigned *investigations* to a treatment condition, not identification procedures. This means that when there were multiple eyewitnesses to a crime, all of the eyewitnesses were tested using the same procedure (to avoid confusing attorneys, judges, and jurors if the case proceeded to trial). In addition, when several crimes were judged to consist of a connected series of related crimes, the cases were also treated as a single investigation. Therefore, all eyewitnesses in these cases were tested using the same procedure. This restriction on our research protocol was necessary to avoid any detrimental effects of our study on the judicial process, but it also introduced the risk of the four conditions being unbalanced with respect to key variables. Therefore, we performed a variety of manipulation checks described below.

Our analysis was limited to 717 photo spreads that satisfied the following criteria: (1) the robberies involved strangers, (2) witnesses who had not previously viewed a photo spread with the suspect, and (3) photo spreads that followed the experimental protocol that should have been used during the investigation. The number of cases in each condition that satisfied these criteria were: Blinded Simultaneous (N = 194), Blinded Sequential (N = 175), Blind Simultaneous (N = 187), Blind Sequential (N = 161).

Results

The raw frequency counts and proportions of suspect IDs and filler IDs (broken down by confidence) and no IDs for the blind and blinded simultaneous and sequential conditions are shown in Table S1. For the blind condition, overall response bias was similar (e.g., the proportion of no IDs was nearly identical for simultaneous and sequential lineups). For the blinded condition, overall response bias appears to be higher for the sequential condition (e.g., the proportion of no IDs was considerably higher for simultaneous lineups than sequential lineups). However, as described in more detail below, this likely reflects the fact that the blinded sequential condition ended up with more guilty suspects (hence more suspect choosing) than the simultaneous procedure.

Manipulation Checks

Experiment Investigator Survey. For each case, an investigating officer filled out a questionnaire that addressed many issues pertaining to the case (e.g., where was the lineup conducted?, is there independent evidence of suspect guilt?, what was the level of confidence expressed by the eyewitness?, etc.). The full questionnaire can be found in Appendix A of W. Wells (2014, "Eyewitness experiment investigator survey"). The first step in the analysis determined the degree to which the randomization procedure generated four treatment groups that were roughly equivalent in terms of key variables. Table S2 shows the results of these 11 comparisons. One result that stands out concerns corroborating evidence (second line). For this questionnaire item, the investigating officer answered "Was any corroborating evidence available in this case?," with the options being "Yes" or "No (eyewitness ID may be the only link between an offender and the crime)." For 3 of the 4 conditions, the values fell in a fairly narrow range of .62 to .70, but for the blinded sequential condition, the value was conspicuously higher (.91). The difference in the proportion of lineups with corroborating evidence was highly significant, $\chi^2(3) = 45.90$, $p < .00001$. This difference remains significant if the alpha level is adjusted to control for the 11 comparisons shown in Table S2 (using the Bonferroni correction, the alpha level is .05 / 11 = .0045). If the existence of corroborating evidence is assumed to be a proxy for likely guilt, then this result suggests that the blinded sequential condition ended up with a much higher proportion of guilty suspects than the other three conditions, which were similar to each other. Fortunately, as described later, the fact that the equal-variance signal-detection model includes a base rate parameter (p_{target}) means that the two lineup procedures can still be differentially evaluated despite the apparent difference in base rates of guilty suspects for the blinded simultaneous and sequential lineups.

Two other effects are also significant even after applying the Bonferroni correction for the inflation of Type I error. One effect is the "Location" in which the lineup was administered. This effect is entirely understandable and reflects the fact that police officers tended to conduct blind lineups (which required another officer) in a police facility. The other effect is "Interpreter Used." There is no obvious reason why this variable would differ across conditions. The other apparent trends (e.g., for Witness Under Influence, Witness Saw Photo, and Witness not Wearing His/Her Glasses) were not significant using the Bonferroni-corrected alpha level of .0045. Nevertheless, in the main article, we also performed comparisons between the blind simultaneous and sequential lineups after excluding the 65 witnesses who fell into these categories. When that was done, the simultaneous lineup advantage was no longer significant.

Lineup Fairness. To examine the potential differences between the four experimental groups, the fairness of photo spreads was examined for a sample of 60 randomly selected photo spreads. This analysis measures the degree to which the suspect "stands out" in the lineup by providing the lineups to mock witnesses to see if they could identify the suspect. Fifteen photo spreads were randomly selected from each of the four experimental conditions. The sample of mock witnesses included 15 HPD police cadets, 15 volunteers from the HPD Positive Interaction Program (PIP), and 19 Sam Houston State University students (N = 49 mock witnesses in all). Each mock witness was provided with the descriptions of the suspect and then attempted to identify that individual from the lineup.

Table S3 shows the proportion of mock witnesses who identified the suspect from the 15 lineups drawn from each of the four conditions. This measure of lineup fairness did not differ significantly for simultaneous lineups (mean = .21) vs. sequential lineups (mean = .25), but it differ significantly for blinded lineups (mean = .28) vs. blind lineups (mean = .18), $t(58) = 2.31$, $p = .024$.

The mean of the blinded lineups (.28) fell significantly above the expected value of one-sixth (.167) for a fair lineup, $t(29) = 3.14$, $p < .01$, but the mean of the blind lineups (.18) did not differ significantly from the expected value for a fair lineup, $t(29) = 0.76$. Three of the 30 blinded lineups had values that fell more than two standard deviations below the mean of .28 (biased away from the suspect), whereas 15 had values that fell more than two standard deviations above the mean of .28 (biased towards the suspect). By contrast, 4 of the 30 blind lineups had values that fell more than two standard deviations below the mean of .18 (biased away from the suspect), whereas 7 had values that fell more than two standard deviations above the mean of .18 (biased towards the suspect). The fact that the blinded lineups were, on average, unfair (biased towards the suspect) is another reason why our primary analysis focused on the blind lineups. The difference in lineup fairness between the blind and blinded conditions may indicate that police investigators who are aware that the lineup they have constructed will soon be scrutinized by a different investigator (the one who will be asked to administer it) take greater care to ensure that they prepare a fair lineup.

