aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2014, Vol. 9(5) 556–578

DOI: 10.1177/1745691614545653
pps.sagepub.com
SAGE

Registered Replication Report: Schooler and Engstler-Schooler (1990)

Proposing Authors: This proposal was initiated by the editors

Contributing authors (alphabetical order): Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., Bornstein, B. H., Bouwmeester, S., Brandimonte, M. A., Brown, C., Buswell, K., Carlson, C., Carlson, M., Chu, S., Cislak, A., Colarusso, M., Colloff, M. F., Dellapaolera, K. S., Delvenne, J.-F., Di Domenico, A., Drummond, A., Echterhoff, G., Edlund, J. E., Eggleston, C. M., Fairfield, B., Franco, G., Gabbert, F., Gamblin, B. W., Garry, M., Gentry, R., Gilbert, E. A., Greenberg, D. L., Halberstadt, J., Hall, L., Hancock, P. J. B., Hirsch, D., Holt, G., Jackson, J. C., Jong, J., Kehn, A., Koch, C., Kopietz, R., Körner, U., Kunar, M. A., Lai, C. K., Langton, S. R. H., Leite, F. P., Mammarella, N., Marsh, J. E., McConnaughy, K. A., McCoy, S., McIntyre, A. H., Meissner, C. A., Michael, R. B., Mitchell, A. A., Mugayar-Baldocchi, M., Musselman, R., Ng, C., Nichols, A. L., Nunez, N. L., Palmer, M. A., Pappagianopoulos, J. E., Petro, M. S., Poirier, C. R., Portch, E., Rainsford, M., Rancourt, A., Romig, C., Rubínová, E., Sanson, M., Satchell, L., Sauer, J. D., Schweitzer, K., Shaheed, J., Skelton, F., Sullivan, G. A., Susa, K. J., Swanner, J. K., Thompson, W. B., Todaro, R., Ulatowska, J., Valentine, T., Verkoeijen, P. P. J. L., Vranka, M., Wade, K. A., Was, C. A., Weatherford, D., Wiseman, K., Zaksaite, T., Zuj, D. V., Zwaan, R. A.

Protocol vetted by: Jonathan W. Schooler
Protocol edited by: Daniel J. Simons

Multilab direct replication of: Study 4 (modified) and Study 1 from Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36–71.

Data and registered protocols: https://osf.io/ybeur/

Citation: Alogna, V. K., Attaya, M. K., Aucoin, P., Bahnik, S., Birch, S., Birt, A. R., ... Zwaan, R. A. (2014). Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*, 556–578.

Abstract

Trying to remember something now typically improves your ability to remember it later. However, after watching a video of a simulated bank robbery, participants who verbally described the robber were 25% worse at identifying the robber in a lineup than were participants who instead listed U.S. states and capitals—this has been termed the "verbal overshadowing" effect (Schooler & Engstler-Schooler, 1990). More recent studies suggested that this effect might be substantially smaller than first reported. Given uncertainty about the effect size, the influence of this finding in the memory literature, and its practical importance for police procedures, we conducted two collections of preregistered direct replications (RRR1 and RRR2) that differed only in the order of the description task and a filler task. In RRR1, when the description task immediately followed the robbery, participants who provided a description were 4% less likely to select the robber than were those in the control condition. In RRR2, when the description was delayed by 20 min, they were 16% less likely to select the robber. These findings reveal a robust verbal overshadowing effect that is strongly influenced by the relative timing of the tasks. The discussion considers further implications of these replications for our understanding of verbal overshadowing.

Keywords

recognition memory, verbal overshadowing, eyewitness, lineup identification, replication

Address correspondence to:
Daniel J. Simons, University of Illinois at Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820
E-mail: dsimons@illinois.edu

If you want to remember something better, practice it. This mantra follows from decades of memory research: repeat the names of people you have just met, study flashcards for your upcoming language test, summarize the chapter you read. Other techniques might be even better, but this type of rehearsal cannot hurt. Or can it?

The results of Schooler and Engstler-Schooler (1990; henceforth S&E-S) suggested that when the to-be-remembered materials are faces, verbal rehearsal hurts rather than helps memory performance. Participants in their study witnessed a video of a simulated bank robbery. Half wrote a description of the robber, and half completed an unrelated writing task. All then tried to pick the robber out of a photo lineup. Those who had provided a written description correctly identified that robber approximately 25% less often than those who performed the unrelated writing task.

This finding, dubbed "the verbal overshadowing effect," suggests that verbally describing a person impairs later recognition memory for that person. Thus, eyewitness recollection may be impaired by asking witnesses to describe what they saw, a result with both practical and theoretical importance. The paper has had a substantial impact on the field: It has been cited more than 600 times and is a staple of psychology textbooks.

Yet the magnitude of the effect remains uncertain. Schooler has noted that the measured effect size of the overshadowing effect reported in later studies is smaller than that in the original report (Schooler, 2011; see also Lehrer, 2010). A meta-analysis of verbal overshadowing studies of lineup recognition performance revealed a significant but smaller (about 12%) effect of verbal description (Meissner & Brigham, 2001). The studies in the meta-analysis used a variety of stimuli, delays, filler tasks, and other materials, and the measured effect size across studies showed substantial heterogeneity, with some studies finding no effect or even an effect in the opposite direction.[1] The studies also might have overestimated the true effect because they had small sample sizes: The collection of studies included more statistically significant results than would be expected based on their power to find statistical significance, indicating a pattern of publication bias in favor of statistically significant results (Francis, 2012).[2] In the presence of publication bias, the true population effect size is difficult to estimate from a meta-analysis. Moreover, some of the differences in methods used across the studies could have moderated the underlying effect. For example, the meta-analysis found less verbal overshadowing with a delay between the verbal description task and the lineup identification task. However, the size of the delay varied substantially across studies.

Verbal overshadowing potentially has broad ramifications, both for our understanding of the mechanisms of memory and for police practices. If asking a witness to verbally describe the person they saw substantially impairs their ability to recognize that person later, then eyewitness identification should be weighted less if the witness had provided a description earlier. Given the importance and influence of this finding, coupled with uncertainty about the size of the effect and the absence of any large-scale direct replications of it, the original study merits a large-scale direct replication to better determine the size of the effect. This registered replication report (RRR) was designed to provide an accurate estimate of the verbal overshadowing effect via a collection of pre-registered, independently conducted direct replications of the original study, all using the same materials and a common, vetted protocol.

Protocol Development to Compare Past and Present Studies

The protocol for a direct replication of the original verbal overshadowing study was developed in collaboration with the lead author of the original article, Jonathan Schooler. Once the protocol was completed, *Perspectives on Psychological Science* publicly announced a call for laboratories interested in participating on May 14, 2013. Based on the rapid response from a large number of labs, we set a deadline for proposals of June 11, 2013. A total of 31 labs joined the initial replication project (RRR1). All labs preregistered the details of their plan to implement the protocol, the editors verified those plans before data collection began, and each lab conducted an independent replication. Of those teams, 22 completed a follow-up experiment (RRR2) that reversed the order of the filler task and the description task.

We conducted RRR2 after discovering an error in the original protocol that went unnoticed throughout the development process. Although we had intended to replicate S&E-S Study 1, the protocol inadvertently reversed the order of the verbal description task and the filler task. In S&E-S Study 1, participants saw the video, did the filler task, then wrote their verbal description and moved to the lineup task. In RRR1, they wrote their description immediately after seeing the video and then did the filler task, thus adding a 20-min delay before the lineup task. Previous evidence suggests that introducing a delay between the verbal description task and the lineup can reduce the overshadowing effect, meaning that the task order of RRR1 might not provide the strongest possible test of the overshadowing effect (e.g., Finger & Pezdek, 1999; Meissner & Brigham, 2001; note, though, that S&E-S Study 4 showed a roughly comparable overshadowing effect using an order comparable to RRR1). After a participating laboratory noticed the error, the editors consulted with Schooler and we collectively decided to conduct RRR2, reversing the task order to match that of S&E-S

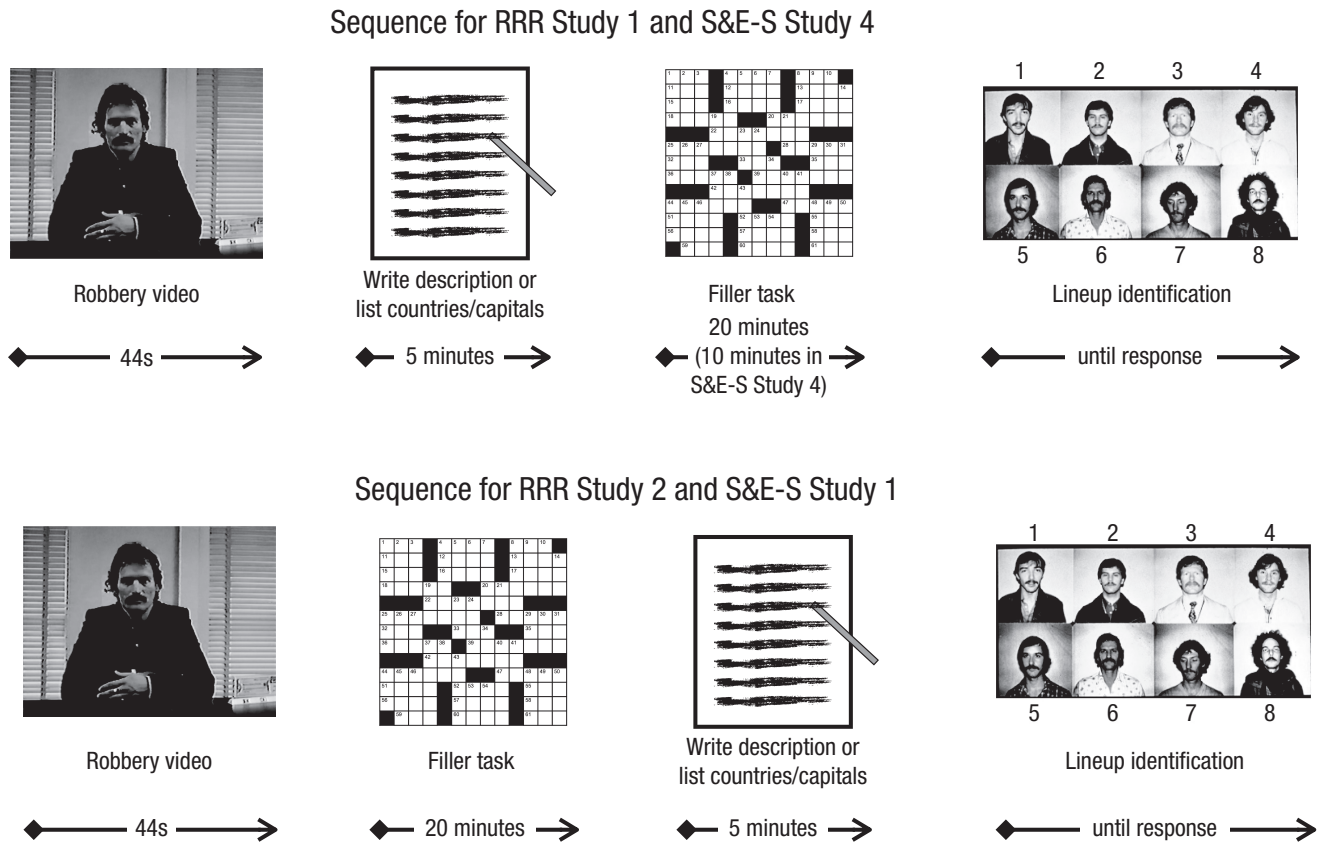

Fig. 1. Illustration of the task sequence for RRR1 (S&E-S Study 4) and RRR2 (S&E-S Study 1). Note that S&E-S used a different filler task, but a crossword puzzle was used in the replication studies at Jonathan Schooler's suggestion. Also, in S&E-S Study 4, the filler task lasted 10 min rather than 20 min. For the replication studies, we kept the duration of the filler task constant.

