Project No: 800010140

Award Number: 2018-R2-CX-0030

Title: Testing a 'not sure' instruction to reduce the harmful impact of system and estimator variables on lineup identification accuracy

Principal Investigator:

Steve Charman

Professor of Legal psychology

charmans@fiu.edu

Florida International University, Department of Psychology

11200 SW 8th St., DM 256

Miami, FL 33199

Award recipient organization:

Florida International University

11200 SW 8th St.

Miami, FL 33199

Project period: 01/01/2019 to 08/31/2021

Amount: $183,059.00

Summary of the project

Eyewitness lineup identification accuracy is affected by numerous variables, including those that are under the control of the legal system, called system variables (e.g., pre-lineup instructions), and those that are not under the control of the legal system, called estimator variables (e.g., the race of the perpetrator). One of the ultimate goals of eyewitness researchers is to develop procedures that (a) minimize false identifications caused by system and estimator variables (while minimizing any decrease in correct identifications), and (b) require few resources for law enforcement to enact. The current project tested the effectiveness of a system variable—providing witnesses with an explicit 'not sure' instruction before viewing the lineup— that potentially meets both of these criteria. Furthermore, a 'not sure' instruction may act as a prophylactic against the harmful effects of system and estimator variables known to inflate false identifications. The specific objective of the proposed research was to test the effectiveness of this 'not sure' instruction. The pilot study (n = 322) examined whether variations in the administration of the 'not sure' instruction (i.e., whether it was described as a 'not sure' option or a 'don't know' option; whether the option was accompanied by a corresponding pre-lineup instruction regarding the option or not) affected identification decision outcomes; results indicated that the instruction was equally effective across these variations. The main two studies examined whether the 'not sure' option would restore any loss in discriminability associated with a system variable (the appearance-change instruction; Study 1) and an estimator variable (the race of the perpetrator relative to the race of the witness: Study 2). Study 1 participants (community members recruited via online research platform Prime Panels; n = 2137) viewed one of four mock crime videos and were randomly assigned (a) to receive either the 'not sure' instruction or not, (b) to receive the appearance-change instruction or not, and (c) to view a

target-present or target-absent lineup. Results indicated that the 'not sure' option both (a) failed to improve eyewitness decision-making (as assessed via changes in empirical discriminability), and (b) failed to reduce the increase in choosing caused by the appearance-change instruction. Study 2 participants (community members, either Black or White; n = 2097) viewed one of four mock crime videos and were randomly assigned (a) to view either a same-race or cross-race perpetrator, (b) to receive a 'not sure' instruction/option or not, and (c) to view either a target-present or target-absent lineup. Results indicated that although witnesses exhibited lower discriminability when making cross-race (cf. same-race) identifications, the 'not sure' option failed to restore that discriminability. In summary, the 'not sure' option failed to improve eyewitness decision-making overall, and failed to act as a prophylactic against the harmful effects of the appearance-change instruction or of cross-race identifications.

Major goals and objectives

To serve the overarching purpose of improving the administration of justice, this research had as its main goal to help develop a lineup procedure that (a) maximizes eyewitness identification accuracy by reducing false identifications (ideally without a concomitant reduction in correct identifications); (b) can act as a prophylactic against the harmful impact of system and estimator variables known to inflate false identifications; and (c) requires few resources to implement.

The specific objectives of the proposed research were to examine whether providing an explicit 'not sure' instruction to eyewitnesses (a) improves overall lineup identification accuracy by improving empirical discriminability, and/or (b) acts as a prophylactic by reducing the harmful impact of other variables (the appearance-change instruction; cross-race identifications).

Research questions

The project had the following 3 research questions:

(1) Does an explicit 'not sure' instruction increase eyewitness accuracy overall (**all studies)**? More specifically, does it reduce false identifications without reducing (or with a smaller reduction in) correct identifications, thereby increasing empirical discriminability?

(2) Does the effectiveness of an explicit 'not sure' instruction depend on methodological variations in its administration **(Pilot study**)? Specifically, does the wording of the option ('not sure' vs. 'don't know') or the presence of a pre-lineup instruction regarding the option affect eyewitness decision-making?