High-Threshold Model

The derivation of Equations 5 and 6 in the main text is straightforward. Adding Equations 2 and 4 yields:

$$F = n_{TP} \, [\, (1\text{-}p) * g * (5/6) \,] + n_{TA} * [\, g * (5/6) \,]$$

Setting $n_{TP} = n_{TA} = n$, where $n = N / 2$ (i.e., assuming equal base rates), we can algebraically solve for p:

$$F = n \, [\, (1\text{-}p) * g * (5/6) \,] + n * [\, g * (5/6) \,]$$

$$F = (1\text{-}p) * [\, n * g * (5/6) \,] + n * g * (5/6)$$

$$F = n * g * (5/6) - p * n * g * (5/6) + n * g * (5/6)$$

$$F = 2 * n * g * (5/6) - p * n * g * (5/6)$$

$$(6 / 5) * F = 2 * n * g - p * n * g$$

$$p * n * g = 2 * n * g - (6 / 5) * F$$

$$p = [\, 1 / (n * g) \,] * [\, 2 * n * g - (6 / 5) * F \,]$$

$$p = 2 - (6 / 5) * F / (n * g)$$

$$p = 2 - 6 * F / (5 * n * g) \qquad [S1]$$

Next, we can use Equation S1 to solve for g. Adding Equations 1 and 3 and setting $n_{TP} = n_{TA} = n$ yields:

$$S = n \, [\, p + (1\text{-}p) * g * (1/6) \,] + n * [\, g * (1/6) \,]$$

$$S = n * p + (1\text{-}p) * n * g * (1/6) + n * g * (1/6)$$

Substituting the expression for p above into this expression yields a rather complex equation that is easily simplified:

$$S = n * [2 - 6 * F / (5 * n * g)] + \{1 - [2 - 6 * F / (5 * n * g)]\} * n * g * (1/6) + n * g * (1/6)$$

$$S = n * [2 - 6 * F / (5 * n * g)] + [1 - 2 + 6 * F / (5 * n * g)] * n * g * (1/6) + n * g * (1/6)$$

$$S = n * 2 - 6 * F / (5 * g) + n * g * (1/6) – 2 * n * g * (1/6) + F / 5 + n * g * (1/6)$$

$$S = n * 2 - 6 * F / (5 * g) + F / 5$$

$$S = n * 2 + F / 5 - 6 * F / (5 * g)$$

This expression can now be solved for g:

$$S - n * 2 - F / 5 = - 6 * F / (5 * g)$$

$$S - n * 2 - F / 5 = - 6 * F / 5 * (1/ g)$$

$$g = - (6 * F / 5) / (S - n * 2 - F / 5)$$

Multiplying the numerator and denominator by -5 yields:

$$g = (6 * F) / (10 * n – 5 * S + F) \qquad [S2]$$

This is Equation 5 in the main article. Substituting Equation S2 for g in Equation S1 allows us to solve for p:

$$p = 2 - 6 * F / (5 * n * [(6 * F) / (10 * n – 5 * S + F)])$$

$$p = 2 – [6 * F / (5 * n)] * [(10 * n – 5 * S + F) / (6 * F)]$$

$$p = 2 – (10 * n – 5 * S + F) / (5 * n)$$

$$p = (10 * n - 10 * n + 5 * S - F) / (5 * n)$$

$$p = (5 * S - F) / (5 * n)$$

This is Equation 6 in the main article.

As noted in the main article, once g and p are known, they can be substituted into Equations 1 and 3 to estimate the number of suspect IDs from target-present and target-absent lineups (nS_{TP} and nS_{TA}, respectively), and these values can be used to estimate suspect ID accuracy, which equals $nS_{TP} / (nS_{TP} + nS_{TA})$.

The equations for g and p (Equations 5 and 6, respectively) allow one to compute suspect ID accuracy for all IDs, but one would also like to use the high-threshold model to compute suspect ID accuracy separately for each level of confidence. To do so, the computational steps involved in computing suspect ID accuracy are identical except that, now, confidence-specific values for S and F are used to compute g and p from Equations 5 and 6 (the value of n in those equations remains unchanged). However, making this move implies additional parameters that allow for different levels of confidence to be expressed when in the above threshold state (which occurs with probability p) or in the guessing state (which occurs with probability g). For example, probability of a high-confidence guess is given by:

$$g * c_{High} = (6 * F_{High}) / (10 * n - 5 * S_{High} + F_{High}) \quad [S3a]$$

where c_{High} is the probability of expressing high confidence in an ID despite being in the guessing state. The probability of instead expressing medium or low confidence in an ID despite being in the guessing state are c_{Med} and c_{Low} such that:

$$g * c_{Med} = (6 * F_{Med}) / (10 * n - 5 * S_{Med} + F_{Med}) \quad [S3b]$$

$$g * c_{Low} = (6 * F_{Low}) / (10 * n - 5 * S_{Low} + F_{Low}) \quad [S3c]$$

where $c_{High} + c_{Med} + c_{Low} = 1$.

Similar considerations apply to Equation 6 such that

$$p * d_{High} = (5 * S_{High} - F_{High}) / (5 * n) \quad [S4a]$$

$$p * d_{Med} = (5 * S_{Med} - F_{Med}) / (5 * n) \quad [S4b]$$

$$p * d_{Low} = (5 * S_{Low} - F_{Low}) / (5 * n) \quad [S4c]$$

where $d_{High,}$ d_{Med} and d_{Low} represent the probability of expressing high, medium or low confidence in a guilty suspect ID despite when in the above-threshold (i.e., recognition) state, and $d_{High} + d_{Med} + d_{Low} = 1$.

Note that the values on the right side of each of these equations are known, so it is possible to directly compute the 6 combined parameter values on the left side of these equations. That is, one can directly compute these six parameters (param_1 through param_6):

$\text{param}_1 = g{*}c_{High}$

$\text{param}_2 = g{*}c_{Med}$,

$\text{param}_3 = g{*}c_{Low}$

$\text{param}_4 = p{*}d_{High}$

$\text{param}_5 = p{*}d_{Med}$

$\text{param}_6 = p{*}d_{Low}$.

There are only 6 degrees of freedom in the data, so this is as far as one can go. There are 6 degrees of freedom because the raw data fall into 7 cells: 3 confidence-specific suspect ID counts, 3 confidence-specific filler ID counts, and the number of no IDs. Still, these 6 combined parameter values are all that are needed to compute confidence-specific estimates of suspect ID accuracy. For example, to compute high-confidence suspect ID accuracy, $g * c_{High}$ is first computed from Equation S3a. There were 17 high-confidence filler IDs ($F_{High} = 17$), 71 high-confidence suspect IDs ($S_{High} = 71$), and, assuming equal base rates, 348 / 2 = 174 target-present lineups and 174 target-absent lineups ($n = 174$). Thus, according to Equation S3a, $g{*}c_{High} = 0.07$. According to Equation S4a, $p{*}d_{High} = 0.39$. To estimate the number of correct and incorrect suspect IDs made with high confidence, one simply substitutes $g{*}c_{High}$ for g in Equations 1 and 3 and substitutes $p{*}d_{High}$ for p in Equation 1. With $n_{TP} = n_{TA} = n$, the estimated correct and incorrect high-confidence suspect IDs come to 68.9 and 2.1, respectively, so estimates of suspect ID accuracy = 68.9 / (68.9 + 2.1) = .97.

Signal-Detection Model Fits

The signal detection model illustrated in Fig. 2 uses a simple decision rule according to which an ID is made if the most familiar person in a lineup exceeds the lowest confidence criterion (c1), with confidence (Low, Medium or High) being determined by the highest criterion that is exceeded. Predictions from a model with a decision rule like that cannot be directly computed (unlike when a signal detection model is fit to a simple list memory experiment, where only a single item is presented for a yes/no decision). Thus, the signal detection model was fit to the experimentally controlled field data and the Houston Police Department field data via Monte Carlo simulation.