Study 1. That way, we could replicate the original study as intended, providing the strongest and clearest test of the verbal overshadowing effect, and we could also examine the effect of task order by comparing the two RRR studies. Critically, this decision was made before data collection from RRR1 was completed or analyzed, making the decision blind to the outcome of RRR1. Moreover, labs were not informed about the results from any other labs until data collection from both studies had been completed.

For the purposes of this report, we treat RRR1 as a fairly direct replication of S&E-S Study 4. The studies used the same task ordering, with the filler task coming after the verbal description. Note that S&E-S Study 4 included another between-subject condition and that the delay was 10 min rather than 20 min. So, RRR1 is not an exact replication of all conditions of S&E-S Study 4. However, the difference in the length of the delay is the only substantive change in procedure from a direct replication of the critical comparisons for a test of verbal overshadowing on face recognition. A benefit arising out of our error is that, by using the same timing in RRR1 and RRR2, we can provide one of the first highly powered direct comparisons of the influence of task order on the verbal overshadowing effect.

Many of the teams consisted of experts on memory and eyewitness accuracy, including some researchers who had previously studied verbal overshadowing. Other labs had experience in conducting other types of cognitive psychology experiments, and still others lacked domain-specific experience but were skilled in experimental methods and were interested in replication efforts more broadly. The participating labs included teams from 11 countries and from a variety of college and university settings. For labs in non-English speaking countries, the associated researchers translated the instructions and other materials and then independently translated them back to English to verify accuracy. In some cases, the audio track on the bank robbery video was dubbed into the native language of the participants. Details of this translation process and any other departures from the standardized procedures are noted by the participating labs in the individual study descriptions (see Appendix). Laboratories were responsible for obtaining any necessary ethics approval from their institutions.

In addition to the lab-based studies, one lab that had participated in RRR1 replicated the procedures of both RRR1 and RRR2 in a large-scale online experiment using participants from Amazon Mechanical Turk. Except as noted in the study description below, it followed the same protocols as the lab-based studies. Given that it adopted a different procedure, it was not included in the meta-analytic effect size estimates, but it is reported alongside the lab-based results for comparison.

Protocol Requirements

Participants

The protocol specified a minimum allowable sample size of 50 participants in each condition, but labs were encouraged to include as many participants as possible. Given that the goal was a direct replication of the original result, the protocol specified that participants be drawn from an undergraduate subject population with all participants between the ages of 18 and 25 years. It further required that participants be able to understand the instructions and have vision adequate to perceive the events in the video and to recognize people. Because the robber depicted in the original video was White, and the verbal overshadowing effect is thought to be weaker with other-race faces than with own-race faces (Fallshore & Schooler, 1995), only White participants were included in the analyses reported here. The sample in each replication study was required to be between 20% and 80% female. Many of the labs collected additional data from participants who did not meet these inclusion criteria, and data from all participants are included in the data files posted on the main project page at Open Science Framework (OSF; https://osf.io/ybeur/).

Testing location

The protocol required in-person testing. Testing could occur individually or in small groups, provided that participants could not see or hear each other when viewing stimuli or responding and that they could not communicate with each other during the study. The protocol specified that the study could not be conducted in a classroom setting. (This stipulation was included to maximize the similarity of the testing context across labs.)

Experimenters

Any trained research assistant, postdoctoral researcher, or faculty member could serve as the experimenter if they had experience collecting experimental psychology data and interacting with participants. No special expertise was required to conduct the study, and the experimenter did not need to be blind to condition assignment (as that would be difficult to achieve and was not the case in the original study).

Materials

Schooler provided a digitized version of the original videotape that was then reformatted as a QuickTime movie file. Schooler also provided a digital version of the original lineup image (an 8-person lineup that included the robber) as well as the text of the instructions given to participants for each task. The original studies used a variety of filler tasks, but Schooler recommended using a crossword puzzle, something he had done in some of his studies. The original crossword puzzle was no longer available, so Schooler selected a comparable crossword puzzle. All of these materials are available from https://osf.io/ybeur/.

Data collection

The study could be conducted by presenting the video using a computer display, television, or projector and by collecting written responses and ratings either on paper or on a computer. Participants were blind to the hypothesis about verbal overshadowing and were unaware of any experimental conditions other than their own. They also were not informed that they were participating in a recognition memory experiment—the study description used for recruiting participants described it more vaguely as a study of perception and memory. Participants were randomly assigned to the experimental and control conditions with the constraint that approximately equal numbers of participants were assigned to each condition. Labs differed in how they implemented the random assignment, and details are provided in the individual study descriptions (see Appendix). Note that the original S&E-S studies assigned participants to conditions in small groups and all members of each group were assigned to the same condition. The replication protocol required individual assignment to condition to eliminate this nonindependence in randomization.

Procedure

Participants were told, "This experiment consists of several tasks. First, please pay close attention to the following video." They then viewed a 44-s video depicting a bank robbery. Participants assigned to the Experimental condition were then asked to write a description of the robber:

> Please describe the appearance of the bank robber in as much detail as possible. It is important that you attempt to describe all of his different facial

features. Please write down everything that you can think of regarding the bank robber's appearance. It is important that you try to describe him for the full 5 minutes.

Participants assigned to the Control condition were asked to "Please name as many countries and their capitals as you can." In the original study, participants were asked to list the states of the United States and their capitals, but for the replication protocol, the control task was changed because participants outside the United States might not be as familiar with states and capitals in the United States.

After 3 min, each group received a reminder to keep working. Participants in the Experimental condition were told, "Please continue describing every detail of the bank robber. It is important that you provide as full a description as possible." Participants in the Control condition were told, "Please continue to list countries and their capitals. It is important that you continue this task for the full 5 minutes." This reminder could be spoken aloud or presented on the computer display. If the reminder was spoken and the testing session included participants from both conditions, the reminder was phrased to be condition blind: "Please keep working. It's important that you continue the task for the full 5 minutes."

After 5 min of writing/typing, participants spent 20 min working on a printed crossword puzzle. Immediately after this filler task, participants viewed a lineup of eight faces and heard/read the following instructions: "Next you will see a lineup with eight faces. Please identify the individual in the lineup who you believe was the bank robber in the video you watched earlier. If you do not believe the bank robber is present please indicate 'not present.'" If the lineup was presented on a computer or projector, the images were numbered from 1 to 8 to allow a keyboard response, and the last sentence of the instructions was modified to end "... please indicate 'not present' by pressing '9' [or '0']. Press 'space' to view the image." Finally, participants were asked to "Please indicate your confidence in your selection from the lineup on a scale from 1 (guessing) to 7 (certain)."

Data collection stopping rules and exclusions

Each lab documented their stopping rules for data collection as part of their OSF preregistration (see Appendix for links), and the editors reviewed these procedures to verify that they ensured random assignment to conditions and that each lab would be able to meet the minimum required sample size after any exclusions necessitated by the protocol requirements. Labs were permitted to exclude participants for any of the following reasons: participants did not meet the age or race requirements for the study, participants did not follow instructions on the experimental or control task, participants did not complete all tasks, or the experimenter/computer incorrectly administered the task or instructions. Labs were permitted to prespecify additional exclusions necessitated by their testing situation (e.g., failure to understand the nature of the video). All decisions about whether or not to exclude data were made prior to examining performance on the recognition task and were based on factors unrelated to the outcome measures. All excluded data are included in the data files along with the reason for exclusion.

Differences between RRR1 and RRR2

All materials and procedures were identical across the two studies except for the following substantive changes (see Fig. 1):

(a) In RRR2, the crossword puzzle filler task followed immediately after the video and preceded the verbal description (experimental) or countries/capitals (control) task. The lineup task immediately followed the experimental/control task;
(b) The minimum required sample size for the study was reduced from 50 to 30 participants in each condition of RRR2 in order to accommodate smaller subject pools available in the spring semester at many universities, thereby permitting participation by more labs;
(c) When necessary, labs were permitted to use paid participant pools; funding was provided from the Association for Psychological Science (APS) via a grant from the Center for Open Science.

As noted above, S&E-S Study 1 used a 20-min filler task, but S&E-S Study 4 used a 10-min filler task. After we identified the error in the task ordering in RRR1, and in consultation with Schooler, we chose to maintain the 20-min filler task across RRR1 and RRR2 in order to make them directly comparable.

Online version of the protocol

In addition to the lab-based protocol adopted by all of the replicating teams, *Perspectives* solicited and APS funded an online version of the study that was conducted by one of the teams that had participated in RRR1 (Michael et al.). The participants for this study were drawn from Amazon Mechanical Turk, with each participant randomly assigned to the task order from RRR1 or RRR2 and to the verbal description or countries/capitals task. We chose to have one lab conduct a single large-scale online experiment

rather than having multiple labs conduct smaller experiments. This approach avoids a duplication of effort and the difficulty of ensuring that a Mechanical Turk participant did not complete multiple verbal overshadowing experiments. We also would not have been able to collect enough independent online replications to conduct a meta-analysis of the online-only studies, so we favored a single, larger-scale study. The results of the Mechanical Turk study were not included in the meta-analysis of the lab-based replications, but they are reported along with the lab results in all tables and figures.

In most respects, the Mechanical Turk study was identical to the lab-based ones: It used the same materials, the same timing and instructions, and the same measures. Due to the constraints of online testing, though, the following changes were made based on consultation between the editors and Schooler: (a) Participants were paid USD 2.00 for participation; (b) participants were excluded for reasons beyond those in the lab task, including a failure to list at least five countries/capitals in the control condition, a self-reported failure to engage appropriately with the filler task, having seen the robbery video before, or reporting participation in a study just like this one; (c) participation was limited to participants from the United States; (d) the crossword puzzle filler task was replaced with a set of Sudoku puzzles; and (e) participants were not given a reminder after 3 minutes to continue writing their description of the robber or listing countries/capitals.

Results

Lab demographics and results

Tables 1 and 2 provide demographic data for each lab, including the number of participants tested in each condition, the number who did not meet the demographic requirements or who were excluded for other reasons, and the number of the included participants who made each type of lineup selection (correct ID, mistaken ID, "not present"). For comparison, the tables include data from the original S&E-S studies. Note that some of the S&E-S numbers were reported in the original journal article and others were in Schooler's dissertation (those that were not reported and are no longer available are marked "NA").

Data analyses: Original and present

The S&E-S data analysis consisted of a χ^2 test comparing the frequency of correct and incorrect identification in the experimental and control conditions. A secondary analysis included a χ^2 comparing the types of errors (selecting the wrong face from the lineup or indicating "not present") across the conditions. Finally, the original study reported a 2 (Condition) × 2 (Correct vs. Incorrect/Miss) ANOVA on confidence ratings. Each lab conducted these analyses for the RRRs, and they are reported on the lab OSF project pages (URLs available along with each lab's project summary in the Appendix). Given that we have access to the full data set for each study, we used a more direct measure of the performance difference between conditions (*the risk difference*; the difference in percentage correct and the difference in percentages of error types) for the meta-analysis. We did not meta-analyze the ANOVAs of confidence ratings (the data are available).

