S1: Deviations from Pre-Registration

Table S1

Preregistration Deviations Table Template

Deviations					
#	Details		Original Wording	Deviation Description	Reader Impact
1	Type	Analysis	In our pre-registration, we stated that one of the analyses that we would conduct was an ANOVA predicting signal detection measures (*d'* and *c*). However, this analysis was not possible because only one observation was collected from each participant.	We replaced this analysis with a binary logistic regression on juror accuracy (accurate / inaccurate). In the logistic regression, we o We still calculate and report the signal detection measures for each condition and these can be found in the supplemental materials and in relevant follow-up analyses in the main manuscript.	The deviation should not impact the interpretation of the results.
	Reason	Plan not possible			
	Timing	After data access			
2	Type	Research Q(s)	In our pre-registration we list four hypotheses and a number of exploratory analyses. The complexity of results means that the paper would become long and difficult to follow if we focused on all of the research questions.	In this paper we focus on how results answer the research questions posed in our study description (and using the analysis from our preregistration) for ease of readability in the manuscript. We do refer the reader to the preregistration for a full list of hypotheses. We also more fully draw out our research questions relating to accuracy, which are exploratory.	The deviation should not impact the interpretation of the results.
	Reason	New knowledge			
	Timing	Other (Please Explain)			

				We had specifically noted implications of one potential interaction (which ended up not being significant) for accuracy, but in the paper examine the implications of other effects (which did end up being significant) for accuracy.	
3	Type	Analysis	In our pre-registration we planned to include all experimental conditions in our ANOVAs. However, the number of variables involved in analyses made them difficult to follow when reported in the paper and increased the length of the paper. In addition, reporting ROC analyses alongside signal detection measures created a long results section with multiple analyses gave the impression of rummaging to produce significant findings.	Our primary analyses in the paper now focus on our key experimental manipulations of interest, however we still report results from larger experimental designs controlling for other experimental variables in the supplemental materials. Primary results remain the same, and any divergence is reported in the manuscript. We now retain signal detection analyses in the paper, but have moved the ROC analyses (which produce equivalent results) into the supplemental materials.	The deviation should not impact the interpretation of the results.
	Reason	Peer review			
	Timing	After results known			
4	Type	Analysis	In our pre-registration we planned to conduct an analysis in Experiment 1 in which we compared a model including witness confidence and witness and lineup race to a model including assessor decisions only, in predicting identification decision accuracy. However, peer review feedback pointed out that this model had an unfair advantage in our Experiment 1	This analysis is now conducted with our Experiment 2 data, where the CRE is not exaggerated and therefore results are more generalizable.	The deviation should not impact the interpretation of the results.

			stimuli where the CRE was exaggerated when compared to the underlying dataset.		

Unregistered Steps

#	Details		Original Wording	Unregistered Step Description	Reader Impact
1	Type	Other (Please Explain)	The original preregistration contained one study.	We conducted a follow up study based on reviewer feedback about our stimuli selection.	This follow-up study provides additional information to help the reader understand the reliability and results of Experiment 1 better.
	Timing	After results known			
2	Type	Select One			
	Timing	Select One			
3	Type	Select One			
	Timing	Select One			

Note. Choose one characteristic for each dropdown menu. If more than one type or reason apply to the same deviation, then replace the dropdown menu with details. Provide one deviation per row, and add or delete rows as needed.

S2: Detail on Mock Crime Videos

We created two primary mock crime videos – one containing a Black defendant and one containing a White defendant. The target was visible for approximately 33 s in each video. The defendant (the Black or White target) enters the room and sits down near where people are studying. The victim's phone starts to ring so he gets up to answer it, placing the laptop on the seat beside him, then leaves the room to take the call. The thief then gets up, looks around the room, picks up the laptop and walks out of the room. After a few moments, the victim returns to the room, notices that his laptop is missing, and asks the person reading if they saw anything. There were two sub-versions of each video. In one version participants were shown a bystander entering the room before the defendant appeared, In the other version, participants were just shown other people in the room continuing their activities (reading and studying). Because the presence of a bystander did not have any significant effect on memory, it is not considered further in this study.