First, initial values were set for each of the model parameters (μ_{Target}, σ_{Target}, c1, c2, and c3). The values of μ_{Lure} and σ_{lure} were always fixed at 0 and 1, respectively. Next, for each of n_{TP} simulated witnesses (where n_{TP} = the number of target-present lineups used in the experiment being fit), a predicted target-present decision was made by randomly drawing 5 values from the lure distribution, randomly drawing 1 value from the target distribution, and then applying the decision rule described above. Once all n_{TP} trials were completed, one set of predicted target-present data had been generated. Similarly, for each of n_{TA} simulated witnesses (where n_{TA} = the number of target-absent lineups used in the experiment being fit), a predicted target-absent decision was made by randomly drawing 6 values from the lure distribution, and then applying the decision rule described above. Once all n_{TA} trials were completed, one set of predicted target-absent data had been generated. These predicted target-present and target-absent data constituted results from one simulated experiment. However, because the simulation involves stochastic processes, the predicted data from a single simulated experiment are noisy. Thus, for each iteration of a fit, predicted values were obtained by running 500 simulated experiments and then averaging the results. Those averaged results were compared to the observed data (computing a chi square goodness-of-fit statistic), at which point the parameters were adjusted and the next iteration of the fit occurred. Using *fminsearch* in MATLAB, the fitting process continued until the chi square was minimized.

Fitting the Model to the Palmer et al. (2013) Experimentally-Controlled Field Data. As noted in the main article, in a large-scale (n = 908) investigation into the relationship between confidence and accuracy (11), experimenters approached participants in parks and shopping malls and asked them to view a target person (the "perpetrator"). Approximately half the participants were tested using a target-present lineup and the other half using a target-absent lineup. Thus, the base rate of target-present lineups was known to be approximately 50%. We first fit the 5-parameter unequal-variance signal-detection model (parameters = μ_{Target}, σ_{Target}, c1, c2, and c3) to the *uncollapsed* data shown in Fig. 3A of the main article. This is how the model would ordinarily be fit to empirical data where it is known which trials involved target-present lineups and which involved target-absent lineups. For a given set of parameter values, the model generates predicted data separately for target-present and target-absent lineups. Thus, for each iteration of this fit, the predicted data were assessed by comparing the predicted values to the uncollapsed observed data in Fig. 3A. The results of the fit are shown in Table S4. The fit was very good, $\chi^2(7) = 4.25$. Unlike what is typically found in basic list memory studies, an equal-variance model turned out to be sufficient (i.e., σ_{target} did not differ appreciably from 1, thus $\sigma_{target} = \sigma_{lure}$).

We next fit the model to the *collapsed* experimentally-controlled field data shown in Fig. 3C of the main article. As noted in the main article, on each iteration of the fit, the predicted data were combined across target-present and target-absent lineups (thereby losing predicted information specific to lineup type) and the chi square goodness-of-fit statistic was computed by comparing the collapsed predicted data to the collapsed observed data from [11]. In other words, we fit the model to these data as if we were fitting it to police department field data, where target-present and target-absent lineups are combined. The model-fitting procedure assumed that target-present and target-absent lineups had been combined in equal proportion (i.e., it assumed equal base rates, which is known to be true of the experimentally-controlled field study), and the fit was very good, $\chi^2(1) = 0.34$ (see predicted values Fig. 3D-E). The results in Table S4 show that, once again, an equal-variance model turned out to be sufficient. Indeed, and remarkably, the estimated parameter values were nearly identical whether the model was fit to the uncollapsed data or to the collapsed data. These results indicate that the model is capable of recovering information about target-present and target-absent lineups even when the data are collapsed (albeit in known proportions) across lineup type. Even more remarkably, as described next, if an equal-variance model is assumed, the model can also recover the true underlying base rate when the target-present and target-absent lineups are combined in an unknown proportion.

Validating Base-Rate Signal-Detection Analyses. We tested the ability of the signal-detection model to recover the underlying base rate of target-present lineups using experimentally-controlled field data (11). For the base-rate analyses described here, we fixed $\sigma_{target} = \sigma_{lure} = 1$, and we added a target base-rate parameter to the model, so it's 5 parameters now were μ_{Target}, p_{Target}, c1, c2, and c3, where p_{Target} estimates $n_{TP} / (n_{TP} + n_{TA})$. When this model was fit to the collapsed data in Fig. 3C, the estimated value of p_{Target} was 0.51, very close to the true target-present base rate. Next, we mixed the target-present and target-absent data in varying proportions, fitting the signal-detection model each time to the collapsed data (aggregated across target-present and target-absent lineups) to see if it could recover the true base rate value. Indeed, the estimated value of p_{Target} very closely tracked the true underlying base rate (Fig. S1). These data suggest that, if the equal-variance model is assumed to also apply to the Houston field data, it can be used to accurately estimate the base-rate of target-present lineups in that study (i.e., it can be used to estimate this real-world base rate). As noted in the main article (and as described in more detail below), when the equal-variance model was fit to the Houston field data, the value of p_{Target} was estimated to be .35. This was true for the separate fits of the model to the blind simultaneous and blind sequential data. In other words, despite being independent fits, they yielded the same estimate of p_{Target}.

Fitting the Model to the Blind Houston Field Data. We first fit a 5-parameter unequal-variance signal-detection model (parameters = μ_{Target}, σ_{Target}, c1, c2, and c3) separately to the blind simultaneous and blind sequential data from the Houston field study. Because the true base rate of target-present lineups is unknown, we performed the fit three times, assuming a 25% base rate, 50% base rate, and 75% base rate. We also fit the model to the data combined across simultaneous and sequential lineups. The results of these fits are shown in Table S5. The chi-square goodness-of-fit statistics show that, as a general rule, the model fit very well. Moreover, a discriminability estimate (d_a) shows that, generally speaking, the simultaneous procedure outperformed the sequential procedure.

However, as noted in the main article, 65 witnesses reported that they (1) encountered a photo of the suspect before being presented with the photo lineup, (2) were under the influence of alcohol when they witnessed the crime, and/or (3) were not wearing their prescribed glasses during the crime. The results of these model fits on the reduced data set (eliminating these 65 witnesses) are shown in Table S6. The fits are still good but now the apparent d_a advantage for simultaneous lineups is limited to the low base rate condition and even slightly reverses at the high (75%) base rate condition.

As described in the main text, based on the results of the model fit to the experimentally-controlled field data (which suggested an equal-variance model, see Table S4), we next made the assumption that an equal-variance model ($\sigma_{target} = \sigma_{lure}$) also applies to the blind lineup data from the Houston field study, thereby allowing us to add a new parameter to the model. As noted above, the new parameter, p_{target}, provides an estimate of the base rate of target-present lineups. Model-based comparisons between the blind simultaneous and sequential lineups were performed by fitting the model to the simultaneous and sequential data concurrently, estimating the following 5 parameters for each lineup procedure: μ_{Target}, c1, c2, c3, and p_{Target}. These 5 parameters were allowed to differ for the simultaneous and sequential lineups, so there were 10 free parameters in all. There were 12 degrees of freedom in the data (6 per lineup type), so this was a 2-degree-of-freedom fit. The fit was very good, $\chi^2(2) = 1.35$, $p = .51$.