"Verbal overshadowing" is defined as the difference in accuracy between the control condition (listing countries and capitals) and the verbal description condition (writing a description of the robber). But that difference can be measured in absolute or relative terms. The difference between 10% accuracy and 15% accuracy could be treated as a 5% increase in accuracy (the difference between the percentages) or it could be treated as a 50% increase in accuracy (15% accuracy is 1.5× as large as 10%). Note how these measures differ when the baseline accuracy is different: 50% and 55% accuracy still differ by 5%, but 55% is only 10% bigger in ratio terms (55 = 1.1 × 50). A ratio measure takes that baseline difference into account. When the baseline accuracy varies widely across studies or when the same difference in magnitude has different meanings (the difference between 50 and 55 has less importance than the difference between 5 and 10), ratio measures are more appropriate. But when accuracy levels are roughly comparable across studies and none are extreme, the raw difference between the percentages is more straightforward.

Given that accuracy levels in these studies were not extreme, we used "risk difference" as our measure of effect size for the meta-analyses: the percentage accuracy for the verbal description condition minus the percentage accuracy for the control condition. Negative effect sizes indicate a cost of verbally describing the robber.

Effect size measurements

For both RRRs, we provide a forest plot showing the accuracy percentages in each condition for each lab, the effect size measured by each lab (with 95% confidence intervals), and the meta-analytic effect size estimate in a random effects model. The top-most data point in each plot shows the effect from S&E-S, and the data point below that shows the effect found in the online Mechanical Turk variant of the study. Neither of those results are included in the meta-analytic effect size estimate at the bottom of each figure; the meta-analysis includes only the preregistered, lab-based replications of the original study. To permit a visual comparison of effects across the RRR studies, the plot for RRR1 identifies

Table 1. Sample Sizes and Data for RRR1

			Verbal Description Condition								Control Condition							
Authors	**Country of Participants**	**Language**	Total *N*	Excluded – Age	Excluded - Race	Excluded - Other	Total Included	Correct	False ID	Not Present	Total *N*	Excluded - Age	Excluded - Race	Excluded - Other	Total Included	Correct	False ID	Not Present
Original study: Schooler and Engstler-Schooler (1990), Study 4	USA	English	37	0	0	0	37	18	NA	NA	38	0	0	0	38	27	NA	NA
Online study (MTurk): Robert B. Michael, Gregory Franco, Mevagh Sanson, Maryanne Garry	USA	English	313	0	0	109	204	112	36	56	313	0	0	127	186	91	59	36
Victoria K. Alogna, Jamin Halberstadt, Jonathan Jong, Joshua C. Jackson, Cathy Ng	New Zealand	English	78	0	17	5	56	32	6	18	81	0	20	4	57	40	10	7
Stacy Birch	USA	English	77	0	13	8	56	37	7	12	79	0	21	8	50	33	13	4
Angela R. Birt, Philip Aucoin	Canada	English	53	0	2	0	51	17	14	20	52	0	2	0	50	18	19	13
Maria A. Brandimonte	Italy	Italian	70	0	0	0	70	34	18	18	70	0	0	0	70	27	32	11
Curt Carlson, Dawn Weatherford, Maria Carlson	USA	English	84	0	0	9	75	32	22	21	79	0	0	4	75	42	19	14
Kimberly S. Dellapaolera, Brian H. Bornstein	USA	English	86	0	12	0	74	29	23	22	86	0	10	0	76	41	20	15
Jean-Francois Delvenne, Charity Brown, Emma Portch, Tara Zaksaite	United Kingdom	English	63	0	2	2	59	40	9	10	67	0	0	6	61	40	13	8
Gerald Echterhoff, René Kopietz	Germany	German	54	10	0	3	41	15	9	17	53	4	0	3	46	24	15	7
Casey M. Eggleston, Calvin K. Lai, Elizabeth A. Gilbert,	USA	English	93	2	0	10	81	40	15	26	78	5	1	3	69	39	14	16
Daniel L. Greenberg, Marino Mugayar-Baldocchi	USA	English	59	0	9	0	50	24	11	15	60	0	10	0	50	28	12	10
Andre Kehn, Kimberly Schweitzer, Bradlee W. Gamblin, Kimberly Wiseman, Narina L. Nunez	USA	English	73	4	6	0	63	33	12	18	75	2	5	0	68	39	15	14
Christopher Koch, Remi Gentry, Jennifer Shaheed, Kelsi Buswell	USA	English	54	2	4	0	48	26	8	14	54	1	5	0	48	22	16	12
Nicola Mammarella, Beth Fairfield, Alberto Di Domenico	Italy	Italian	117	2	0	5	110	45	31	34	115	0	0	5	110	58	35	17
Shannon McCoy, Arielle Rancourt	USA	English	75	2	11	0	62	26	19	17	73	2	10	0	61	35	15	11
Abigail A. Mitchell, Marilyn S. Petro	USA	English	71	0	9	0	62	34	11	17	57	0	6	0	51	26	9	16
Robin Musselman, Michael Colarusso	USA	English	65	0	14	0	51	12	21	18	59	1	8	0	50	20	17	13

(continued)

Table 1. (continued)

Authors	Country of Participants	Language	Verbal Description Condition								Control Condition							
			Total *N*	Excluded – Age	Excluded - Race	Excluded - Other	Total Included	Correct	False ID	Not Present	Total *N*	Excluded - Age	Excluded - Race	Excluded - Other	Total Included	Correct	False ID	Not Present
Christopher R. Poirier, Matthew K. Attaya, Kathleen A. McConnaughy, Jessica E. Pappagianopoulos, Griffin A. Sullivan	USA	English	56	0	0	1	55	24	12	19	59	0	0	4	55	31	12	12
Eva Rubínová, Marek Vranka, Štěpán Bahník	Czech Republic	Czech	80	0	0	12	68	36	17	15	58	0	0	8	50	23	18	9
Kyle J. Susa, Jessica K. Swanner, Christian A. Meissner	USA	English	69	0	13	1	55	23	11	21	76	0	18	3	55	23	18	14
W. Burt Thompson	USA	English	66	2	11	3	50	28	11	11	65	1	14	0	50	21	18	11
Joanna Ulatowska, Aleksandra Cislak	Poland	Polish	59	0	0	4	55	40	4	11	68	2	0	11	55	38	7	10
Kimberley A. Wade, Ulrike Körner, Melissa, F. Colloff, Melina A. Kunar	United Kingdom	English	60	0	0	0	60	34	15	11	60	0	0	0	60	39	4	17
Simon Chu, John E. Marsh, Faye Skelton	United Kingdom	English	50	0	0	0	50	27	7	16	51	0	0	0	51	32	6	13
John E. Edlund, Austin Lee Nichols	USA	English	64	4	9	0	51	28	15	8	70	1	8	0	61	26	23	12
Fiona Gabbert, Tim Valentine	United Kingdom	English	83	6	22	0	55	34	6	15	83	9	21	0	53	31	7	15
Fábio P. Leite	USA	English	63	1	7	3	52	20	21	11	65	1	4	10	50	28	17	5
Alex H. McIntyre, Stephen R. H. Langton, Peter J. B. Hancock	United Kingdom	English	55	3	0	0	52	37	7	8	56	4	1	0	51	37	10	4
Robert B. Michael, Gregory Franco, Mevagh Sanson, Maryanne Garry	New Zealand	English	184	0	67	30	87	49	20	18	191	0	55	38	98	55	30	13
Matthew A. Palmer, Aaron Drummond, James D. Sauer, Daniel V. Zuj, Lauren Hall, Liam Satchell, Glenys Holt, Miriam Rainsford	Australia	English	65	4	5	0	56	32	13	11	58	6	1	0	51	30	11	10
Verkoeijen, P. P. J. L., Bouwmeester, S., Zwaan, R. A.	Netherlands	Dutch	56	0	1	4	51	26	14	11	59	0	5	0	54	30	12	12
Christopher A. Was, Dale Hirsch, Rachel Todaro, Connie Romig	USA	English	71	0	4	0	67	36	11	20	71	0	3	0	68	38	16	14

Table 2. Sample Sizes and Data for RRR2

			Verbal Description Condition								Control Condition							
Lab	**Country of Participants**	**Language**	Total *N*	Excluded - Age	Excluded - Race	Excluded - Other	Total Included	Correct	False ID	Not Present	Total N	Excluded - Age	Excluded - Race	Excluded - Other	Total Included	Correct	False ID	Not Present
Original study: Schooler and Engstler-Schooler (1990), Study 1	USA	English	44	0	0	0	44	17	16	11	44	0	0	0	44	28	10	6
Online study (MTurk): Robert B. Michael, Gregory Franco, Mevagh Sanson, Maryanne Garry	USA	English	302	0	0	98	204	94	47	63	313	0	0	130	183	104	49	30
Victoria K. Alogna, Jamin Halberstadt, Jonathan Jong, Joshua C. Jackson, Cathy Ng	New Zealand	English	70	0	20	0	50	25	7	18	67	0	17	0	50	31	15	4
Stacy Birch	USA	English	83	1	25	4	53	27	13	13	73	1	14	4	54	36	10	8
Angela R. Birt, Philip Aucoin	Canada	English	33	0	3	0	30	8	13	9	32	0	1	0	31	16	7	8
Maria A. Brandimonte	Italy	Italian	50	0	0	0	50	23	7	20	50	0	0	0	50	24	17	9
Curt Carlson, Dawn Weatherford, Maria Carlson	USA	English	81	4	0	2	75	32	26	17	79	3	0	1	75	50	12	13
Kimberly S. Dellapaolera, Brian H. Bornstein	USA	English	82	0	10	0	72	26	25	21	82	0	15	0	67	38	13	16
Jean-Francois Delvenne, Charity Brown, Emma Portch, Tara Zaksaite	United Kingdom	English	48	0	1	1	46	13	10	23	50	0	1	1	48	26	11	11
Gerald Echterhoff, René Kopietz	Germany	German	58	10	0	5	43	15	8	20	66	15	0	5	46	26	11	9
Casey M. Eggleston, Calvin K. Lai, Elizabeth A. Gilbert	USA	English	49	1	0	5	43	15	8	20	44	0	0	3	41	15	17	9
Daniel L. Greenberg, Marino Mugayar-Baldocchi	USA	English	37	0	7	0	30	15	7	8	38	0	7	1	30	19	8	3
Andre Kehn, Kimberly Schweitzer, Bradlee W. Gamblin, Kimberly Wiseman, Narina L. Nunez	USA	English	55	1	5	0	49	23	7	19	58	1	7	0	50	24	13	13
Christopher Koch, Remi Gentry, Jennifer Shaheed, Kelsi Buswell	USA	English	35	1	3	1	30	11	9	10	32	1	1	0	30	14	8	8
Nicola Mammarella, Beth Fairfield, Alberto Di Domenico	Italy	Italian	50	0	0	0	50	15	13	22	54	4	0	0	50	26	8	16
Shannon K. McCoy, Arielle Rancourt	USA	English	45	1	3	0	41	13	7	21	44	1	2	0	41	19	8	14
Abigail A. Mitchell, Marilyn S. Petro	USA	English	57	0	9	2	46	13	13	20	52	1	2	3	46	22	11	13
Robin Musselman, Michael Colarusso	USA	English	38	0	8	0	30	7	10	13	40	0	10	0	30	14	12	4
Christopher R. Poirier, Matthew K. Attaya, Griffin A. Sullivan, Kathleen A. McConnaughy, Jessica E. Pappagianopoulos	USA	English	46	0	1	1	44	13	13	18	49	0	0	8	41	24	9	8
Eva Rubínová, Marek Vranka, Štěpán Bahník	Czech Republic	Czech	56	0	0	4	52	16	19	17	54	0	0	4	50	17	20	13
Kyle J. Susa, Jessica K. Swanner, Christian A. Meissner	USA	English	53	0	3	0	50	11	10	29	58	0	8	0	50	23	13	14
W. Burt Thompson	USA	English	51	1	12	0	38	20	5	13	51	1	11	0	39	24	6	9
Joanna Ulatowska, Aleksandra Cislak	Poland	Polish	51	0	0	4	47	27	8	12	55	0	0	8	47	35	7	5
Kimberley A. Wade, Ulrike Körner, Melissa, F. Colloff, Melina A. Kunar	United Kingdom	English	61	0	0	1	60	26	19	15	60	0	0	0	60	36	19	5

the subset of labs that completed both studies and separates those from the subset that completed only the first study.