(3) Does the 'not sure' instruction reduce the harmful impact of other variables known to harm identification accuracy (**Studies 1-2)**? Specifically, the appearance-change instruction, a common pre-lineup instruction often given to witnesses, has been shown to increase false identifications without a concomitant increase in correct identifications (e.g., Charman & Wells, 2007; Molinaro et al., 2013). **Study 1** examined whether a 'not sure' option would ameliorate the deleterious effect of the appearance-change instruction. In addition, much research has shown that identification accuracy is harmed when witnesses make cross-race, as opposed to same-race, identifications (e.g., Meissner & Brigham, 2001). **Study 2** examined whether a 'not sure' option would ameliorate the deleterious effects of cross-race identifications.

Research design, methods, analytical and data analysis techniques

Pilot study

The pilot study conformed to a 2 (target status: target-present vs. target-absent) X 3 (instruction: 'don't know' vs. 'not sure' vs. none) X 2 (pre-lineup instruction regarding response option: given vs. not given) between-subjects design. (Note that pre-lineup instruction could not

be manipulated for participants who did not actually receive the option; consequently, the total number of conditions was 2*3*2 – 2 = 10). Participants were randomly assigned to all conditions.

Participants (undergraduate students: n = 322) viewed one of four mock crimes (carjacking, store theft, graffiti, or bike theft), were either given a pre-lineup instruction regarding the withholding response option that would be presented to them or not, viewed a target-present or target-absent lineup, were either presented with an option to withhold their decision ('not sure,' don't know,') or not, and made an identification decision and provided their confidence.

To determine if a withholding response affected responding, a series of chi-square tests were conducted comparing target-present hit rates and target-absent choosing rates of witnesses who received a withholding response option ('not sure' and 'don't know' options combined) to witnesses who did not receive a withholding option. To determine whether the wording of the response option affected responding, a series of chi-square tests were conducted comparing target-present hit rates and target-absent choosing rates across response option conditions ('not sure' vs. 'don't know). To determine whether the presence of a pre-lineup instruction regarding the withholding option affected responding, a series of chi-square tests were conducted comparing target-present hit rates and target-absent choosing rates across pre-lineup instruction condition (pre-lineup instruction vs. no pre-lineup instruction).

Study 1

Study 1 conformed to a 2 (target status: target-present vs. target-absent) X 2 (response option: 'not sure' vs. none) X 2 (appearance-change instruction: present vs. absent) between-subjects design. Participants were randomly assigned to all conditions.

Participants (recruited online via Prime Panels: $n = 2137$) viewed one of four mock crimes (carjacking, store theft, graffiti, or bike theft), were either given the appearance-change instruction (ACI) or not, were presented with either a target-present or target-absent lineup, were either given the option of responding 'not sure' (accompanied by a corresponding pre-lineup instruction) or not, and made an identification decision and provided their confidence.

To ensure participants were responding at above-chance levels, I examined for each target (a) whether the correct identification rate among target-present choosers was greater than expected by chance (i.e., 1/6); and (b) whether the hit rate was greater than the false identification rate.

To determine if the ACI and/or the 'not sure' option affected decision-making, I conducted a series of logistic regressions predicting (a) hits, (b) false alarms, and (c) overall choosing as a function of ACI condition, 'not sure' condition, and the ACI x 'not sure' interaction. I also conducted ROC analyses to examine whether the ACI and/or the 'not sure' option affected empirical discriminability.

Study 2

Study 2 conformed to a 2 (target status: target-present vs. target-absent) X 2 (response option: 'not sure' vs. none) X 2 (perpetrator race: cross-race or same-race) between-subjects design. Participants were randomly assigned to all conditions.

Participants (Black and White, recruited online via Prolific: $n = 2097$) viewed one of four mock crimes (all of which depicted a man stealing a laptop from a woman) in which they viewed either a same-race perpetrator or a cross-race perpetrator, were presented with either a target-present or target-absent lineup, were either given the option of responding 'not sure'

(accompanied by a pre-lineup instruction) or not, and made an identification decision and provided their confidence.

To ensure participants were responding at above-chance levels, I examined for each target (a) whether the correct identification rate among choosers was greater than expected by chance (i.e., 1/6); and (b) whether the hit rate was greater than the false identification rate.

To determine if the race of the perpetrator relative to the race of the witness (i.e., same-race or cross-race) and/or the 'not sure' option affected decision-making, I conducted a series of logistic regressions predicting (a) hits, (b) false alarms, and (c) overall choosing as a function of perpetrator race condition, 'not sure' condition, and the perpetrator race x 'not sure' interaction. I also conducted ROC analyses to examine whether the race of the perpetrator relative to the race of the witness and/or the 'not sure' option affected empirical discriminability.