We next eliminated 3 parameters by constraining *c1*, *c2* and *c3* to be equal for the two lineup formats. The resulting change in the goodness-of-fit chi square was not significant, $\chi^2(3) = 1.94$, $p = .59$, indicating that these parameters did not differ for the simultaneous and sequential formats. Thus, *c1*, *c2*, and *c3* were constrained to be equal across lineup format for the subsequent fits. Next, we eliminated another parameter by constraining p_{Target} to be equal for the two lineup formats. Even when free to differ, the estimated value of this parameter was 0.35 for both lineup formats, so the goodness-of-fit chi square did not change at all when this constraint was added. Thus (obviously), the estimated value of p_{Target} did not differ for the simultaneous and sequential formats. Finally, we tried eliminating one additional parameter by constraining μ_{Target} to be equal for the two lineup formats. This constraint resulted in a far worse fit, $\chi^2(1) = 8.12$, $p =$

.004, indicating that the mean of the target distribution differed significantly for the simultaneous and sequential lineup formats. With all other parameters constrained to be equal, μ_{Target} was estimated to be 2.94 for the blind simultaneous lineups and 2.03 for the blind sequential lineups. In other words, simultaneous lineups were estimated to be diagnostically superior to sequential lineups (i.e., the ability to discriminate innocent from guilty suspects was higher for the simultaneous procedure). The estimated values of μ_{Target} fit reported in the main article (SIM = 2.87 vs. SEQ = 2.06) differed slightly only because, for that fit, *c1*, *c2*, and *c3* were allowed to differ for the two lineup formats. The results are essentially the same either way. However, as noted in the main article, 65 witnesses reported that they (1) encountered a photo of the suspect before being presented with the photo lineup, (2) were under the influence of alcohol when they witnessed the crime, and/or (3) were not wearing their prescribed glasses during the crime. When these witnesses are eliminated from the model fits, the difference in μ_{Target} still favored the simultaneous procedure (SIM = 2.67 vs. SEQ = 2.13), but the difference was no longer significant (p = .11).

Lap 1 vs. Lap 2 Choosing in the Blind Sequential Condition

When the sequential lineup is used in mock-crime laboratory studies, participants are typically permitted to view the photos only once (usually with the first ID being the one that counts). In actual practice, the police permit second viewings if the witness requests it. This was also true of the Houston Police Department field study. Investigators were instructed to allow witnesses to view sequential photo spreads a second time only if the witness requested to view the photos again. In other words, investigators did not give this option up front when giving the instructions. If the witness requested to view the photo spread or they requested to view a single photo again, the investigator showed the entire set of photos a second time, in the same order as the first showing. Thus, we can examine the data for witnesses who viewed the photos only once (1 lap) to see how the results compared to witnesses who viewed the photos more than once (2+ laps).

Investigators were asked to complete a survey item that asked how many times the witnesses viewed the photo spread. For 24 of the 161 blind sequential photo spreads, this information was not recorded, and we assumed that this was because only one viewing occurred. This counting rule resulted in 96 witnesses viewing the photos once and 65 viewing the photos more than once (59 twice, 5 three times, and 1 four times). In other words, 40% of the witnesses (65 out of 161) requested more than one viewing. For the analyses discussed next, we consider the data separately for those who viewed the photo spread once vs. those who viewed it more than once. The relevant frequency counts and proportions are presented in Table S7.

Responding was considerably more conservative for those who viewed the sequential photo spread only once compared to those who viewed it more than once. For example, for the Lap 1 witnesses, 45 out of 96 made no ID (.47), whereas for the Lap 2+ witnesses, only 14 out of 65 made no ID (.22), a difference that was highly significant, $\chi^2(1) = 10.72$, $p = .0011$. For those who made an ID, Lap 1 witnesses identified a suspect 26 times and identified a filler 25 times (.51 suspect IDs), whereas Lap 2+ witnesses identified a suspect 20 times and identified a filler 31 times (.39 suspect IDs). On the surface, it appears that the Lap 2+ witnesses were less accurate, but the fact that they were also more liberal in their responding means that the two numbers cannot be meaningfully compared to assess relative discriminability. Instead, a measure that is not influenced by response bias (e.g., d') is needed.

We fit the equal-variance signal-detection model, with p_{target} fixed at .35 (free parameters = μ_{target}, *c1*, *c2*, and *c3*), separately to the Lap 1 and Lap 2+ sequential Houston field data broken down by confidence (i.e., we fit the model to the frequency data shown in Table S7). With all parameters free to vary, the model fit the data well, $\chi^2(4) = 1.76$, $p = .779$. The optimal parameter estimates are shown in Table S8. If anything, μ_{target} (i.e., d') was actually higher, not lower, for the Lap 2+ sequential witnesses compared to the Lap 1 sequential witnesses. However, the difference was not remotely close to being significant, $\chi^2(1) = 0.11$, $p = .744$. In other words, the large apparent difference in μ_{target} is an illusion because the fit is scarcely affected by constraining the parameters to be equal (in which case the estimated value of μ_{target} was 2.06). Thus, with these data, no difference in discriminability was detected between the Lap 1 and Lap 2+ witnesses. However, the other parameter estimates suggest that the Lap 2+ witnesses may have been much more liberal about making an ID in general (i.e., *c1* is noticeably lower for the Lap 2+ witnesses) while being more conservative about making a high-confidence ID (i.e., *c3* is noticeably higher for the Lap 2+ witnesses). Indeed, when the *c1*, *c2*, and *c3* parameters were constrained to be equal across the Lap 1 and Lap 2+ sequential witnesses, the fit was significantly worse, $\chi^2(3) = 19.64$, $p < .001$. This result indicates that the main difference between the Lap 1 and Lap 2+ witnesses is in their placement of the various decision criteria, not in their ability to discriminate innocent from guilty suspects. It also indicates that the trends favoring the simultaneous procedure discussed earlier did not arise merely because the Lap 2+ witnesses were included in the sequential analysis.

Analyses of Blinded Lineups

The data from the blinded simultaneous and sequential conditions were analyzed in the same way that the data from the blind conditions were analyzed in the main article. Here, we first describe the basic empirical trends in the data and then present signal-detection analyses of the data. The results of these analyses are all similar to the results from the analyses of the blind lineups, but various complexities involved with this data set (unbalanced corroborating evidence for simultaneous and sequential formats prior to any IDs being made, unfair lineups, and poor signal-detection model fits) warrant caution in interpreting the results. We present the results for the sake of completeness and to underscore the fact that the results are similar the results presented above and therefore offer no reason to question the interpretation of the blind lineup data.