For RRR1 (Fig. 2), the meta-analysis showed a small effect of verbally describing the robber relative to listing countries and capitals. Whereas the original study showed a −22% difference between the verbal description condition and the control condition (verbal description − control), the meta-analytic effect across 31 larger scale replications was substantially smaller: −4.01% [95% confidence interval: −7.15% to −0.87%]. The original study had a larger absolute effect size than any of the replication studies, but that estimate also was the least precise because of its smaller sample size. All of the replication effect size estimates, including the online Mechanical Turk study, fell between −17.54% and 14.00%. The differences in the estimated effect size among the studies (i.e., heterogeneity) were consistent with what would be expected by chance ($\tau = 0$, $I^2 = 0\%$, $H^2 = 1.00$, $Q_{30} = 29.302$, $p = .502$).[3] Taken together, these studies reveal only a small effect of verbal descriptions on lineup accuracy when the task order required participants to provide their verbal description immediately after witnessing the crime video and then view the lineup after a 20-min delay (see also Finger & Pezdek, 1999 and Meissner & Brigham, 2001, for evidence that the verbal overshadowing effect is smaller with a delay between the description and lineup task).

For RRR2 (Fig. 3), the meta-analysis revealed a substantially larger effect of verbally describing the robber relative to listing countries and capitals. The original study showed a −25.00% difference between the verbal description condition and the control condition, and the meta-analysis of 22 studies showed a difference of −16.31% [95% confidence interval: −20.47% to −12.14%]. All 22 studies, as well as the online Mechanical Turk study, showed an effect in the same direction, with effect sizes ranging from −28.99% to −10.61%. The differences in the estimated effect size among the studies (i.e., heterogeneity) were entirely consistent with what would be expected by chance ($\tau = 0$, $I^2 = 0\%$, $H^2 = 1.00$, $Q_{21} = 15.25$, $p = .810$). Taken together, these studies reveal a robust and consistent effect of verbal descriptions on lineup accuracy when the task order requires participants to wait 20 min before providing the verbal description and then immediately try to identify the person they saw in a lineup.

When participants did not correctly select the robber from the lineup, they could make one of two types of error: selecting someone else from the lineup (false identification) or electing not to select anyone (miss). S&E-S Study 1 reported no difference in the proportion of errors that were false identifications between the verbal description condition and the control condition.[4] This breakdown of the errors into two categories was not reported in S&E-S Study 4, and those data are no longer available.

To explore the difference in error types across the replication studies, we considered only error trials and calculated the proportion of those trials for which participants selected the wrong person from the lineup. In RRR1, the meta-analysis showed reliably higher proportions of false identifications in the control condition than in the verbal description condition (Fig. 4). Across the 31 lab replication studies, the meta-analytic effect size was −11.53% [−16.36% to −6.70%]; negative numbers mean that the false alarm rate was larger in the control condition. The Mechanical Turk replication showed a difference of −22.97%, which is consistent with the pattern of the lab studies. The heterogeneity across studies was largely consistent with what would be expected by chance ($\tau = 0.0462$, $I^2 = 11.41\%$, $H^2 = 1.13$, $Q_{30} = 34.72$, $p = .253$). The difference appears to be driven by a greater tendency for participants in the verbal description condition to respond "not present."

This pattern was similar in RRR2 (Fig. 5), with a meta-analytic difference of −15.49% [−22.91% to −8.06%]. The Mechanical Turk study showed a similar effect of −19.29%. Although the overall pattern and size of the effect was consistent across studies, the results from individual labs were more heterogeneous in RRR2 ($\tau = 0.1113$, $I^2 = 39.77\%$, $H^2 = 1.66$, $Q_{21} = 34.06$, $p = .036$), ranging from a minimum of −50.95% to a maximum of 12.47%. Note, however, that the minimum required sample size in RRR2 was smaller than in RRR1, meaning that the effect size estimates from each lab are less precise.

Conclusions

The results of this large-scale, multiple-lab direct replication of S&E-S Study 4 and S&E-S Study 1 show that verbally describing the robber in a video can impair successful selection of that person from a subsequent lineup. The effect was larger when the verbal description happened immediately before the lineup selection than when it happened immediately after viewing the video. For RRR1, all of the replication studies produced a smaller effect size estimate than S&E-S Study 4, but the sample size in S&E-S was small enough that its large confidence interval included most of the replication studies. For RRR2, the original result from S&E-S Study 1 was close to that of the average replication study.

Although S&E-S reported no difference across conditions in proportion of errors that were false identifications as opposed to responding "not present," both replication studies found a robust difference, with a higher proportion of false identification errors in the control condition than in the verbal description condition;

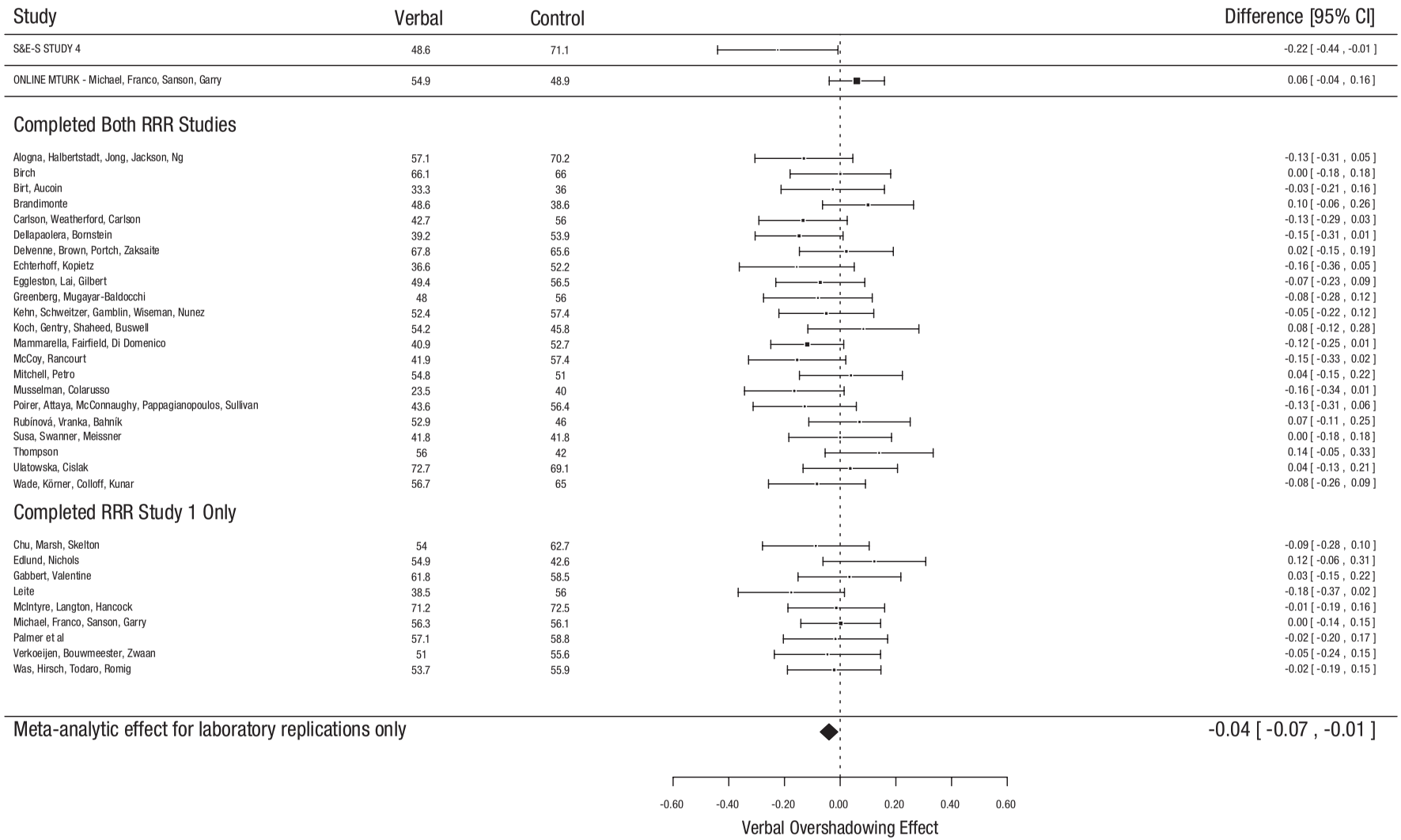

Fig. 2. Forest plot of the verbal overshadowing effect in RRR1, with negative effects indicating lower accuracy for participants who verbally described the robber (Verbal – Control). The data are listed in alphabetical order by the name of the first author from each replicating team. For each team, the figure shows the percentage correct for the verbal description condition and the control condition and a forest plot of the difference in proportions correct (bigger effect size markers reflect bigger samples). The Difference column provides the values used in the forest plot. The study replicates the task ordering from S&E-S Study 4.