Expected applicability of the research

Very few studies have examined the effects of providing witnesses with an explicit 'not sure' option on their identification decisions; those studies that have been conducted have shown conflicting findings regarding the option's efficacy. Results from the current studies provide further evidence that the 'not sure' option fails to either improve identification outcomes on its own, or to ameliorate the harmful impact of the appearance-change instruction or cross-race identifications. That said, the 'not sure' option also did not harm outcomes. Consequently, two applied conclusions can be drawn from this research. First, these null results suggest that real-world identifications obtained in the absence of a 'not sure' option are just as reliable as real-world identifications obtained in the presence of a 'not sure' option. Second, lab-based studies rarely provide witnesses with a 'not sure' option, whereas real-world witnesses are able to respond 'not sure.' That a 'not sure' option failed to affect identification outcomes thus suggests

that we are able to generalize lab findings to the real world despite this methodological difference in how witnesses are able to respond.

Furthermore, the 'not sure' option did reduce choosing overall (in the pilot study; for cross-race identifications in Study 2), and is thus consistent with a recent National Academy of Sciences recommendation for "a rigorous exploration of methods that can lead to more conservative responding (such as witness instructions) but do not compromise discriminability." Because the 'not sure' option is a witness instruction that leads to more conservative responding and doesn't compromise discriminability, it is potentially useful as a pre-lineup instruction that could be administered to witnesses in the real-world.

Changes in approach from original design and reason for change

While running this project, two major changes in methodology from the original design were required. First, due to quarantining required as a result of the Covid-19 pandemic, my original plan to collect data from community members around Miami for Studies 1 and 2 was unfeasible. Consequently, I collected data using online research platforms (Prime Panels for Study 1; Prolific for Study 2). This change in data collection strategy resulted in a significantly decreased cost per participant to collect data; as a consequence, the second major change in methodology from the original design was that instead of collecting data from only 600 participants per study, I was able to collect data from approximately 2200 people per study for the same cost.

Two other minor changes also occurred. First, because data for Study 2 were collected online instead of from the general Miami area, data were collected from White and Black participants (instead of Hispanic and Black participants). Second, when conducting ROC

analyses, I constructed full ROC curves instead of partial ROC curves, due to a recent published article recommending that statistical approach (Smith et al., 2020).

Outcomes

Activities/accomplishments

A total of three studies were conducted examining the effects of providing witnesses with an explicit 'not sure' option on their identification decisions. The pilot study, which involved collecting data in person from n = 322 participants, showed that identification outcomes did not vary as a function of the wording of the option ('not sure' vs. 'don't know'), nor as a function of whether a pre-lineup instruction was administered in addition to the option. These data were presented at a talk at the American Psychology-Law Society conference in March of 2020. Study 1, which involved collecting data online via Prime Panels from n = 2137 participants, showed that the 'not sure' option failed to ameliorate the increase in false identifications caused by the appearance-change instruction. Portions of these data were presented at talks at the American Psychology-Law Society conference in March of 2022 and March of 2023. Study 2, which involved collecting data online via Prolific from n = 2097 participants, showed that the 'not sure' option also failed to restore the loss in empirical discriminability caused by cross-race identifications.

Results and findings

Pilot study

Question 1: Did giving witnesses any option to withhold an identification response affect the hit/false alarms? To address this first research question, I first compared the hit rate among participants who received no withholding option to the hit rate among participants who received either of the withholding response options (i.e., the 'not sure' and 'don't know'

conditions were combined) among participants who received no pre-lineup instructions regarding the withholding option. The hit rate of witnesses who did not receive a withholding response option (.24) did not differ significantly from the hit rate of witnesses who did receive a withholding response option (.35), $\chi 2$ (df = 1) = 1.40, p = .237. Similarly, the overall target-present choosing rate of witnesses who did not receive a withholding response option (.62) did not differ significantly from that of witnesses who did receive a withholding response option (.56), $\chi 2$ (df = 1) = .34, p = .557.