Confidence in Suspect IDs and Filler IDs from Blinded Lineups. For the blinded lineups (collapsed across simultaneous and sequential), the results were similar to the results observed for the blind lineups shown in Figure 1A-D. Suspect IDs and filler IDs occurred with approximately equal frequency, whereas no IDs occurred with somewhat greater frequency (Fig. S2A). As with the blind data, most filler IDs were made with low confidence and most suspect IDs were made with high confidence (Fig. S2B). In other words, the proportion of IDs for each level of confidence that were suspect IDs – that is, suspect IDs / (suspect IDs + filler IDs) – increased dramatically with confidence (Fig. S2C). For suspect IDs, the proportion of cases with corroborating evidence of guilt also increased as confidence in the ID increased (Fig. S2D).

Fitting the Model to the Blinded Houston Lineup Data. As with the blind data, we first fit a 5-parameter unequal-variance signal-detection model (parameters = μ_{Target}, σ_{Target}, c1, c2, and c3) separately to the blinded simultaneous and blinded sequential data from the Houston field study. Because the true base rate of target-present lineups is unknown, we performed the fits three times, assuming a 25% base rate, 50% base rate or 75% base rate. We also fit the model to the data combined across simultaneous and sequential lineups. The results of these fits are shown in Table S9. The chi-square goodness-of-fit statistics show that, as a general rule, the model fit the simultaneous data very well but provided a poor fit to the sequential data. It is not clear why the data from the blinded sequential condition are so poorly fit by the model, but the parameter estimates for the fits described below generally seem sensible and interpretable despite the poor fit to that condition.

Base Rate Estimates. Based on the results of the model fit to the experimentally-controlled field data (which suggested an equal-variance model), and again following the same steps we followed for the blind lineup data, we next made the assumption that an equal-variance model ($\sigma_{target} = \sigma_{lure}$) also applies to the blinded Houston field data. Making this assumption allowed us to obtain an estimate of the base rate of target-present lineups (p_{Target}), separately for the simultaneous and sequential lineups. Model-based comparisons between the blinded simultaneous and sequential lineups were performed by fitting the data from both lineup procedures concurrently.

The first model included the following parameters: μ_{Target}, c1, c2, c3, and p_{Target}, which were all allowed to differ for simultaneous and sequential lineups (thus, there were 10 parameters in all). Mainly because the equal-variance model was not able to accurately characterize the blinded sequential data (as discussed above), the simultaneous fit of the full model to the blinded simultaneous and sequential data was poor, $\chi^2(2) = 12.09$, $p = .002$. We next investigated the effect of constraining the parameters to be equal across lineup formats. Unlike the model fits to the blind simultaneous and sequential data, when the confidence criteria c1, c2 and c3 were constrained to be equal across the blinded simultaneous and sequential lineup formats, the fit was worse, $\chi^2(3) = 7.85$, $p = .049$. The difference was barely significant, and given the number of significance tests performed, it might be regarded as a Type I error. Thus, we conservatively constrained the confidence parameters to be equal despite this barely significant result.

Next, p_{Target} was constrained to be equal for the two lineup formats. Unlike the model fits to the blind simultaneous and sequential data, the fit was significantly worse, $\chi^2(6) = 32.94$, $p < .0001$. In other words, p_{Target} was significantly greater for blinded sequential lineups compared to blinded simultaneous lineups (.23 for simultaneous and .43 for sequential). Remarkably, the fact the estimated value of p_{Target} was significantly greater for blinded sequential lineups is consistent with the entirely independent corroborating evidence estimates discussed earlier, which also suggest that more guilty suspects ended up in blinded sequential lineups than in blinded simultaneous lineups. In other words, two entirely independent analyses converge on the notion that an unusually high number of guilty suspects ended up in the blinded sequential condition (for reasons that are unclear).

Model-Based Confidence-Accuracy Analyses. The best-fitting unequal-variance model to the combined data (Table S9) yielded the confidence-accuracy predictions shown in Fig. S3A (averaged across simultaneous and sequential lineups), which closely resemble the corresponding predictions for the blind data shown in Fig. 4A of the main article. Similarly, the confidence-accuracy relationship predicted by this best-fitting equal-variance model (again averaged across simultaneous and sequential lineups) with p_{Target} fixed at the estimated values described above exhibits a strong relationship between the confidence associated with a suspect ID and the accuracy of that ID (Fig. S3B), just as was true of the lineups from the blind condition (Fig. 4B). Thus, with regard to the relationship between confidence and accuracy, the results are very similar for the blind and blinded lineups.

Discriminability for Blinded Simultaneous and Sequential Lineups. For the next fit, we compared an equal-variance model with $c1$, $c2$, $c3$, and μ_{Target} constrained to be equal across lineup format (with p_{Target} free to vary) to an equal-variance model with $c1$, $c2$, and $c3$

constrained to be equal across lineup format (with μ_{Target} and p_{Target} free to vary). This test asks whether discriminability differs for the blinded simultaneous and sequential lineup formats and is therefore a key test. The result indicated a significant simultaneous advantage, $\chi^2(1) = 4.31$, $p = .038$. The estimates of μ_{Target} for the blinded simultaneous and sequential lineups were 3.10 and 2.27, respectively. Thus, as with the fits to the blind lineups, the fits to the blinded lineup data suggest a discriminability advantage for simultaneous lineups.

Based on the values shown in Table S9, the argument could be made that the *c1* confidence parameter should be free to differ when comparing μ_{Target} for the blinded simultaneous and sequential lineups (whereas they were constrained to be equal across lineup format in the test described above). When that is done, the estimate for μ_{Target} still favors the simultaneous procedure (SIM = 2.90, SEQ = 2.48), but the difference is now only marginally significant, $\chi^2(1) = 2.88$, $p = .090$.

Corroborating Evidence of Guilt Analyses for Simultaneous vs. Sequential Lineups. Next, we describe a non-model-based comparison of simultaneous and sequential lineups based on corroborating evidence of guilt. Directly evaluating lineup performance based on evidence of guilt associated with identified suspects from blinded simultaneous and sequential lineups, as we did for the blind conditions, would not be useful here because it would obviously be higher for sequential lineups. However, some evidence of the relative diagnostic performance of the two lineup types can be obtained by examining whether corroborating evidence of guilt is higher for suspect IDs compared to no IDs within each lineup type. To the extent that suspect IDs for a given lineup type are associated with more corroborating evidence of suspect guilt compared to no IDs, eyewitness decisions would be diagnostic of guilt. In other words, this difference score (i.e., corroborating evidence against identified suspects minus corroborating evidence against suspects in lineups where no ID was made) provides an independent estimate of the ability to discriminate innocent from guilty suspects (an estimate that was directly quantified in the equal-variance signal-detection model fits summarized above).