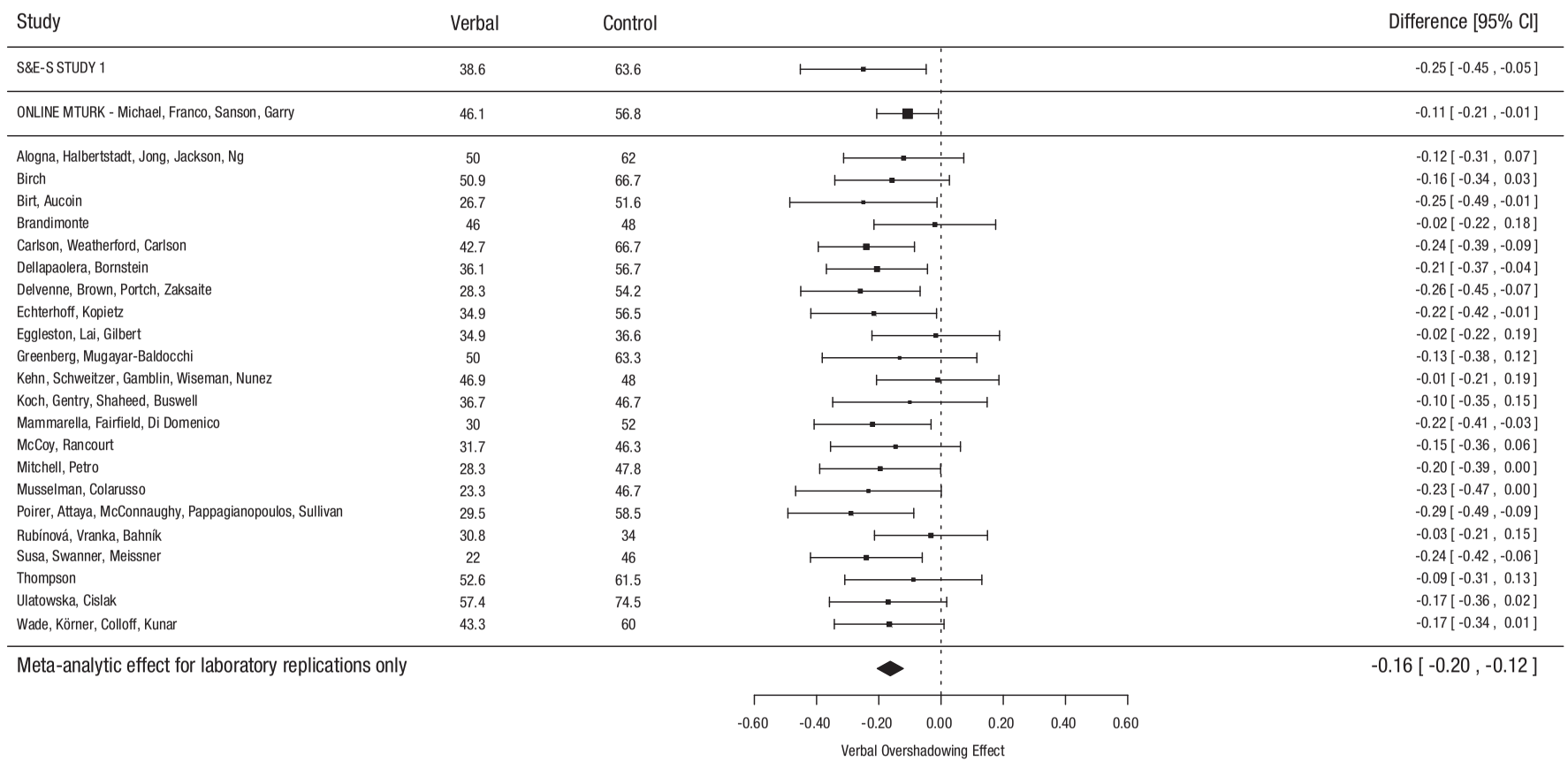

Fig. 3. Forest plot of the verbal overshadowing effect in RRR2, with negative effects indicating lower accuracy for participants who verbally described the robber (Verbal – Control). The data are listed in alphabetical order by the name of the first author from each replicating team. For each team, the figure shows the percentage correct for the verbal description condition and the control condition and a forest plot of the difference in proportions correct (bigger effect size markers reflect bigger samples). The Difference column provides the values used in the forest plot. The study replicates the task ordering from S&E-S Study 1.

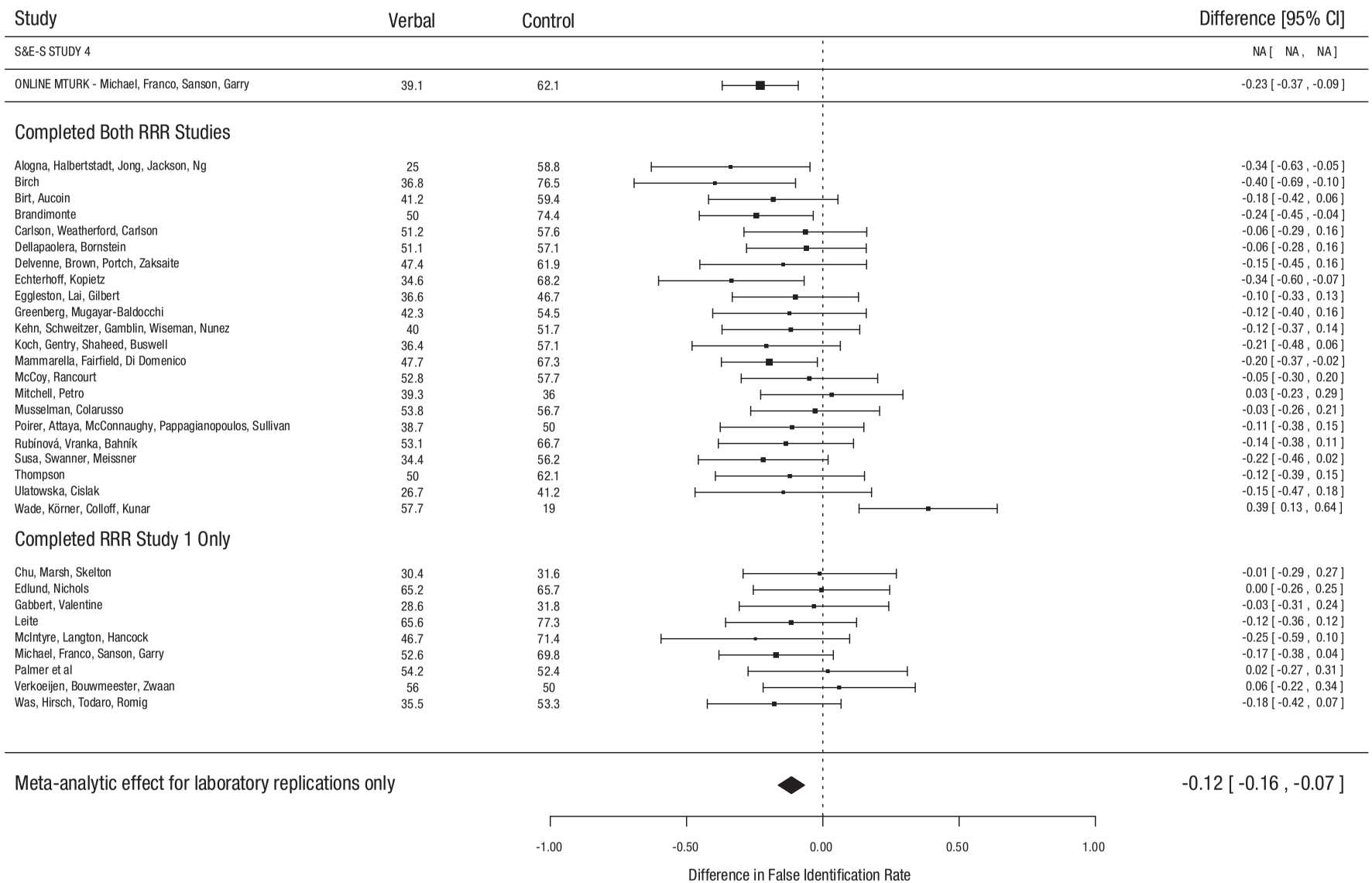

Fig. 4. Forest plot of the difference between the verbal description condition and the control condition in the proportion of error trials that were false identifications in RRR1. Negative effects (Verbal – Control) constitute evidence that people who verbally described the robber were proportionally more likely to select "not present" (they were relatively less likely to make the error of selecting the wrong person from the lineup). The data are listed in alphabetical order by the name of the first author from each replicating team. For each team, the figure shows the percentage of errors that were false identifications for the verbal description condition and the control condition and a forest plot of the difference in the proportion of false identifications (bigger effect size markers reflect bigger samples). The Difference column provides the values used in the forest plot. The study replicates the task ordering from S&E-S Study 4.

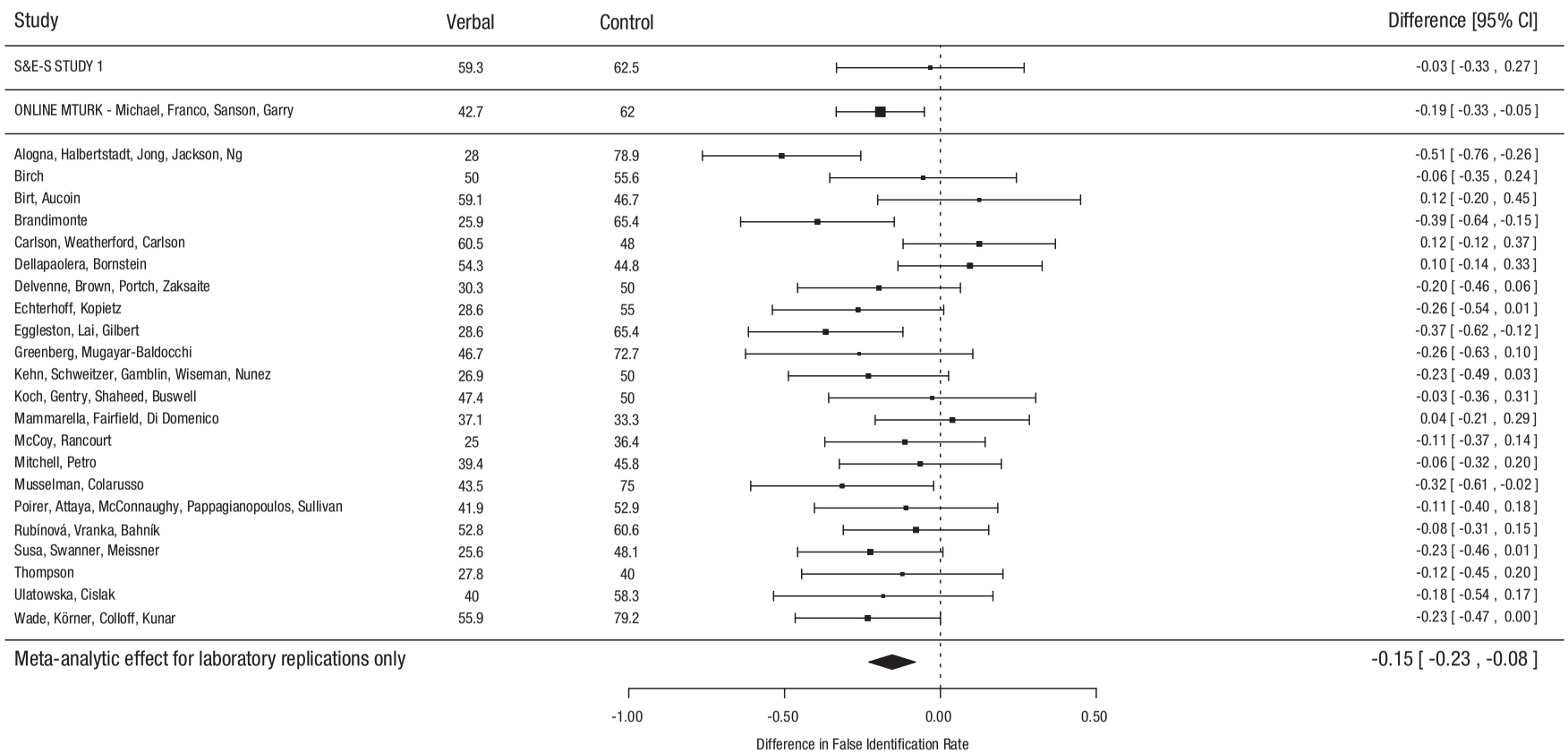

Fig. 5. Forest plot of the difference between the verbal description condition and the control condition in the proportion of error trials that were false identifications in RRR2. Negative effects (Verbal – Control) constitute evidence that people who verbally described the robber were proportionally more likely to select "not present" (they were relatively less likely to make the error of selecting the wrong person from the lineup). The data are listed in alphabetical order by the name of the first author from each replicating team. For each team, the figure shows the percentage of errors that were false identifications for the verbal description condition and the control condition and a forest plot of the difference in the proportion of false identifications (bigger effect size markers reflect bigger samples). The Difference column provides the values used in the forest plot. The study replicates the task ordering from S&E-S Study 1.

participants in the verbal description condition were more likely to say "not present" than were those in the control condition. This difference in the types of errors across conditions might reflect a difference in the response bias—the willingness to select someone from the lineup—induced by the critical manipulation (see Clare & Lewandowsky, 2004; see also Chin & Schooler, 2008 for further discussion). Alternatively, the pattern of errors might reflect a memory distortion caused by eliciting a verbal description. Further studies that include both target present and target absent lineups could help distinguish between these alternatives.