S3: Mock Crime Study Supplemental Analyses with Balanced Sample, and Controlling for Target Presence

Balanced Sample

ANOVA using lineup race, witness race and the interaction between witness race and lineup race to predict witness accuracy. (replications of results from our primary analysis are marked with a *):

*Significant main effect of lineup race (*B* = -3.49, *SE* = 1.25, Wald = 7.82, *p* =.01, OR = .03 CI[.003,.35]).

Significant main effect of witness race (*B* = -2.55, *SE* = 1.23, Wald = 4.29, *p* =.04, OR = 0.08 CI[.01,.87]), such that identification decisions made by Black witnesses (*M*=0.59 CI[0.46,0.72], *SE* = 0.06). were more accurate than identification decisions made by White witnesses (*M*=0.57 CI[0.45,0.70], *SE* = .06).

*Significant interaction between lineup race and witness race (*B* = 1.76, *SE* = 0.77, Wald = 5.22, *p* =.02, OR = 5.80 CI[1.28,26.23]).

Controlling for Target Presence

Logistic regression using lineup race, witness race, target presence, and all two-way interactions between these variables to predict witness accuracy (replications of results from our primary analysis are marked with a *):

*Significant interaction between participant race and lineup race (*B* = 1.44, *SE* = 0.71, Wald = 4.10, *p* =.04, OR = 4.20 CI[1.05,16.89]).

Significant interaction between target presence and lineup race, such that there was a significant main effect of target race in the target present line up (*B* = -3.20, *SE* = 0.64, Wald

= 24.70, p <.001, OR = 0.04 CI[0.01,0.14]), but not in the target absent line up (B = 0.11, SE = 0.49, Wald = 0.05, p =.82, OR = 0.18 CI[1.12,0.43]).

S4: Signal Detection Statistics for Mock Crime Study

Table S3. Signal Detection Statistics by Witness Race and Lineup Race.

	White Witness		Black Witness	
Condition	*d′*	*c*	*d′*	*c*
Lineup Race				
White	1.16	-0.39	0.62	-0.61
Black	0.10	0.21	1.07	0.37

S5: Juror Assessments of Accuracy Likelihood Supplemental Analyses (E1)

Controlling for Identification Type

ANOVA using witness race (White, Black), lineup race (White, Black), juror race (White, Black), and identification decision type (identification, rejection) to predict the jurors rating of the likelihood an identification was correct. The following effects were significant (replications of results from our primary analysis are marked with a *):

*Significant main effect of witness race ($F(1,1128)=22.69$, $MSe = 9985.28$, $p < .001$, $\eta_p^2 = .02$).

*Significant main effect of target race ($F(1,1128)=10.64$, $MSe = 4682.29$, $p = .001$, $\eta_p^2 = .01$).

Significant main effect of identification decision type ($F(1,1128)=8.70$, $MSe = 3828.10$, $p = .003$, $\eta_p^2 = .01$), such that assessors rated lineup rejections ($M = 68.18$, CI[66.49, 69.86], $SE = .86$) as more likely to be accurate than identifications ($M = 65.57$, CI[63.79, 67.34], $SE = 0.91$).

Controlling for Target Presence and Identification Detail

ANOVA using witness race (White, Black), lineup race (White, Black), juror race (White, Black), target presence (present, absent), and identification detail (identification, identification + detail) to predict the jurors rating of the likelihood an identification was

correct. The following effects were significant (replications of results from our primary analysis are marked with a *):

*Significant main effect of witness race ($F(1,1112)=24.12$, $MSe = 10437.58$, $p < .001$, $\eta_p^2 = .02$).

*Significant main effect of target race ($F(1,1112)=8.89$, $MSe = 3846.94$, $p = .003$, $\eta_p^2 = .01$).

*Significant interaction between target race and juror race ($F(1,1112)=4.08$, $MSe = 1766.26$, $p = .04$, $\eta_p^2 = .004$).