Unexpectedly, the false alarm rate was too low to conduct reliable statistical tests on false identifications (i.e., among participants who received no pre-lineup instructions regarding a withholding response, only 5/102 = .05 made a false identification of the a priori innocent suspect). Consequently, I examined whether providing witnesses with an option to withhold an identification response affected the overall target-absent choosing rate. Witnesses who received a withholding response option were less likely to make an identification from a target-absent lineup (.26) than were witnesses who did not receive a withholding response option (.47), $\chi 2$ (df = 1) = 5.12, p = .024. However, the wording of the withholding response option did not matter: Whether witnesses received the 'not sure' response option (.33) or the 'don't know' response option (.19) had no significant effect on the target-absent choosing rate, $\chi 2$ (df = 1) =1.22, p = .270.

Question 2: Did the wording of the withholding response option affect the hit rate/false alarm rate? To address this second research question, I first compared the hit rates across withholding response option separately for witnesses who received a pre-lineup instruction regarding the withholding response option and for those who did not receive a pre-lineup instruction regarding the withholding option. Among witnesses who did not receive a pre-lineup

instruction, the hit rate of witnesses who received a 'not sure' response option (.38) did not differ from the hit rate of witnesses who received a 'don't know' response option (.30), $\chi 2$ (df = 1) = .30, p = .587. Similarly, among witnesses who received a pre-lineup instruction, the hit rate of witnesses who received a 'not sure' response option (.27) did not differ from the hit rate of witnesses who received a 'don't know' response option (.19), $\chi 2$ (df = 1) = .52, p = .471. The wording of the withholding option also did not affect overall target-present choosing rates, either among witnesses who did not receive a pre-lineup instruction ('not sure': .66; 'don't know': .43), $\chi 2$ (df = 1) = 2.67, p = .102, or among witnesses who did receive a pre-lineup instruction ('not sure': .55; 'don't know': .46), $\chi 2$ (df = 1) = .41, p = .522.

Again, because false alarms were too infrequent to conduct reliable analyses, I examined whether the wording of the withholding option affected the overall target-absent choosing rate. Wording of the withholding option had no significant effects on the overall target-absent choosing rate, whether witnesses did not receive a corresponding pre-lineup instruction ('not sure': .33; 'don't know': .19), $\chi 2$ (df = 1) = 1.35, p = .246, or did receive a corresponding pre-lineup instruction ('not sure': .35; 'don't know': .50), $\chi 2$ (df = 1) = 1.35, p = .246.

Question 3: Does providing a pre-lineup instruction regarding the withholding option (in addition to the response option itself) affect the hit rate/false alarm rate? To address this third research question, I first compared the hit rates and target-present choosing rates for witnesses who received pre-lineup instructions regarding the withholding option to witnesses who did not receive pre-lineup instructions regarding the withholding option. Because no differences were found in the earlier analyses regarding the wording of the response option, these analyses collapsed across witnesses who received a 'not sure' option and witnesses who received a 'don't know' option. Witnesses who did not receive a withholding response option were

omitted from this analysis. The hit rate of witnesses who received a pre-lineup instruction (.24) did not differ significantly from that of witnesses who did not receive a pre-lineup instruction (.35), $\chi 2$ (df = 1) = 1.62, p = .203. Similarly, the overall target-present choosing rate of witnesses who received a pre-lineup instruction (.51) did not differ significantly from that of witnesses who did not receive a pre-lineup instruction (.56), $\chi 2$ (df = 1) = .35, p = .555.

Once again, because the false alarm rate was too low to conduct reliable statistical analyses, I instead examined whether providing witnesses with a pre-lineup instruction regarding the withholding response option affected the overall target-absent choosing rate. The overall target-absent choosing rate of witnesses who received the pre-lineup instruction (.43) did not differ significantly from that of witnesses who did not receive the pre-lineup instruction (.26), $\chi 2$ (df = 1) = 3.37, p = .066.

Conclusion of pilot study. Three conclusions are drawn from this pilot study. First, giving people an option to withhold a decision decreased overall choosing rates among witnesses provided with a target-absent lineup. Second, the wording of the withholding option—whether phrased as 'not sure' or 'don't know'—had no significant impact on any outcomes measures (hit rates, choosing rates). Third, whether a corresponding pre-lineup instruction was provided to witnesses in addition to the withholding option itself had no significant effect on any outcome measures. Thus, it is concluded that a withholding option does make witnesses less likely to make an identification, but that specific variations in the manner in which the withholding option is presented make no significant difference.