Effect of delay

The only published meta-analysis of the verbal overshadowing effect (Meissner & Brigham, 2001) found that the effect of providing a verbal description is reduced with a delay between providing the description and completing the lineup identification task (see also Finger & Pezdek, 1999). The present results are consistent with that conclusion. The studies included in the meta-analysis varied in the length of the delay and the materials used, meaning that the inference of a delay effect depended on averaging across a number of other differences among the studies. The comparison of RRR1 to RRR2 provides more compelling support for the conclusion that, keeping all other aspects of the protocol constant, task order alone moderates the effect of providing a verbal description on lineup accuracy.

Future research is needed to better understand the memory process responsible for this difference. Switching the task order affects two aspects of the design: the delay between witnessing the robbery and providing a description and the delay between providing a description and completing the lineup identification task. Because the lengths of these delays are confounded, it is impossible to separate the contributions of these two types of delay to the difference between the two studies. Future research could systematically vary the delays between the witnessed event, the description tasks, and the identification task to see which contribute to the change in the effect of verbally describing the robber.

A better understanding of how each type of delay affects lineup identification accuracy would be of both theoretical and practical importance. If verbally describing a person impairs subsequent lineup identification accuracy under some delays but not others, then those differences could inform police practices. For example, perhaps the effect of providing a verbal description depends critically on having the lineup identification task immediately follow the description. If so, then verbal overshadowing would have less practical relevance: In most cases, the verbal description witnesses provide to police does not immediately precede the lineup task. However, if the effect instead depends only on the presence of a delay between witnessing an event and describing the suspect, then the verbal overshadowing effect could have broad practical importance: Witnesses rarely provide a verbal description immediately after witnessing a crime, so verbal overshadowing could come into play in most eyewitness situations.

Reliability of effect size over time

One motivation for this RRR was the claim that the verbal overshadowing effect had declined in size over the past 20 years, the so-called "decline effect" (Schooler, 2011). Assessing whether or not an effect has diminished in size depends critically on having a robust estimate of the effect size, both initially and later. The RRR was designed to provide a robust estimate of the effect, one that could be compared with that from the original study. However, the original studies used small samples, meaning that the estimates were not precise. For example, the confidence interval around the effect size for S&E-S Study 4 ranged from −44% to −0.79%. Although that original effect size estimate falls outside the confidence intervals of our meta-analytic effect size for that study, it is unclear whether the effect actually declined in size or whether the original estimate was just an inaccurate estimate of the effect. Moreover, RRR2 produced a meta-analytic effect size that was in line with that of the original study, providing no compelling evidence for a change in the true effect over time with that task order.

By providing a precise meta-analytic estimate of the true effect size, the RRR studies provide guidance on the sample sizes needed to reliably detect the effect of providing a verbal description on lineup identification performance. An analysis of the sample sizes of earlier verbal overshadowing studies suggested that they were, on the whole, substantially underpowered (Francis, 2012). The results of this RRR are largely consistent with that conclusion. Only by combining across many larger scale studies could we detect the effect of providing a verbal description in RRR1. The confidence intervals around an individual lab's effect size estimate are large (see the intervals around individual lab studies in Figures 1 and 2—even those studies with the largest samples do not provide a highly precise estimate of the effect size). Even the Mechanical Turk study, with nearly 200 participants in each condition, produced a confidence interval with a range of approximately 12%. In other words, it could not have reliably detected a significant difference from no effect with a true effect size of about 4%. In fact, all of the confidence intervals for the individual replications in RRR1 included 0. Had we simply tallied the number of studies providing clear evidence for an effect in RRR1, we

would have concluded in favor of a robust failure to replicate—a misleading conclusion. Moreover, our understanding of the size of the effect would not have improved. The purpose of the RRR approach is to better understand the true size of important effects and not to make global succeed/fail judgments about individual replication studies.

Summary

RRR1 and RRR2 combine the results of multiple, independent, direct replications to determine the size of the verbal overshadowing effect. In doing so, they provide clear evidence for verbal overshadowing, particularly in the original task order used by S&E-S Study 1 (description after a delay and just before line-up). Moreover, the effect size estimates the RRRs provide can guide future research on verbal overshadowing, both by suggesting new experimental questions and by indicating the sample sizes needed to test those questions.

Appendix: Individual Lab Details

Below, each lab briefly describes the characteristics of their sample and notes any substantial departures from the standard protocol. Each lab description identifies the authors and their affiliations in the order of their contributions to the project. Each also provides a link to that lab's OSF project page for the study where readers can see all of the details of the study including more complete method and results descriptions as well as the raw data. Labs are listed in the same order as in the tables and figures.

Amazon Mechanical Turk variant

Robert B. Michael, Victoria University of Wellington
Gregory Franco, Victoria University of Wellington
Mevagh Sanson, Victoria University of Wellington
Maryanne Garry, Victoria University of Wellington
OSF: https://osf.io/ez4w3/

For both RRR Study 1 and 2, participants were recruited from Amazon Mechanical Turk (MTurk) and were paid USD 2.00 Participants were randomly assigned to Study 1 or Study 2 and to the Control condition or the Experimental condition. We first collected 800 subjects with no prescreened exclusion criteria. We then collected an additional 350 subjects with prescreened exclusion criteria. These participants were required to self-report race as White and age between 18 and 25 to be eligible. We used custom software (see Turkitron.com) to track MTurk workers, preventing subjects from taking the study multiple times. All participants were from the United States. We tracked and excluded participants who did or reported any of the following: (a) failed to complete the experiment, (b) failed to follow instructions, (c) failed an attention check, (d) failed to give at least five countries and capitals in the control condition, (e) failed to engage appropriately with the filler task, (f) had seen the robbery video before, or (g) had already participated in a study just like this one.

Because the experiment was run online, subject behavior was not subject to the same degree of control as a lab-based experiment. Specifically, MTurk workers have the freedom to engage in other tasks or communicate with other people. We aimed to reduce this undesirable activity by providing instructions to MTurk workers before they began the experiment. These instructions asked that workers complete the experiment in an environment free from distraction, that they give the experiment their full attention, and that they have functioning audio. We also followed these instructions up with a series of questions at the end of the experiment. These questions asked whether the worker did in fact follow the instructions, with the assurance that they would receive compensation regardless of their answers.

We also embedded an attention check question. This question requested that subjects select "No" as their response to the question and that they remember the word "horse" to be entered on the following page. If subjects selected "Yes" as their response or failed to enter the word "horse", they were tagged for exclusion. At the end of the experiment, we asked participants whether they had seen the video of the robbery before and if they had participated in a study like this one before. A response of "Yes" to either of these questions resulted in an exclusion tag. Our filler task was a series of Sudoku puzzles. We asked subjects at the end of the experiment whether they gave this task their full attention. A response of "No" to this question resulted in an exclusion tag. Because of technical limitations, we did not give our subjects a reminder at the 3-min mark of the experimental or control task. Our procedures, other than the deviations listed above, followed the approved protocol.

Labs that completed both RRR1 and RRR2

Victoria K. Alogna, University of Otago
Jamin Halberstadt, University of Otago
Jonathan Jong, Institute of Cognitive and Evolutionary Anthropology, University of Oxford
Joshua C. Jackson, University of Otago
Cathy Ng, University of Otago
OSF: https://osf.io/sqzuq/

For RRR Study 1, participants were 159 first and second year psychology undergraduates at the University of Otago. One hundred and twenty-two took part during the school term, in exchange for course credit; the remainder took part after their classes had ended and were remunerated NZ$15 for travel expenses. Eight participants were excluded due to computer software failures, and one participant did not complete the experiment. Of the remaining 150 participants (53 men, 97 women, mean age = 20.14, *SD* = 1.71), 113 reported their race

as "European" (i.e., Caucasian). Per our registered experimental protocol, the analyses are based only on these participants, though all data are available online. An additional 75 participants contributed data to a third "suppression" condition, in which participants were instructed to try not to think about the robber's face. This condition is described in more detail on our OSF page.

For RRR Study 2, participants were 137 first and second year psychology undergraduates (43 men, 94 women, mean age = 20.43, *SD* = 3.99) at the University of Otago. One hundred and thirty-one of these students took part during the school term in exchange for course credit; the remainder were remunerated NZ$15 as reimbursement for travel expenses. One hundred reported their race as "European" (i.e., Caucasian). Per our registered experimental protocol, the analyses are based only on these participants, though all data are available online. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Stacy L. Birch, College at Brockport SUNY
OSF: https://osf.io/9zu4g/
For RRR Study 1, participants were recruited from the introductory psychology participant pool at the College at Brockport, State University of New York. They participated as part of one option for course credit. Of the 156 participants in Study 1, 106 met inclusion criteria for the study according to their responses on the demographic form. For RRR Study 2, 159 participants were recruited from the introductory psychology pool at the College at Brockport (none of whom had participated in Study 1). All received participation credit, but only 107 met inclusion criteria for the study based on responses to the demographic questions. All data are available on the OSF page for our studies. Our procedures followed the approved protocol and did not deviate from the preregistered plan.

Angela R. Birt, Mount Saint Vincent University
Philip Aucoin, Mount Saint Vincent University
OSF: https://osf.io/y3xtf/
For RRR Study 1, participants were recruited from undergraduate courses at Mount Saint Vincent University in Halifax, Nova Scotia, Canada. At the discretion of course instructors, most (92.30%) received course points for participating. Participants in RRR Study 2 were recruited in the same way as Study 1 but were compensated $8 for participating. Recruitment in both studies did not include restrictions on race or age; therefore, the overall samples included participants who did not meet inclusion criteria for this RRR. Data from participants excluded from analyses are included on our OSF page. Following our preregistered plans for both studies, we included a self-report question of visual acuity, a brief test to measure potential effects of demand characteristics on the results, and collected data on reaction times for making identifications. Otherwise, our procedures followed the standard protocols. Analyses of the additional data can be found on our OSF page.

Maria A. Brandimonte, Suor Orsola Benincasa University of Naples
OSF: https://osf.io/gfyyd/
For RRR Study 1, 140 participants were recruited from the participant pool at Suor Orsola Benincasa University of Naples in Italy, and they participated as part of one option for course credit. None of them was enrolled in a psychology course. For RRR Study 2, 100 participants were recruited. All participants were White. All participants were able to understand the instructions and had vision adequate to watch the video and see the images. Given that our participants were not native English speakers, instructions were translated into Italian and then translated back into English independently by the two labs participating in this replication effort from Italy (ours and a lab at the University of Chieti). We also replaced the English audio track with an Italian translation. Both laboratories used the same translated materials and dubbed video on which there had been full agreement. Finally, we added an additional question at the end of the study to verify that participants understood that the video depicted a bank robbery. No participant was excluded. In all other respects, our procedure followed the standard protocol.