For blinded simultaneous lineups, the proportion of cases with corroborating evidence against the suspect given that no ID was made was .54, whereas the proportion of cases with corroborating evidence against the suspect given that a suspect ID was made was .75, an increase that was significant, $\chi^2(1) = 5.00$, $p = .025$. Thus, by this measure, suspect IDs from the blinded simultaneous lineups were diagnostic of guilt. By contrast, for blinded sequential lineups, the proportion of cases with corroborating evidence against the suspect given that no ID was made was .93, whereas the proportion of cases with corroborating evidence against the suspect given that a suspect ID was made was .90. Thus, there was no indication of diagnosticity associated with suspect IDs in blinded sequential lineups. However, the high level of corroborating evidence against suspects across all blinded sequential lineups (0.91) would have made the detection of an increase associated with suspect IDs somewhat difficult. Thus, while these data are consistent with a simultaneous superiority effect (as was true of the equal-variance signal-detection fits described above), the findings are less compelling than those from the blind lineup analyses. On balance, it seems fair to conclude that for both the blind and the blinded lineups, (1) there is no evidence whatsoever for a sequential superiority effect, (2) there are many indications of a simultaneous superiority effect, and (3) the significance of the simultaneous superiority effect depends on the details of how the data are analyzed. In other words, there is no evidence of a sequential superiority effect, and, if anything, the data point to a simultaneous advantage (as would be predicted by recent lab-based ROC analyses).

Recoded Corroborating Evidence

For the blind and blinded simultaneous and sequential lineups, five independent raters coded information in the fill-in text boxes where the investigating officer listed what the corroborating evidence was whenever it was deemed to be present. For this work, the coders included one of the authors (William Wells), two faculty members who have significant experience conducting research with criminal investigators, a Ph.D. student who has research experience working with criminal investigators, and a former Houston Police Department robbery investigator (who has his PhD and has done work with inter-rater reliability checks). During this process, the raters identified situations in which the investigator indicated there was corroborating evidence but the team's coding of the text box information suggested it should not be counted as corroborating evidence. The raters also identified situations in which investigators included information in the fill-in text box but nevertheless did not check the box to indicate there was corroborating evidence.

For the independent coding, we used the coding that was recommended by 3 or more of the 5 coders. That is, if 3 raters said the text box information should be counted as corroborating evidence, then we counted it as such. The relationship between corroborating evidence and confidence in a suspect ID is shown in Fig. S4A. The results are similar to the results observed for the blind lineups shown in Fig. 1D. The results for the recoded blinded lineups are shown in Fig. S4B. Again, the relationship remains essentially the same as for the originally coded data.

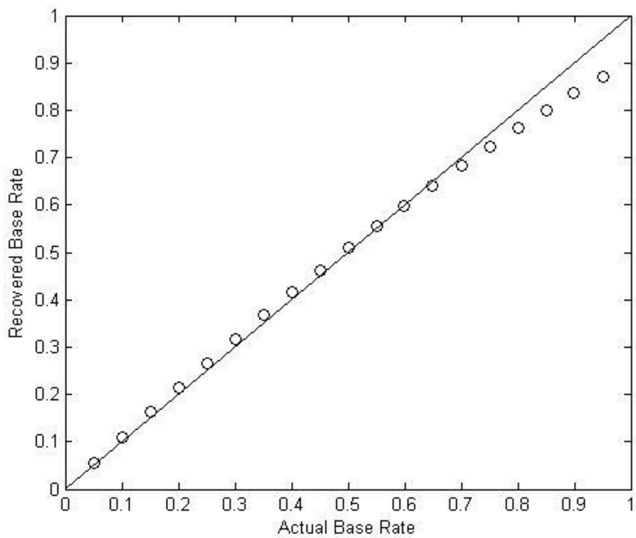

Fig. S1. Estimated value of p_{Target} (y-axis) as a function of the true base rate of target-present lineups (x-axis). The estimated values were based on a 5-parameter fit of the equal-variance signal-detection model to data from the experimentally-controlled field study (10) that had been aggregated across target-present and target-absent lineups in different proportions across fits. For example, at the lower left, the data were aggregated in such a way that 5% of the data were from target-present lineups and 95% from target-absent lineups. Although some inaccuracy is apparent at very high base rates, for the most part the model is able to accurately recover the underlying base rate from data that have been aggregated across target-present and target-absent lineups. Thus, all the model sees are the number of suspect IDs and filler IDs (across 3 levels of confidence) and no IDs. It has no information about how many of each came from target-present lineups or target-absent lineups. Even so, it can recover the true base rate with remarkable accuracy, raising the possibility that it might be able to do the same with police department field data (which are also aggregated across target-present and target-absent lineups in unknown proportion).

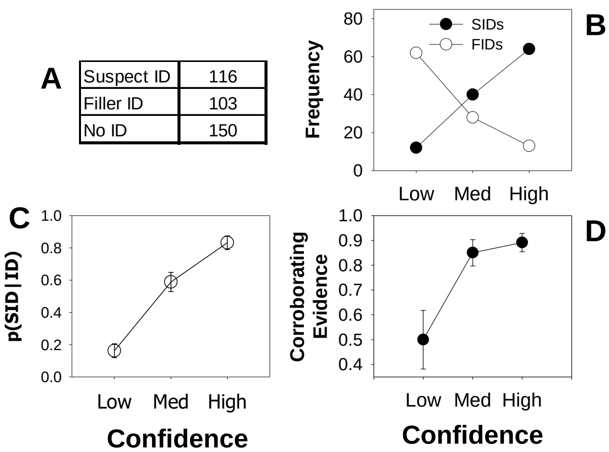

Fig. S2. (**A**) Frequency counts of eyewitness decision outcomes in the Houston field study for 194 blinded simultaneous and 175 blinded sequential lineups combined. (**B**) Frequency of Suspect IDs (SIDs) and Filler IDs (FIDs) in **A** exhibited opposite trends as a function of confidence (low, medium or high), $\chi^2(2) = 69.2$, $p < .0001$. (**C**) For IDs made to a suspect or filler, the probability that it was a suspect ID increased dramatically with confidence. (**D**) Proportion of suspect IDs rated by the investigating officer as having independent corroborating evidence of guilt increased with confidence in the ID. According to a Cochran–Armitage trend test, the effect was significant (one-tailed), $Z = 2.80$, $p < .01$. In all cases, error bars are standard errors.