Study 1

I first analyzed whether participants were able to identify each of the four perpetrators from a target-present lineup at above-chance levels. This was assessed in two ways. First, I

examined whether each of the four perpetrators was identified by more than 1/6th (16.7%) of the target-present choosers. Two of the four perpetrators were identified at above-chance levels (carjacking: 50.1%, $z = 11.58$, $p < .001$; store theft: 27.0%, $z = 3.93$, $p < .001$). The remaining two perpetrators were identified at chance levels (graffiti: 21.2%, $z = 1.47$, $p = .14$; bike theft: 11.4%, $z = -.195$, $p = .05$).

Second, I examined whether the probability of identifying the perpetrator from a target-present lineup was greater than the estimated false identification rate (equal to the mean probability of identifying a filler from a target-absent lineup: carjacking: 7.6%; store theft: 11.9%; graffiti: 10.4%; bike theft: 10.5%). The hit rate was higher than the mean false identification rate, demonstrating above-chance performance, for two perpetrators (carjacking: 28.4%, $z = 13.16$, $p < .001$; store theft: 20.5%; $z = 4.36$, $p < .001$), but not for the other two perpetrators (graffiti: 12.0%, $z = .92$, $p = .36$; bike theft: 8.0%, $z = -1.42$, $p = .16$). Because both analyses indicated that the graffiti perpetrator and the bike theft perpetrator resulted in chance-level responding, prohibiting a proper test of the study's hypotheses, participants who viewed those videos were omitted from subsequent analyses, resulting in a final sample size of $n = 1097$.

Among participants who received the 'not sure' option, 12% responded 'not sure'; this proportion did not differ significantly as a function of whether participants received the ACI (9.8%) or did not receive the ACI (12.3%), $\chi 2$ (df = 1) = .47, $p = .575$. To examine whether witnesses who responded 'not sure' had worse memories on average than witnesses who made a decision, a multivariate analysis of variance (MANOVA) was conducted on witnesses' responses to the various post-identification memory questions. The MANOVA indicated that, as expected, those witnesses who responded 'not sure' self-reported a significantly worse memory overall than witnesses who made a decision from the lineup, $F(10, 537) = 2.12$, $p = .021$.

A logistic regression was conducted to determine whether the likelihood of identifying the perpetrator among participants who viewed a target-present lineup differed as a function of condition. Correct identifications were regressed onto ACI condition and 'not sure' option condition in the first step of the model, and the interaction was added in the second step of the model. Neither the main effect of ACI, b = .227, SE = .269, Wald's $\chi 2$ (1) = .713, p = .398, e^b = 1.255, nor the main effect of 'not sure' instruction, b = -.043, SE = .269, Wald's $\chi 2$ (1) = .026, p = .872, e^b = .958, predicted correct identifications. However, ACI condition significantly interacted with the 'not sure' condition, b = -1.001, SE = .414, Wald's $\chi 2$ (1) = 5.835, p = .016, e^b = .368. Breaking down the interaction revealed that when no ACI was given, whether witnesses received the 'not sure' option or not did not significantly affect their likelihood of correctly identifying the perpetrator, b = -.043, SE = .269, Wald's $\chi 2$ (1) = .026, p = .872, e^b = .958. However, when the ACI was given, participants given the 'not sure' option were significantly less likely to correctly identify the perpetrator (.14) than when the 'not sure' option was not given (.31), b = -1.044, SE = .315, Wald's $\chi 2$ (1) = 10.992, $p < .001$, e^b = .352.

Another logistic regression was conducted to determine whether the likelihood of making a false identification of an innocent suspect among participants who viewed a target-absent lineup differed as a function of condition. Because no a priori innocent suspect was used, overall target-absent choosing was regressed onto ACI condition and 'not sure' condition in the first step of the model, and their interaction was added in the second step of the model. Replicating past research, witnesses given the ACI were significantly more likely to make an identification from a target-absent lineup (.63) than were witnesses not given the ACI (.53), b = .528, SE = .252, Wald's $\chi 2$ (1) = 4.375, p = .036, e^b = 1.695. Whether witnesses received the 'not sure' option or not was not related to the likelihood of making an identification from a target-absent lineup, b = -

.086, *SE* = .245, Wald's $\chi 2$ (1) = .123, *p* = .726, e^b = .918. The 'not sure' option also did not interact with the ACI, *b* = -2.58, *SE* = .350, Wald's $\chi 2$ (1) = .546, *p* = .460, e^b = .772.