Significant interaction between witness race and juror race ($F(1,1112)=4.69$, $MSe = 2028.75$, $p = .03$, $\eta_p^2 = .004$). Jurors rated identification decisions of Black witnesses ($M = 70.51$, CI[68.68, 72.34], $SE = 0.93$) as significantly more likely to be accurate than identification decisions of White witnesses ($M = 63.97$ CI[62.11, 65.84], $SE = 0.95$), but this difference was more pronounced in Black jurors ($M_{black\ witness} = 72.97$, CI[69.96, 75.98], $SE = 1.54$; $M_{white\ witness} = 63.55$, CI[60.46, 66.64], $SE = .1.58$, $p<.001$) than in White jurors ($M_{black\ witness} = 68.05$, CI[65.98, 70.13], $SE = 1.06$; $M_{white\ witness} = 64.38$, CI[62.32, 66.48], $SE = 1.06$, $p = .01$).

Significant main effect of target presence, such that $F(1,1112)=5.33$, $MSe = 2306.08$, $p = .02$, $\eta_p^2 = .01$, such that assessors rated decisions from target absent lineups ($M = 68.13$, CI[66.47, 69.78], $SE = .84$) as more likely to be correct than decisions from target present lineups ($M = 65.23$, CI[63.37, 67.09], $SE = 0.95$).

Significant interaction between target race and target presence ($F(1,1112)=24.72$, $MSe = 10699.21$, $p < .001$, $\eta_p^2 = .02$), such that in White lineups there was no significant difference between rated accuracy for decisions from target absent lineups and target present lineups ($p=.058$) but in Black lineups, assessors rated decisions from target absent lineups ($M = 69.62$, CI[67.30, 71.93], $SE = 1.18$) as more likely to be correct than decisions from target

present lineups (M = 61.05, CI[58.25, 63.85], SE = 1.42) (F(1,581)=21.41, MSe = 10696.08, p <.001, η_p^2 = 0.04).

A significant interaction between witness race, lineup race, and target presence (F(1,1112)=5.58, MSe = 2416.20, p =.02, η_p^2 = .01), illustrated in Figure S1, below.

A significant interaction between lineup race, target presence, and videos seen (F(1,1112)=9.27, MSe = 4013.71, p =.002, η_p^2 = .01), illustrated in Figure S2, below.

Figure S1

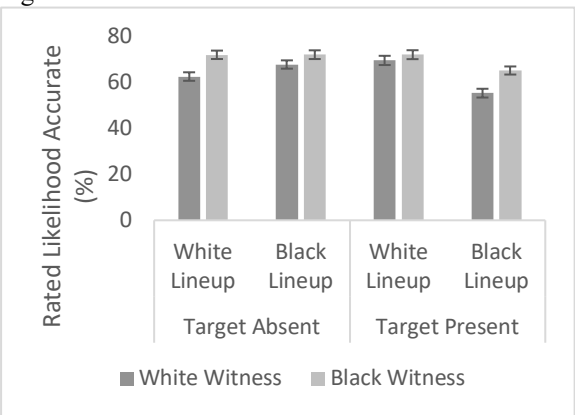
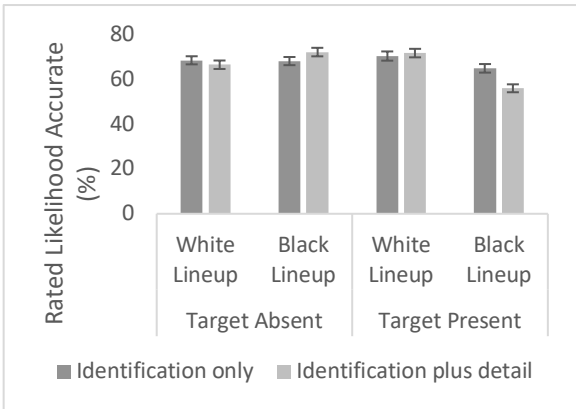


Figure S2



Note: Error Bars +/- one standard error.