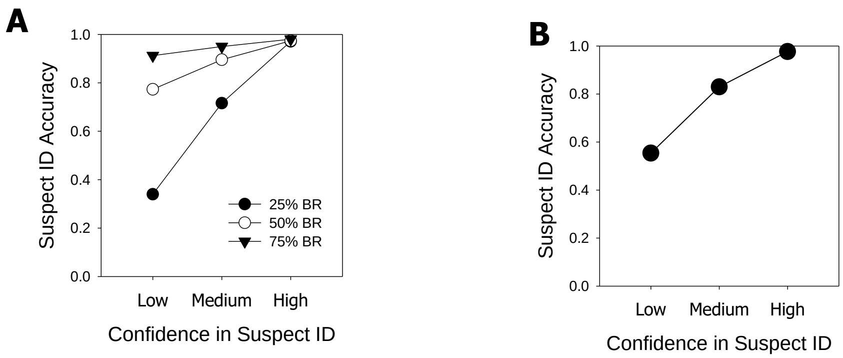

Fig. S3. (**A**) Model-based estimates of the posterior probability of guilt associated with suspects identified from blinded lineups in the Houston field study for three different assumptions regarding base rates (BR). The estimates are averaged across simultaneous and sequential lineups because the results were nearly identical for the two lineup formats. (**B**) Model-based estimate of the posterior probability of guilt associated with suspects identified from blinded lineups (averaged across simultaneous and sequential lineups) in the Houston field study assuming an equal-variance signal-detection model (as suggested by fits to the experimentally-controlled field data) and including "target-present base rate" as a free parameter (estimated to be .23 for simultaneous lineups and .43 for sequential lineups).

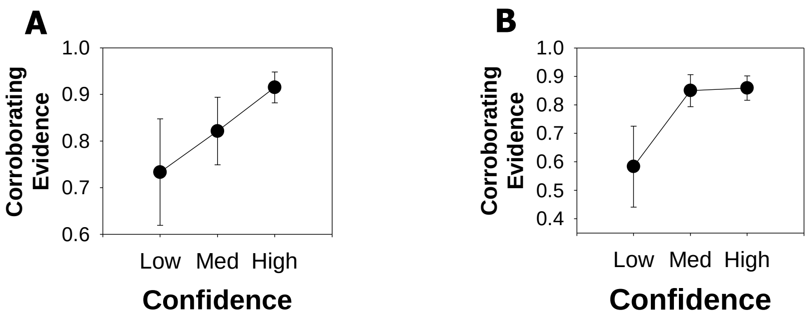

Fig. S4. Recoded corroborating evidence analyses. (**A**) For the blind lineups, proportion of suspect IDs rated by the investigating officer as having independent corroborating evidence of guilt increased with confidence in the ID. According to a Cochran–Armitage trend test, the effect was significant (one-tailed), $Z = 2.71$, $p < .01$. (**B**) For the blinded lineups, proportion of suspect IDs rated by the investigating officer as having independent corroborating evidence of guilt increased with confidence in the ID. According to a Cochran–Armitage trend test, the effect was marginally significant (one-tailed), $Z = 1.42$, $p = .077$. In all cases, error bars are standard errors.

Table S1. Frequency counts and proportions of eyewitness decisions in the blind and blinded simultaneous and sequential conditions.

Frequency Counts

Blind Simultaneous	High	Med	Low	No ID
Suspect ID	49	13	6	71
Filler ID	9	15	24	

Blind Sequential	High	Med	Low	No ID
Suspect ID	22	15	9	59
Filler ID	8	18	30	

Blinded Simultaneous	High	Med	Low	No ID
Suspect ID	30	11	7	96
Filler ID	7	17	26	

Blinded Sequential	High	Med	Low	No ID
Suspect ID	34	29	5	54
Filler ID	6	11	36	

Proportions

Blind Simultaneous	High	Med	Low	No ID
Suspect ID	0.26	0.07	0.03	0.38
Filler ID	0.05	0.08	0.13	

Blind Sequential	High	Med	Low	No ID
Suspect ID	0.14	0.09	0.06	0.37
Filler ID	0.05	0.11	0.19	

Blinded Simultaneous	High	Med	Low	No ID
Suspect ID	0.15	0.06	0.04	0.49
Filler ID	0.04	0.09	0.13	

Blinded Sequential	High	Med	Low	No ID
Suspect ID	0.19	0.17	0.03	0.31
Filler ID	0.03	0.06	0.21	

Table S2. Differences and similarities between treatment groups. Because 11 post hoc statistical tests were performed, the Bonferroni-adjusted alpha level is .05 / 11 = .0045. Significant differences were observed for Interpreter Used, Corroborating Evidence, and Location in which the lineup was administered.

	Photo Spread Administration Method				
Case Characteristics	Blinded Simultaneous N = 194	Blinded Sequential N = 175	Blind Simultaneous N = 187	Blind Sequential N = 161	*Statistical Test*
Interpreter Used	6 (3%)	13 (7%)	16 (9%)	23 (14%)	χ^2=14.99 *p=.002*
Corroborating Evidence	121 (62%)	160 (91%)	130 (70%)	105 (65%)	χ^2=45.90 *p=.000*
Witness Under Influence	17 (9%)	13 (7%)	3 (2%)	13 (8%)	χ^2=10.03 *p=.018*
Witness Saw Photo	21 (11%)	16 (9%)	23 (12%)	5 (3%)	χ^2=10.04 *p=.018*
Witness not Wearing His/Her Glasses	7 (4%)	10 (6%)	6 (3%)	17 (11%)	χ^2=11.01 *p=.012*
Location					χ^2= 52.94 *p=.000*
Police Facility[b]	46 (24%)	34 (20%)	86 (46%)	51 (32%)	
Witness Home	50 (26%)	61 (35%)	56 (30%)	47 (29%)	
Witness Work/school	48 (25%)	46 (26%)	30 (16%)	31 (19%)	
Public Setting	40 (21%)	31 (18%)	11 (6%)	26 (16%)	
Other	10 (5%)	3 (2%)	4 (2%)	6 (4%)	
Serial Investigation	82 (42%)	77 (44%)	76 (41%)	50 (31%)	χ^2=6.98 *p=.073*
Victim	166 (86%)	157 (90%)	166 (89%)	144 (89%)	χ^2=1.98 *p=.576*
Suspect in Position #1[a]	28 (14%)	19 (11%)	18 (10%)	13 (8%)	χ^2=4.14 *p=.247*
Good Viewing Opportunity	65 (34%)	68 (39%)	62 (33%)	55 (34%)	χ^2=1.75 *p=.630*
Weapon Used	136 (70%)	118 (67%)	123 (66%)	109 (68%)	χ^2=0.83 *p=.841*

Table S3. Proportion of mock witnesses who identified the suspect.

	Blinded		**Blind**	
	SIM	**SEQ**	**SIM**	**SEQ**
1	0.02	0.02	0.06	0.02
2	0.02	0.1	0.08	0.06
3	0.08	0.1	0.08	0.06
4	0.14	0.12	0.08	0.1
5	0.16	0.16	0.1	0.1
6	0.18	0.22	0.1	0.12
7	0.2	0.27	0.12	0.16
8	0.22	0.31	0.14	0.2
9	0.31	0.31	0.16	0.22
10	0.31	0.33	0.16	0.27
11	0.35	0.39	0.16	0.29
12	0.35	0.41	0.2	0.29
13	0.39	0.41	0.22	0.31
14	0.44	0.45	0.38	0.35
15	0.73	0.92	0.43	0.47
mean	**0.26**	**0.30**	**0.16**	**0.20**
sd	0.18	0.22	0.11	0.13

Table S4. Parameter values of the signal detection model that minimize the chi-square goodness-of-fit statistic when the model is fit to the uncollapsed data shown in Fig. 3A and to the collapsed data shown in Fig. 3C. The model fit to the collapsed data assumed that target-present and target-absent lineups were combined in equal proportion (i.e., the base-rate parameter, p_{Target}, was fixed to be .50). The similarity of the parameter estimates is striking.