Finally, whether participants made any identification from the lineup was regressed onto ACI condition, 'not sure' condition, the target-present vs. target-absent status of the lineup, and their various 2-way and 3-way interactions. None of the interactions were significant (all *p*s > .40). The main effect of 'not sure' option was also not significant, *b* = -.181, *SE* = .125, Wald's $\chi 2$ (1) = 2.075, *p* = .150, e^b = .835. Choosing was significantly higher among witnesses who viewed a target-present (.65) vs. target-absent (.58) lineup, *b* = .325, *SE* = .125, Wald's $\chi 2$ (1) = 6.726, *p* = .010, e^b = 1.384, and among witnesses who were provided with the ACI (ACI: .66; no ACI: .58), *b* = .360, *SE* = .126, Wald's $\chi 2$ (1) = 8.207, *p* = .004, e^b = 1.433.

To examine whether the ACI or the 'not sure' option affected empirical discriminability, ROC analyses were conducted. I first constructed full ROC curves (see Smith et al., 2020) for participants given the ACI and for participants not given the ACI, collapsed across 'not sure' condition, and compared their area under the curve (AUC) to determine if empirical discriminability differed. The ACI did not significantly affect discriminability, *D* = .745, *p* = .456 (AUC_{ACI}: .56; $AUC_{no\ ACI:}$ .59). Simple main effect analyses similarly demonstrated that the ACI did not significantly affect discriminability among only witnesses who were provided a 'not sure' option, *D* = 1.59, *p* = .11 (AUC_{ACI}: .53; $AUC_{no\ ACI}$: .61), nor among only witnesses who were not given a 'not sure' option, *D* = -.61, *p* = .54 (AUC_{ACI}: .60; $AUC_{no\ ACI}$: .57).

To examine whether the 'not sure' option affected discriminability, full ROC curves were constructed for participants given the 'not sure' option and those not given the 'not sure' option, collapsed across ACI condition. The 'not sure' option did not significantly affect discriminability, *D* = .316, *p* = .752 ($AUC_{no\ not\ sure}$: .58; $AUC_{not\ sure}$: .57). Simple main effect

analyses similarly demonstrated that the 'not sure' option did not affect discriminability among only witnesses who were given the ACI, $D = 1.348$, $p = .178$ ($AUC_{no\ not\ sure}$: .60; $AUC_{not\ sure}$: .53)., nor among only witnesses not given the ACI, $D = -.836$, $p = .403$ ($AUC_{no\ not\ sure}$ .67; $AUC_{not\ sure}$: .61).

Conclusion of Study 1: Although witnesses used the 'not sure' option when their memories were poor, as expected, providing witnesses with the 'not sure' option did not improve empirical eyewitness discriminability. In addition, although the appearance-change instruction increased false identifications, as expected based on past research, the 'not sure' option failed to protect against that harmful effect, contrary to hypotheses. In fact, when the ACI was given to witnesses, the 'not sure' option reduced correct identifications.

Study 2

I first analyzed whether participants were able to identify each of the four perpetrators from a target-present lineup at above-chance levels. This was assessed in two ways. First, I examined whether each of the four perpetrators was identified by more than $1/6^{th}$ (16.7%) of the target-present choosers. All four perpetrators were identified at above-chance levels (Target 1: 72.0%, $z = 20.97$, $p < .001$; Target 2: 47.2%, $z = 10.30$, $p < .001$; Target 3: 67.9%, $z = 18.63$, $p < .001$; Target 4: 74.7%, $z = 21.90$, $p < .001$).

Second, I examined whether the probability of identifying the perpetrator from a target-present lineup was greater than the estimated false identification rate (equal to the mean probability of identifying a filler from a target-absent lineup: Target 1: 10.3%; Target 2: 9.2%; Target 3: 7.4%; Target 4: 7.5%). The hit rate was higher than the mean false identification rate, demonstrating above-chance performance, for all four perpetrators (Target 1: 55.6%, $z = 23.98$, $p < .001$; Target 2: 30.0%; $z = 11.38$, $p < .001$; Target 3: 46.1%, $z = 24.35$, $p < .001$; Target 4:

55.8%, $z = 29.88$, $p < .001$). Consequently, data from all witnesses ($n = 2097$) were included in all analyses.