S6: Juror Assessment Accuracy Supplemental Analyses (E1)

Controlling for Identification Type

ANOVA using witness race (White, Black), lineup race (White, Black), juror race (White, Black), and identification decision type (identification, rejection) to predict signed confidence juror accuracy. The following effects were significant (replications of results from our primary analysis are marked with a *):

*Significant main effect of lineup race ($F(1,1123)=4.05$, $MSe = 5200.71$, $p = .04$, $\eta_p^2 = 0.004$).

*Significant main effect of witness race ($F(1,1123)=8.88$, $MSe = 11407.93$, $p = .003$, $\eta_p^2 = .01$).

Significant main effect of juror race ($F(1,1123)=5.16$, $MSe = 6624.27$, $p = .02$, $\eta_p^2 = .01$), such that Black jurors ($M = 43.72$, CI[39.73, 47.70], $SE = 2.03$) showed higher levels of signed confidence accuracy than White jurors ($M = 39.31$, CI[36.74, 41.89], $SE = 1.31$).

*Significant interaction between target race and witness race ($F(1,1123)=49.65$, $MSe = 63771.82$, $p < .001$, $\eta_p^2 = .04$).

Controlling for Target Presence and Identification Detail

ANOVA using witness race (White, Black), lineup race (White, Black), juror race (White, Black), target presence (present, absent), and identification detail (identification,

identification + detail) to predict signed confidence juror accuracy. The following effects were significant (replications of results from our primary analysis are marked with a *):

*Significant main effect of witness race ($F(1,1107)=12.90$, $MSe = 14559.95$, $p < .001$, $\eta_p^2 = .01$).

Significant main effect of juror race ($F(1,1107)=4.77$, $MSe = 5384.68$, $p = .03$, $\eta_p^2 = .004$), such that Black jurors ($M = 43.72$, CI[39.73, 47.70], $SE = 2.03$) showed higher levels of signed confidence accuracy than White jurors ($M = 39.31$, CI[36.74, 41.89], $SE = 1.31$).

*Significant interaction between lineup race and witness race ($F(1,1107)=71.39$, $MSe = 80576.90$, $p < .001$, $\eta_p^2 = .06$).

*Significant interaction between lineup race and juror race ($F(1,1107)=3.90$, $MSe = 4401.64$, $p = .049$, $\eta_p^2 = .004$).

Significant interaction between lineup race and target presence ($F(1,1107)=142.47$, $MSe = 160798.05$, $p < .001$, $\eta_p^2 = .11$), such that in Target Absent lineups signed confidence juror accuracy was significantly higher for Black lineups ($M = 51.74$, CI[47.61, 55.87], $SE = 2.10$) than White lineups ($M = 30.38$, CI[26.29, 34.46], $SE = 2.08$; $F(1,577)=52.40$, $MSe = 66070.90$, $p < .001$, $\eta_p^2 = 0.08$) and in Target Present lineups signed confidence juror accuracy was significantly higher for White lineups ($M = 56.13$, CI[52.11, 60.15], $SE = 2.04$) than for Black lineups ($M = 26.33$, CI[22.26, 30.40], $SE = 2.07$; $F(1,558)=104.49$, $MSe = 124025.51$, $p < .001$, $\eta_p^2 = 0.16$).

S7: Signal Detection Statistics for Experiment 1

Table S7. Signal Detection Statistics for Each Condition by Juror Race

Condition	Black Juror		White Juror	
	d′	*c*	*d′*	*c*
Witness Race				
Black	0.82	-0.95	0.51	-1.02
White	0.58	-0.72	0.27	-0.82
Lineup Race				
Black	1.27	-1.02	0.47	-0.94
White	0.31	-0.78	0.32	-0.91
Target Presence				
Absent	0.64	-0.90	0.34	-0.92
Present	0.78	-0.76	0.44	-0.92
Identification Detail				
Identification + Detail	0.84	-0.74	0.48	-0.85
Identification Only	0.53	-0.95	0.30	-0.99