Curt A. Carlson, Texas A&M University - Commerce
Dawn R. Weatherford, Arkansas State University
Maria A. Carlson, Texas A&M University - Commerce
OSF: https://osf.io/s73uq/
For RRR Study 1, participants were recruited from the psychology participant pool at Texas A&M University—Commerce, and they participated as part of one option for course credit. For RRR Study 2, participants were recruited from the psychology participant pool at Arkansas State University, and they also participated for course credit. For each participant pool, we used a prescreening process that allowed us to recruit only participants who met the specified inclusion criteria, so only participants who failed to complete the study were excluded. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Kimberly S. Dellapaolera, University of Nebraska-Lincoln
Brian H. Bornstein, University of Nebraska-Lincoln
OSF: https://osf.io/qrz2g/
For RRR Study 1, participants were recruited from the psychology participant pool at the University of Nebraska–Lincoln, and they participated as part of one option for course credit. For RRR Study 2, participants were recruited from the same participant pool; participants could only participate in one of the two studies. We recruited from our participant pool without specifying restrictions on race or age, so our total sample for Study 1 included an additional 22 participants and our total sample for Study 2 included an additional 25 participants who did not meet inclusion criteria. All data are provided on our OSF page. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Jean-Francois Delvenne, University of Leeds
Charity Brown, University of Leeds
Emma Portch, University of Leeds
Tara Zaksaite, University of Leeds
OSF: https://osf.io/vucan/
For RRR Study 1, 93 participants were recruited from the participant pool at the University of Leeds (they participated as part of one option for course credit) and 37 participants were recruited from the broad campus and were compensated £5 for participating. For RRR Study 2, 43 participants were recruited from our participant pool and 55 participants were recruited from the broad campus and were compensated £5 for participating. For our participant pool, we used a prescreening process that allowed us to recruit only participants who met the specified inclusion criteria, so only participants who failed to complete the study (i.e., 10 in Study 1; 4 in Study 2) were excluded. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Gerald Echterhoff, University of Münster
René Kopietz, University of Münster
OSF: https://osf.io/dmuqj/
For RRR Study 1, participants were recruited from the introductory psychology participant pool at the University of Münster in Germany, and they participated as part of one option for course credit. For RRR Study 2, approx. 20% of the participants were recruited from the participant pool and the remaining 80% were recruited from the broader campus and were compensated €6 for participating. We recruited from our participant pool and community without specifying restrictions on race or age, so our total sample included an additional 55 participants—20 in Study 1 and 35 in Study 2—who did not meet inclusion criteria for this RRR. Data from those participants are included on our OSF page. Additionally, data on the OSF page includes participants who did not understand the nature of the event as well as a sample ($N = 36$) from our initial attempt to run Study 1 with the original English-language version of the video.

Given that our participants were not native English speakers, the second author translated all instructions to German, and a bilingual student assistant independently translated them back to English to verify the accuracy of the translation. Based on a small, informal pretest, we initially assumed that our participants would be able to understand the video with the original sound track and therefore did not dub it. However, we added an additional question at the end of the study to verify that participants understood that the video depicted a bank robbery.

Because many participants' did not understand the nature of the event depicted in the original video, we changed the protocol to replace the English audio track with a German translation. We informed the editors about this modification and excluded all participants who watched the original version of the video from the final sample. Based on our preregistered plan, we excluded any participants who did not understand the nature of the video. Because of the need for this change to Study 1, we were unable to reach the preregistered 50 participants per condition (final sample: $n = 46$ in the control and $n = 41$ in the experimental condition). Similarly, because of the need to mainly recruit participants outside the psychology department for Study 2, we did not reach our goal of 50 participants per condition after exclusion due to age or failure to understand the nature of the event (final sample: $n = 46$ in the control and $n = 43$ in the experimental condition). In all other respects, our procedure followed the standard protocol.

Casey M. Eggleston, University of Virginia
Calvin K. Lai, University of Virginia
Elizabeth A. Gilbert, University of Virginia
OSF: https://osf.io/b4g79/
For RRR Studies 1 and 2, participants were recruited from the introductory psychology participant pool at the University of Virginia, and they participated as part of one option for course credit. Of the 180 participants who partook in Study 1, 25 were excluded prior to data analysis based on a priori criteria (e.g., failing to meet the target study demographics, improperly answering the attention catch question), and 5 participants were unexpectedly excluded for failing to sign a proper consent form. Of the 94 participants who partook in Study 2, 10 were excluded prior to data analysis based on a priori criteria. Data from all participants who completed the study and gave informed consent are provided on our OSF page. Our procedures otherwise followed the approved protocol and did not deviate from our preregistered plan.

Daniel L. Greenberg, College of Charleston
Marino A. Mugayar-Baldocchi, College of Charleston
OSF: https://osf.io/sieea/
For both RRR Study 1 and RRR Study 2, participants were recruited from the introductory psychology participant pool at the College of Charleston, and they participated as part of one option for course credit. For both studies, recruitment was conducted without specifying restrictions on race or age, so our total sample included participants who did not meet inclusion criteria for this RRR (19 in Study 1 and 14 in Study 2). Data from those participants are included on our OSF page but were excluded from the analyses reported here. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Andre Kehn, University of North Dakota
Kimberly Schweitzer, University of Wyoming
Bradlee W. Gamblin, University of North Dakota
Kimberly Wiseman, University of Wyoming
Narina L. Nunez, University of Wyoming
OSF: https://osf.io/mkz84/
For RRR Study 1 and 2, participants were recruited from the psychology participant pools at the University of North Dakota and the University of Wyoming, and they participated to receive course credit or extra credit. We oversampled for both studies in

order to reach the minimum participant numbers. Participants were excluded if they did not meet the age or race requirements (n = 10 in Study 1, n = 7 in Study 2). Further, participants were also excluded if they did not complete the study. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Christopher Koch, George Fox University
Remi Gentry, George Fox University
Jennifer Shaheed, George Fox University
Kelsi Buswell, George Fox University
OSF: https://osf.io/bym2a/
For RRR Study 1, participants were recruited from general psychology courses at George Fox University for research participation credit. A total of 109 participants completed the study. However, 13 participants were removed from the analysis for not meeting the RRR inclusion criteria. The remaining participants (62 females and 34 males) were equally divided between the control and experimental conditions. Participant age ranged between 18 to 23 years with a mean of 19.27 (SD = 1.14). For RRR Study 2, 46 participants were recruited from general psychology courses for research participation credit. An additional 21 volunteers were recruited by asking participants who had just completed the study to suggest other people who might be willing to volunteer ("snowball" recruiting). Five participants were removed from the analysis for not meeting the RRR inclusion criteria, and two were removed for invalid responses. The remaining participants (45 females and 15 males) were equally divided between the control and experimental conditions. Participant age ranged between 18 to 23 years with a mean of 20.08 (SD = 1.83). Data from all participants are reported on our OSF page. Other than the use of snowball recruiting to meet the specified sample size for Study 2, our procedure for both studies followed the approved protocol and did not deviate from our preregistered plan.

Nicola Mammarella, University of Chieti
Beth Fairfield, University of Chieti
Alberto Di Domenico, University of Chieti
OSF: https://osf.io/edsrz/
For both RRR Study 1 and 2, participants were recruited from an introductory psychology course participant pool at the University of Chieti in Italy, and they participated for course credit. In both studies, we recruited without specifying restrictions on race or age. Of the 232 participants in Study 1, 12 did not meet the inclusion criteria. Of the 104 participants in Study 2, 4 did not meet the exclusion criteria. Data from all participants are provided on our OSF page. Given that our participants were not native English speakers, one of the authors translated all instructions to Italian, and a second author independently translated them back to English to verify the accuracy of the translation. We also replaced the English audio track with an Italian translation. In all other respects, our procedure followed the standard protocol.

Shannon K. McCoy, University of Maine
Arielle Rancourt, University of Maine
OSF: https://osf.io/ejj7d/
For RRR Study 1 and Study 2, participants were recruited from the introductory psychology participant pool at the University of Maine, and they participated for course credit. We recruited from our participant pool without specifying restrictions on race or age, so our total sample included an additional 32 participants who did not meet inclusion criteria for the RRR (n = 25 from Study 1; n = 7 from Study 2). Data from those participants are included on our OSF page. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Abigail A. Mitchell, Nebraska Wesleyan University
Marilyn S. Petro, Nebraska Wesleyan University
OSF: https://osf.io/zqnjb/
For both RRR Study 1 and Study 2, participants were recruited from the Nebraska Wesleyan Psychology Department's participant pool. Students participated as part of one option for course credit. We recruited without specifying age or race restrictions. For Study 1, 128 were recruited, however 15 did not meet the inclusion criteria. For Study 2, 109 participated, but data from 17 were excluded due to not meeting inclusion criteria. Our procedures followed the approved protocol and did not deviate from our preregistered plan, which included open-ended debriefing questions concerning perceptions of the study.

Robin Musselman, Lehigh Carbon Community College
Michael Colarusso, Lehigh Carbon Community College
OSF: https://osf.io/ybfmu/
For RRR Study 1, 101 participants were recruited from Introduction to Psychology courses at Lehigh Carbon Community College, and they participated in most cases for extra credit in their course (whether participants received course credit was determined by the course instructors and was not under the experimenters' control). For RRR Study 2, 60 participants were recruited from Lehigh Carbon Community College, and 15 were recruited from Cedar Crest College, with students receiving extra credit for participating. We recruited without specifying restrictions on race, so our total sample included an additional 41 participants (23 in Study 1 and 18 in Study 2) whose data are reported on our OSF page. Our procedures followed the protocol and we did not deviate from our preregistered plan, with the exception of recruiting at a neighboring college to meet our specified sample size for Study 2.

Christopher R. Poirier, Stonehill College
Matthew K. Attaya, Stonehill College
Kathleen A. McConnaughy, Stonehill College
Jessica E. Pappagianopoulos, Stonehill College
Griffin A. Sullivan, Stonehill College
OSF: https://osf.io/zgmex/

For RRR Study 1 and Study 2, participants were recruited from the psychology department's participant pool at Stonehill College, and they participated as part of one option for course credit. We used a prescreening process that allowed us to recruit only participants who met the specified inclusion criteria; however, a participant in Study 2 was excluded because he identified as both White and Black during the study. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Eva Rubínová, Masaryk University
Marek A. Vranka, Charles University in Prague
Štěpán Bahník, University of Würzburg
OSF: https://osf.io/ikuh7/
For both RRR Studies, participants were recruited from our laboratory subject pool consisting of students of Czech universities, and they were compensated 100 CZK (approx. $5) for participation. Our participant database allows us to use a prescreening process, so we invited only participants who met the specified inclusion criteria. Given that our participants were not native English speakers, the authors translated all instructions to Czech, and an independent bilingual speaker translated them back into English to verify the accuracy of the translation. We also replaced the English audio track with its Czech translation and used a Czech crossword puzzle similar to the one used in the original study. We did not include any comprehension checks as all of our participants were native or fluent Czech speakers. The study was run on computers (except for the crossword puzzle and robber description/capitals listing, which were completed on paper), and we added some procedural instructions to be able to run the study without additional instructions from the experimenter during the main part of the session. Following our preregistered plan, we added three questions at the end of the session to check participants' knowledge of the experiment. In Study 1, based on the answers, we excluded 15 participants who stated that they (a) knew about the experimental procedure or hypothesis from other participants, (b) knew the tested hypothesis, or (c) knew what the verbal overshadowing effect is; 5 more participants were excluded because of technical issues. In Study 2, participants also had to write down the hypothesis and/or what verbal overshadowing effect is, and we excluded only those who answered correctly (n = 7 excluded). One participant was excluded because she did not provide any description of the robber. In all other respects, our procedure followed the standard protocol.