Among participants who were given the 'not sure' option, 19.3% responded 'not sure'; this did not differ significantly as a function of whether participants made a same-race (18%) or cross-race (21%) identification, $\chi 2$ (df = 1) = 1.15, $p = .284$. A logistic regression was conducted to determine whether the likelihood of identifying the perpetrator among participants who viewed a target-present lineup differed as a function of condition. Correct identifications were regressed onto race condition (i.e., same-race vs. cross-race) and 'not sure' option condition in the first step of the model, and the interaction was added in the second step of the model. Participants were significantly more likely to make correct identifications when the perpetrator was the same race (.502), as opposed to a different race (.439), as them, $b = -.254$, $SE = .124$, Wald's $\chi 2$ (1) = 4.183, $p = .041$, $e^b = .775$. The main effect of the 'not sure' option condition was not significant, $b = -.191$, $SE = .124$, Wald's $\chi 2$ (1) = 2.349, $p = .125$, $e^b = .826$, nor was the race x 'not sure' interaction, $b = .385$, $SE = .249$, Wald's $\chi 2$ (1) = 2.392, $p = .122$, $e^b = 1.470$.

Another logistic regression was conducted to determine whether the likelihood of making a false identification of an innocent suspect among participants who viewed a target-absent lineup differed as a function of condition. Because no a priori innocent suspect was used, overall target-absent choosing was regressed onto race condition (same-race vs. cross-race) and 'not sure' option condition in the first step of the model, and their interaction was added in the second step of the model. The 'not sure' option significantly interacted with the race condition, $b = -.677$, $SE = .249$, Wald's $\chi 2$ (1) = 7.417, $p = .006$, $e^b = .508$. Simple main effect analyses indicated that witnesses given the 'not sure' option were not significantly more likely to make a same-race identification from a target-absent lineup (.500) compared to witnesses not given the

'not sure' option (.437), b = .269, SE = .175, Wald's $\chi 2$ (1) = 2.369, p = .124, e^b = 1.309. However, witnesses given the 'not sure' option were significantly less likely to make a cross-race identification from a target-absent lineup (.506) compared to witnesses not given the 'not sure' option (.595), b = -.408, SE = .177, Wald's $\chi 2$ (1) = 5.331, p = .021, e^b = .665.

Finally, whether participants made an identification from the lineup was regressed onto race condition, 'not sure' condition, the target-present vs. target-absent status of the lineup, and their various 2-way and 3-way interactions. Only the race x 'not sure' option interaction was significant: b = -.474, SE = .184, Wald's $\chi 2$ (1) = 6.600, p = .010, e^b = .623 (all other interactions: $p > .16$). Follow-up analyses indicated that, for same-race witnesses, choosing among witnesses given the 'not sure' option (.585) was not significantly greater than choosing among witnesses not given the 'not sure' option (.574), b = .045, SE = .125, Wald's $\chi 2$ (1) = .130, p = .718, e^b = 1.046. However, for cross-race witnesses, choosing among witnesses given the 'not sure' option was significantly lower (.597) than choosing among witnesses not given the 'not sure' option (.689), b = -.404, SE = .130, Wald's $\chi 2$ (1) = 9.649, p = .002, e^b = .668.

To examine whether the ACI or the 'not sure' instruction affected empirical discriminability, ROC analyses were conducted. I first constructed full ROC curves (see Smith et al., 2020) for same-race identifications and cross-race identifications, collapsed across 'not sure' condition, and compared their area under the curve (AUC) to determine if empirical discriminability differed. Consistent with past research, this analyses revealed that empirical discriminability was significantly higher for same-race identifications (AUC = .74) compared to cross-race identifications (AUC = .70), D = 2.12, p = .034. However, simple main effect analyses demonstrated that this cross-race effect was only present when the 'not sure' option was not

given ($AUC_{same\text{-}race}$ = .75; $AUC_{cross\text{-}race}$ = .67), D = 2.34, p = .019; when the 'not sure' option was given, the cross-race effect disappeared ($AUC_{same\text{-}race}$ = 73; $AUC_{cross\text{-}race}$ = .72), D = .47, p = .638.

To examine whether the 'not sure' option affected discriminability, full ROC curves were constructed for participants given the 'not sure' option and those not given the 'not sure' option, collapsed across race condition. Empirical discriminability as measured by AUC did not significantly differ across 'not sure' option conditions, D = -.752, p = .452 ($AUC_{no\ not\ sure}$: .71; $AUC_{not\ sure}$: .73). Simple main effect analyses also showed a lack of a significant effect of the 'not sure' option among only same-race witnesses, D = -.367, p = .713 ($AUC_{no\ not\ sure}$: .75; $AUC_{not\ sure}$: .73), and among only cross-race witnesses, D = 1.52, p = .130 ($AUC_{no\ not\ sure}$ .67; $AUC_{not\ sure}$: .72).