S8: ROC Analyses

Same Race vs. Cross Race Identifications

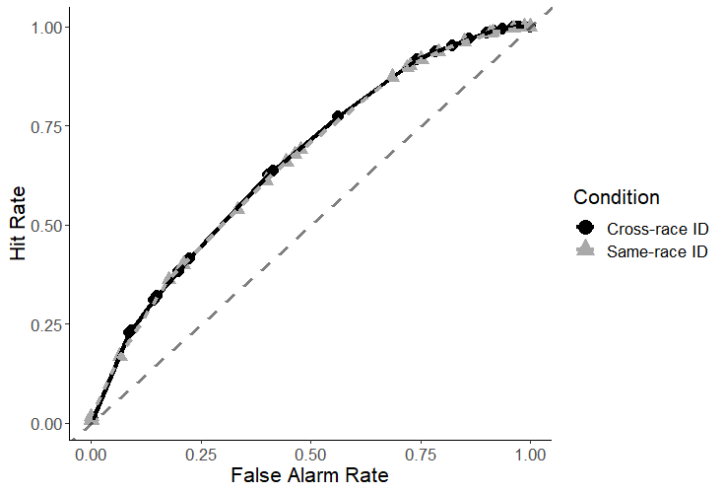
To provide an additional test of whether judgments relating to same-race identification decisions had better discriminability than judgments relating to cross-race identification decisions, we created ROC curves separately for same-race and cross-race identification decisions by aggregating participants responses by their confidence in their assessment, then plotting the cumulative hit rates against the false alarm rates at each level of confidence. Eight participants were not included in the ROC analyses because they did not rate their confidence in their assessment. ROC curves were created using the *fullROC* package v0.1.0 in R 4.2.3. The hit rate was defined as the number of times that participants said that correct identification decisions were accurate divided by the total number of correct identification decisions shown. The false alarm rate was defined as the number of times that participants said that incorrect identification decisions were accurate divided by the total number of incorrect identification decisions shown. The area under the curves thus provides a measure of how well participants discriminated between accurate and inaccurate identification decisions (i.e., discriminability).

Results were consistent with the idea that people are poorer at assessing the evaluations of cross-race identification decisions because of a bias towards assuming that identification decisions are accurate, and reduced accuracy of cross-race identifications. Figure S8(1) shows that jurors' ability to discriminate between correct and incorrect witnesses was similar for same-race identification decisions (bootstrapped AUC = 0.56, CI[0.50, 0.62] and cross-race identification decisions (bootstrapped AUC = 0.53, CI[0.48, 0.57]. A bootstrapped comparison of the AUC revealed no significant difference in juror discrimination accuracy between same-race identification decisions and cross-race identification decisions (p = .35). In sum, these analyses suggest that people are not better at assessing the accuracy of same-

race identification decisions because they are more sensitive to which identification decisions are accurate and which are not, but because they exhibit a general bias towards rating identification decisions as accurate and this assumption is more often correct for same-race identifications than it is for cross-race identifications.

Figure S8(1)

ROC Curve for Assessments of Cross-Race and Same-Race Identification Decisions.



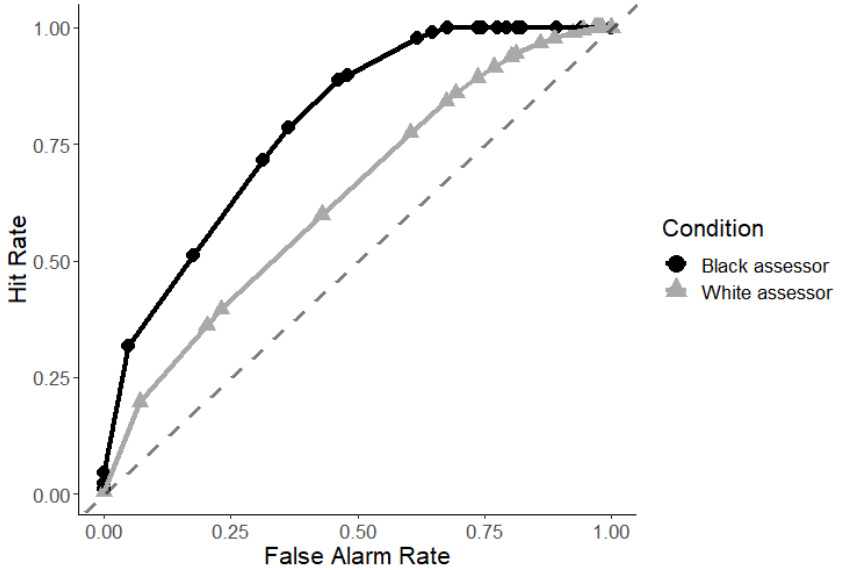
Note. The dashed line represents chance-level performance (hit rate = false alarm rate).