Commented [LM1]: Sure is!

Parameters	Palmer et al. (2013)	Palmer et al. (2013) collapsed
c3	2.57	2.58
c2	2.04	2.04
c1	1.51	1.51
μ_{Target}	1.81	1.83
σ_{Target}	0.98	0.97
χ^2	4.25	0.34
df	7	1
p	0.751	0.560

Table S5. Parameter values of the unequal-variance signal detection model that minimize the chi-square goodness-of-fit statistic when the model is fit to the full blind Houston field data set assuming target-present based rates of 25%, 50% and 75%. Note that in order for the model to provide an acceptable fit, the estimated mean of the target distribution (μ_{Target}) decreased and the estimated standard deviation of the target distribution (σ_{Target}) increased as the base rate of target-present lineups increased. The suspect ID accuracy scores shown in Fig. 4A of the main article were derived from the best-fitting parameter values for the combined fits (Simultaneous + Sequential). For all fits, the mean and standard deviation of the filler distribution were set to 1 and 0, respectively. The bottom row shows a discriminability estimate, $d_a = \mu_{Target}$ / sqrt[(1^2 + σ_{Target}^2)/2]. d_a takes into account the unequal variance of the underlying distributions and is equal to *d'* when the two variances are equal (i.e., when σ_{Target} = 1).

	Simultaneous			Sequential			Combined		
Parameters	25%	50%	75%	25%	50%	75%	25%	50%	75%
c3	2.26	2.26	2.26	2.29	2.3	2.3	2.27	2.28	2.28
c2	1.74	1.74	1.74	1.72	1.7	1.7	1.74	1.72	1.72
c1	1.31	1.34	1.34	1.2	1.21	1.21	1.25	1.28	1.27
μ_{target}	2.54	2.33	1.30	2.36	1.46	0.84	2.53	1.88	1.04
σ_{target}	0.28	1.85	2.45	0.56	1.45	1.68	0.40	1.69	2.08
χ^2	4.13	1.40	1.26	0.00	0.86	0.80	1.59	2.06	1.84
p	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
d_a	3.46	1.57	0.69	2.91	1.17	0.61	3.32	1.35	0.64

Table S6. Parameter values of the unequal-variance signal detection model that minimize the chi-square goodness-of-fit statistic when the model is fit to the reduced blind Houston field data set (eliminating 65 witnesses based on their questionnaire responses about encountered a photo of the suspect, being under the influence of alcohol, and/or not wearing their prescribed glasses) assuming target-present based rates of 25%, 50% and 75%. For all fits, the mean and standard deviation of the filler distribution were set to 1 and 0, respectively. The bottom row again shows a discriminability estimate, $d_a = \mu_{Target}$ / sqrt[($1^2 + \sigma_{Target}{}^2$)/2]. Note that the one case where the sequential procedure yields a slight advantage is in the 75% target-present base rate condition.

	Simultaneous			Sequential			Combined		
Parameters	25%	50%	75%	25%	50%	75%	25%	50%	75%
c3	2.44	2.44	2.45	2.19	2.20	2.21	2.31	2.32	2.32
c2	1.84	1.81	1.81	1.73	1.69	1.68	1.79	1.75	1.75
c1	1.36	1.38	1.38	1.17	1.18	1.18	1.27	1.29	1.28
μ_{target}	2.63	2.01	0.95	2.32	1.53	0.86	2.50	1.76	0.91
σ_{target}	0.25	2.10	2.59	0.35	1.49	1.76	0.33	1.77	2.14
χ^2	0.47	1.47	1.26	0.00	1.85	1.75	0.21	3.31	2.98
p	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
d_a	3.60	1.22	0.48	3.10	1.21	0.60	3.36	1.22	0.54

Table S7. Frequency counts and proportions of eyewitness decisions in the blind sequential condition, separated according to witnesses who viewed the photo spread only once and those who viewed it more than once.

Frequency Counts

Lap 1 Sequential	High	Med	Low	No ID
Suspect ID	14	7	5	45
Filler ID	7	7	11	

Lap 2+ Sequential	High	Med	Low	No ID
Suspect ID	8	8	4	14
Filler ID	1	11	19	

Proportions

Lap 1 Sequential	High	Med	Low	No ID
Suspect ID	0.15	0.07	0.05	0.47
Filler ID	0.07	0.07	0.11	

Lap 2+ Sequential	High	Med	Low	No ID
Suspect ID	0.12	0.12	0.06	0.22
Filler ID	0.02	0.17	0.29	

Table S8. Parameter values of the signal detection model that minimize the chi-square goodness-of-fit statistic when the model is fit simultaneous to the Lap 1 and Lap 2+ blind sequential frequency data shown in Table S7. For this fit., the base-rate parameter, p_{Target}, was fixed to be .35.

Parameters	Lap 1 Sequential	Lap 2+ Sequential
c3	2.19	2.72
c2	1.78	1.67
c1	1.43	0.93
μ_{Target}	1.97	2.51
σ_{Target}	1	1
χ^2	1.76	
df	4	
p	0.779	

Table S9. Parameter values of the unequal-variance signal detection model that minimize the chi-square goodness-of-fit statistic when the model is fit to the blinded Houston field data assuming target-present based rates of 25%, 50% and 75%. Note that in order for the model to provide an acceptable fit, the estimated mean of the target distribution (μ_{Target}) decreased and the estimated standard deviation of the target distribution (σ_{Target}) increased as the base rate of target-present lineups increased (trends that are also evident in the fits to the blind data in Table S5). The suspect ID accuracy scores shown in Fig. S4A were derived from the best-fitting parameter values for the combined fits (Simultaneous + Sequential). For all fits, the mean and standard deviation of the filler distribution was set to 1 and 0, respectively.

	Simultaneous			Sequential			Combined		
Parameters	25%	50%	75%	25%	50%	75%	25%	50%	75%
c3	2.40	2.40	2.40	2.43	2.52	2.54	2.43	2.46	2.47
c2	1.85	1.84	1.84	1.85	1.82	1.77	1.87	1.81	1.81
c1	1.44	1.44	1.43	1.18	1.22	1.20	1.32	1.34	1.33
μ_{target}	2.71	1.21	0.33	2.44	2.26	1.49	2.47	1.85	1.05
σ_{target}	1.15	2.32	2.52	0.05	0.99	1.62	0.10	1.52	1.94
χ^2	0.24	0.35	0.22	9.25	13.00	14.75	0.73	11.42	11.27
p	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75