Conclusion of Study 2: As expected, results demonstrated a cross-race effect: Same-race identifications exhibited greater empirical discriminability than cross-race identifications. However, the 'not sure' option, although reducing cross-race false identifications, failed to restore empirical discriminability, contrary to hypotheses.

Limitations

Due to the Covid-19 pandemic, data were collected using online platforms instead of in person, as originally planned. Consequently, participant compliance with instructions, as well as their level of engagement with the study (e.g., attention paid to stimuli) were uncontrolled. Despite efforts to ensure compliance and maintain high levels of engagement (such as various manipulation checks, using online platforms available only to prescreened participants, etc.), results could have been affected by the online data collection format. However, it should be noted that numerous published eyewitness identification studies have recently collected data using the same (or similar) online platforms as those used in the current studies. Furthermore,

statistical analyses of the current project's data indicate that participants were mostly responding at above-chance levels, indicating that participants were, on average, paying at least a moderate amount of attention.

Nonetheless, participant performance differed across studies. In particular, analyses of Study 1 participant data suggested that although performance was at above-chance levels for two targets, it was not above chance levels for two others, effectively cutting the sample size in that study in half. Consequently, the null results observed in Study 1 (i.e., that the 'not sure' instruction failed to reduce the increase in false identifications caused by the appearance-change instruction), could be partly due to relatively low engagement from Study 1 participants.

For this reason, Study 2 collected data using a different online research platform: Prolific. These data suggest a higher level of engagement from the participants, and analyses indicated that their performance overall was similar to the performance of witnesses in lab-based eyewitness identification studies. Thus, the Study 2 finding (that the 'not sure' option failed to improve eyewitness discriminability) is not likely to be attributable to low-engagement witnesses.

One other limitation should be acknowledged. Study 1—which examined whether the 'not sure' option would restore any discriminability lost as a result of using the appearance-change instruction (ACI)—was predicated on the assumption that the ACI would in fact harm discriminability. This assumption was based on the results of prior studies that had examined the effects of the ACI, all of which had consistently shown that while the ACI increases false identifications, it fails to increase correct identifications—a pattern that would indicate that the ACI harmed discriminability. And in fact, Study 1 results demonstrated a similar finding: The ACI increased false identifications but not correct identifications. However, ROC analyses

indicated that the ACI did not significantly harm discriminability, but rather simply resulted in a more liberal choosing criterion. It is possible that the decrease in sample size that resulted from at-chance responding in Study 1 reduced the statistical power required to find an effect of the ACI on discriminability. Alternatively, it is possible that ACI in fact does not affect discriminability, but simply makes witnesses more willing to choose. Therefore, instead of examining whether the 'not sure' option restored a loss in discriminability caused by the ACI, I instead examined whether the 'not sure' option was able to counteract the more liberal decision criterion caused by the ACI; results indicated that it did not.

Artifacts

Data sets generated

Three data sets were generated. Data set 1 contains data regarding the pilot study, which addressed whether the effectiveness of a 'not sure' option is dependent on (a) the exact wording of the option (i.e., 'not sure' vs. 'don't know'), and (b) how it was administered (i.e., as only a listed response option, or as a pre-lineup instruction and listed response option). Data set 2 contains data regarding Study 1, which addressed whether the 'not sure' option is able to reduce the harmful impact of the appearance-change instruction. Data set 3 contains data regarding Study 2, which addressed whether the 'not sure' option is able to reduce the harmful impact of cross-race identifications.

Dissemination activities

Three talks at a were given at the flagship conference in the field of legal psychology (the American Psychology-Law Society) regarding the results of this project:

Charman, S. D., Shambaugh, L., Cahill, B., Wylie, K., Hunter, M., & Blouir, M. Testing whether a 'not sure' option protects witnesses against the negative impact of other pre-lineup instructions. *American Psychology-Law Society,* Philadelphia, PA.

Charman, S. D., Shambaugh, L., & Hunter, M. Does a 'not sure' option protect witnesses from the effects of the appearance change instruction? *American Psychology-Law Society,* Denver, CO.

Mook, A., & Charman, S. D. "Not sure" or "Don't know": Examining methodological variations in the administration of a 'not sure' response option. *American Psychology-Law Society,* New Orleans, LA.