Superior Performance of Black Jurors

To examine the superior performance of Black jurors in assessing identification decisions relating to Black lineups, we created ROC curves following the procedure described above. The ROC curves show that Black jurors (bootstrapped AUC = 0.57, CI[0.49, 0.65]) were better at discriminating between correct and incorrect identification decisions from Black lineups than White jurors were (bootstrapped AUC = 0.51, CI[0.45, 0.56]) (Figure S8(2)). A bootstrapped comparison, revealed that discriminability did not significantly differ for White and Black jurors (p = .23), however this lack of significance is likely the result of a relatively small number of observations compared to the recommended sample size for ROC analyses (300-500 participants) ($n_{white\ jurors}$ = 390, hits = 40%, false alarms = 41.5%; $n_{black\ jurors}$ = 193, hits = 44.6%, false alarms = 34.7%).

Figure S8(2)

ROC Curve for Black Juror (Assessor) and White Juror (Assessor) Assessments of Identification Decisions from Black Lineups.



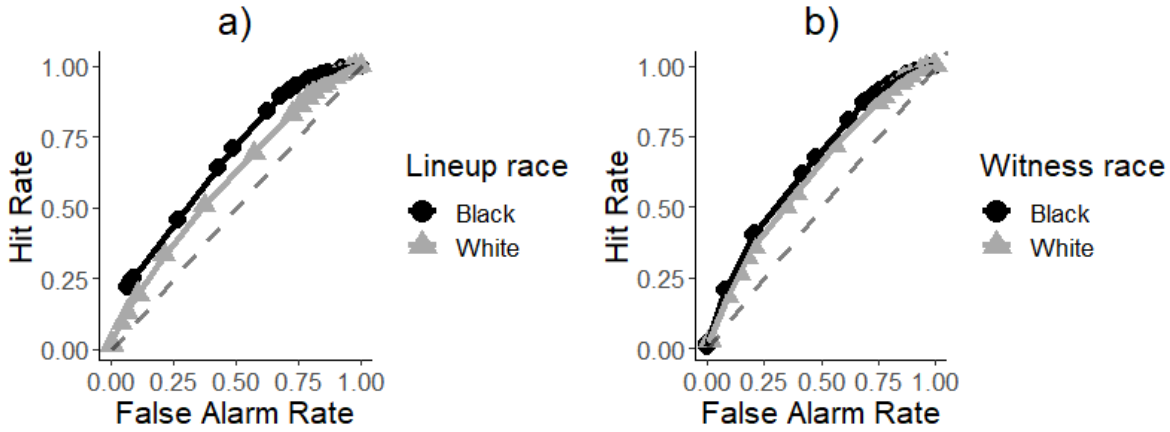
Note. The dashed line represents chance-level performance (hit rate = false alarm rate).

Lineup Race and Witness Race

ROC analyses suggested that lineup race did not significantly affect discriminability (bootstrapped $AUC_{black\ lineup}$ = 0.53, CI[0.48, 0.57], bootstrapped $AUC_{white\ lineup}$ = 0.55, CI[0.50, 0.60], p = .52; see Figure S8(3)a), and that discriminability did not significantly differ based on whether the witness who made the identification was White (bootstrapped AUC = 0.55, CI[0.50, 0.59]) or Black (bootstrapped AUC = 0.53, CI[0.48, 0.58], p = .62).

Figure S8(3)

ROC Curve for Assessments of Identification Decisions by Lineup Race and Witness Race



Note. **The dashed line represents chance-level performance (hit rate = false alarm rate).**

S9: Descriptive Statistics Relating to Experiment 2 Stimuli

		Accuracy (%)		Confidence	
		M	CI	*M*	CI
Full Dataset (White witnesses)	Lineup Race				
	White	65.79	[54.88, 76.70]	71.05	[65.26, 76.84]
	Black	33.33	[22.18,44.49]	65.28	[60.56, 69.99]
Experiment 2 Stimuli	White	77.50	[63.97, 91.03]	71.50	[63.58, 79.42]
	Black	37.50	[21.82, 53.18]	63.50	[56.57, 70.43]

S10: Juror Assessments of Accuracy Likelihood Supplemental Analyses (E2)

Identification Decision Type

ANOVA using lineup race (White, Black) and identification decision type (identification, rejection) to predict the jurors rating of the likelihood an identification was correct. The following effects were significant (replications of results from our primary analysis are marked with a *):

*Significant main effect of lineup race ($F(1,653)=5884.66$, $MSe = 2657118.40$, $p < .001$, $\eta_p^2 = 0.90$).

Significant interaction between lineup race and identification decision type ($F(1,653)=27.11$, $MSe = 12241.44$, $p < .001$, $\eta_p^2 = .04$), such that the main effect of lineup race is significant for identifications ($F(1,313)=28.18$, $MSe = 14470.69$, $p < .001$, $\eta_p^2 = 0.08$), but not for lineup rejections ($F(1,340)=3.06$, $MSe = 1205.55$, $p = .08$, $\eta_p^2 = 0.01$).

Target Presence

*Significant main effect of lineup race ($F(1,653)=6.31$, $MSe = 2940.57$, $p = .01$, $\eta_p^2 = 0.01$).

Significant interaction between lineup race and target presence ($F(1,653)=5.58$, $MSe = 2599.58$, $p = .02$, $\eta_p^2 = 0.01$), such that the main effect of lineup race is significant for decisions relating to identification decisions from target present lineups ($F(1,327)=12.43$, $MSe = 5542.91$, $p < .001$, $\eta_p^2 = .04$), but not for decisions relating to identification decisions from target absent lineups ($F(1,326)=0.01$, $MSe = 5.25$, $p = .92$, $\eta_p^2 = 0.00$).

S11: Juror Assessment Accuracy Supplemental Analyses (E2)

Identification Decision Type

Logistic regression using lineup race, identification decision type, and the interaction between lineup race and identification decision type to predict assessor accuracy. The following effects were significant (replications of results from our primary analysis are marked with a *):

*Significant main effect of lineup race (*B* = -1.74, *SE* = 0.29, Wald = 36.09, *p* <.001, OR = 0.18 CI[0.10,0.31]).

Significant interaction between lineup race and identification decision type (*B* = 0.97, *SE* = 0.42, Wald = 5.31, *p* =.02, OR = 0.38 CI[0.17,0.87]), such that the significant main effect of lineup race was significant for both identifications (*B* = -2.7, *SE* = 0.31, Wald = 78.77, *p* <.001, OR = 0.07 CI[0.04,0.12]) and lineup rejections (*B* = -1.74, *SE* = 0.29, Wald = 36.09, *p* <.001, OR = 0.18 CI[0.10,0.31]), but larger for identifications.

Target Presence

Logistic regression using lineup race, target presence, and the interaction between lineup race and target presence to predict assessor accuracy. The following effects were significant (replications of results from our primary analysis are marked with a *):

*Significant main effect of lineup race (*B* = -0.49, *SE* = 0.25, Wald = 4.02, *p* =.045, OR = 0.61 CI[0.38,0.99]).

Significant interaction between lineup race and target presence (*B* = -2.94, *SE* = 0.40, Wald = 53.33, *p* <.001, OR = 0.05 CI[0.02,0.12]), such that the significant main effect of lineup race was significant for both target present (*B* = -3.43, *SE* = 0.32, Wald = 116.03, *p* <.001, OR = 0.03 CI[0.02,0.06]) and target absent lineups (*B* = -0.49, *SE* = 0.25, Wald = 4.02, *p* =.045, OR = 0.61 CI[0.38,0.99]), but larger in target present lineups.