Kyle J. Susa, University of Texas at El Paso
Jessica K. Swanner, Iowa State University
Christian A. Meissner, Iowa State University
OSF: https://osf.io/5vunt/
For RRR Studies 1 and 2, participants were recruited from the introductory psychology participant pool at Iowa State University, and they participated as part of one option for course credit. For RRR Study 1, 145 participants were recruited. For RRR Study 2, 111 participants were recruited. In accordance with IRB approval, we did not restrict our participants by race or age—however, only participants who met the inclusion criteria were evaluated in our analyses. In RRR Study 1, 35 participants did not meet the inclusion criteria, in RRR Study 2, 11 participants did not meet the inclusion criteria. Data from all participants are reported on our OSF page. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

W. Burt Thompson, Niagara University
OSF: https://osf.io/4ijas/
For RRR Study 1, participants were recruited from psychology classes at Niagara University, and they participated in return for course credit. The primary sample consists of the first 100 participants, 50 per condition, who met all criteria for inclusion in the study. An additional 31 participants either did not meet one or more of the study criteria (e.g., age, ethnicity) or were tested after the primary data set had been collected. For RRR Study 2, the primary sample consists of the first 77 students who met all of the criteria for the study: 38 in the description (experimental) condition, and 39 in the capitals (control) condition. An additional 25 students were tested but did not meet all study criteria. All participants were recruited from Niagara University psychology and criminal justice classes. Fifteen of the participants were compensated $5, and the others received course credit. For both studies, our procedures followed the approved protocol and did not deviate from our preregistered plan.

Joanna Ulatowska, Academy of Special Education, Warsaw, Poland
Aleksandra Cislak, University of Social Sciences and Humanities, Warsaw, Poland
OSF: https://osf.io/bzhvf/
For RRR Study 1, participants were recruited among social sciences students through study advertisements and personally by research assistants at the campus of Academy of Special Education in Poland. For RRR Study 2, participants were recruited at the campuses of Academy of Special Education and University of Social Sciences and Humanities in Poland. They participated in return for a gift voucher (25 PLN, approximately $8.16). We only recruited undergraduate students ages 18 to 25 who claimed to speak English. Given that our participants were not native English speakers, all of the instructions were translated to Polish by one of the experimenters and then translated back to English by a fluent English speaker. The independent translator was blind to the study topic. All Polish instructions were also verified using Google Translate. At the end of the study, we asked an additional question to verify that participants understood that the video depicted a bank robbery. Based on our preregistered plan, we excluded any participants who did not understand the nature of the video. In Study 1, 15 participants (10 women) were excluded from further analyses as they did not understand the sense of robber's words, and two

more women were excluded as they exceeded the age limit. In Study 2, 12 participants (11 women) were excluded from further analyses as they did not understand the sense of robber's words. Data from those participants are included on our OSF page. In all other respects, our procedure followed the standard protocol.

Kimberley A. Wade, University of Warwick
Ulrike Körner, Heinrich-Heine-University Düsseldorf
Melissa F. Colloff, University of Warwick
Melina A. Kunar, University of Warwick
OSF: https://osf.io/dbxv4/
For RRR Study 1, 68 of the participants were recruited from the introductory psychology participant pool at the University of Warwick, and they participated as part of one option for course credit. The other 52 participants were recruited from the broader campus via a university-wide participant pool and were compensated £3. For RRR Study 2, participants were recruited from across the University of Warwick campus via the university-wide participant pool and were compensated £3. For our participant pool, we used a prescreening process that allowed us to recruit only participants who met the specified inclusion criteria, so only 1 participant in Study 2 who failed to complete the study was excluded. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Labs that completed only RRR1

Simon Chu, Ashworth Research Centre
John E. Marsh, University of Central Lancashire
Faye Skelton, University of Central Lancashire
OSF: https://osf.io/qu3zp/
For RRR Study 1, 79 participants were recruited from the undergraduate psychology participant pool at the University of Central Lancashire and participated as one option in return for course credit. Twenty-two participants from across the broader university campus were also recruited through a university online bulletin board. Participants recruited from outside the psychology department volunteered their time. We used a prescreening process that allowed us to recruit only participants who met the specified inclusion criteria, so only participants who failed to complete the study were excluded. Owing to time constraints, we were forced to close the study before meeting our original recruitment target of 120. Our experimental procedure followed the approved protocol.

John E. Edlund, Rochester Institute of Technology
Austin Lee Nichols, University of Navarra
OSF: https://osf.io/ybswb/
For RRR Study 1, participants were recruited from the introductory psychology participant pool at the Rochester Institute of Technology, and they participated as part of one option for course credit. Due to our limited participant pool, we were unable to complete RRR Study 2. We recruited from our participant pool without specifying restrictions on race or age, so our total sample for Study 1 included an additional 22 participants who did not meet the inclusion criteria for this RRR. Data from those participants are included on our OSF page. Our total included sample consisted of 61 participants in the control condition and 51 participants in the experimental condition. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Fiona Gabbert, Goldsmiths University of London
Tim Valentine, Goldsmiths University of London
OSF: https://osf.io/rmdz7/
For RRR Study 1, participants were recruited as part of their Research & Methods laboratory class on χ^2 analysis at Goldsmiths University of London. They were not required to take part, but everyone did. No compensation was given. We recruited without specifying restrictions on race or age, so our total sample included an additional 58 participants who did not meet inclusion criteria for this RRR. Our procedures followed the approved protocol and did not deviate from our preregistered plan. Due to having tested all of our first year psychology students for Study 1, we were unable to complete RRR Study 2.

Fábio P. Leite, Ohio State University at Lima
OSF: https://osf.io/kmibs/
For RRR Study 1, 128 participants were recruited from the introductory psychology participant pool at the Ohio State University at Lima, and they participated as part of one option for course credit. We recruited from our participant pool without specifying restrictions on race or age. Twenty six participants did not meet inclusion criteria for this RRR, and their data are included on our OSF page. Due to our limited participant pool, we were unable to complete RRR Study 2. The incomplete data set for Study 2 is also included on our OSF page. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Alex H. McIntyre, University of Stirling
Stephen R. H. Langton, University of Stirling
Peter J. B. Hancock, University of Stirling
OSF: https://osf.io/3rn5f/
For RRR Study 1, 103 participants were recruited from the introductory psychology participant pool at the University of Stirling in Scotland, and they participated as part of one option for course credit. A further 7 participants were excluded due to age criteria, and 1 was excluded in line with race criteria. Data from the excluded participants are included on our OSF page. For RRR Study 2, we were unable to recruit the required sample of 30 participants in each group and just 24 participants were recruited from the participant pool. Data from all participants are available on our OSF page. Our procedures followed the approved protocols and did not deviate from our preregistered plan.

Robert B. Michael, Victoria University of Wellington
Gregory Franco, Victoria University of Wellington
Mevagh Sanson, Victoria University of Wellington
Maryanne Garry, Victoria University of Wellington
OSF: https://osf.io/bnzrj/
Participants were recruited from the introductory psychology participant pool at Victoria University of Wellington and participated for course credit. For our participant pool, we used a prescreening process that allowed us to exclude, post-hoc, participants who did not meet the specified inclusion criteria. We also excluded subjects who failed to complete the experiment or experienced procedural difficulties, such as sound malfunctions on the video. The results we report are from a dataset with strict exclusion criteria, but we have additional datasets available on our OSF page with less strict exclusion criteria that may be of interest to researchers. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Matthew A. Palmer, University of Tasmania
Aaron Drummond, Flinders University
James D. Sauer, University of Portsmouth
Daniel V. Zuj, University of Tasmania
Glenys A. Holt, University of Tasmania
Miriam Rainsford, University of Tasmania
Lauren Hall, Flinders University
Liam Satchell, University of Portsmouth
OSF: https://osf.io/d97bt/
For RRR Study 1, 107 participants were recruited from three locations: the University of Tasmania (comprising 19 recruited from the introductory psychology participant pool who received course credit and 17 from the broader campus community who were compensated $10); Flinders University (19 recruited from the broader campus community who were compensated $15); and The University of Portsmouth (52 recruited from the broader campus community who volunteered their time). Due to our limited participant pool, we were unable to complete RRR Study 2 (we recruited 29 participants from the introductory psychology participant pool at the University of Tasmania). We recruited from our participant pool and community without specifying restrictions on race or age, so our total sample included an additional 16 participants in Study 1 and 23 participants in Study 2 who did not meet inclusion criteria for this RRR. Data from all participants are included on our OSF page. Due to experimenter error, 41 participants in Study 1 received a version of the response questionnaire in which subjects made their identification response and identification confidence rating on the same page, rather than different pages. This had minimal effect on identification accuracy and the results of the main analyses. Details of these extra analyses are included on our OSF page.

Peter P. J. L. Verkoeijen, Erasmus University Rotterdam
Samantha Bouwmeester, Erasmus University Rotterdam
Rolf A. Zwaan, Erasmus University Rotterdam
OSF: https://osf.io/wtbkp/
The results of RRR Study 1 were obtained by exactly executing the sampling plan and procedure described on our lab's project page at the OSF. We tested 115 Dutch-speaking Erasmus University undergraduates (most of whom were psychology undergraduates) who took part in the experiment to meet their course requirements. Ten participants did not meet the inclusion criteria: 6 of them were non-White, 1 of them heard about crucial experiment characteristics prior to participation (note that 3 other participants also indicated they heard about the experiment before, but 2 of them reported the characteristics of a different unrelated experiment and 1 of them only heard about the crossword puzzle; these 3 participants were not excluded), and 3 of them failed to adhere to the experimental instructions. After exclusion, the sample consisted of 105 participants, with 51 participants in the experimental (i.e., verbal overshadowing) condition and 54 in the control condition. Because our participants were not native English speakers, we used translated instructions. To obtain the Dutch instructions, one of the members of the research team (Verkoeijen) translated the English instructions from the approved protocol and a colleague at the Department of Psychology of the Erasmus University Rotterdam checked whether the translated versions matched the meaning of their English counterparts. The translations were adjusted based on this feedback.

Christopher A. Was, Kent State University
Dale Hirsch, Kent State University
Rachael Todaro, Kent State University
Connie Romig, Kent State University
OSF: https://osf.io/fub7j/
For RRR Study 1, 145 participants were recruited from the educational psychology participant pool at Kent State University, and they participated as part of one option for course credit. For our participant pool, we used a prescreening process that allowed us to recruit only participants who met the specified inclusion criteria. Ten participants who failed to complete the study were excluded from analyses. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Acknowledgments

Geoff Cumming, Daniël Lakens, Joanne Yaffe, and John Protzko all provided helpful guidance on the choice of a meta-analytic approach. Chris Meissner, Maryanne Garry, Robert Michael, and Kim Wade spotted the erroneous task order in the protocol for RRR1 and prompted a second replication study for this manuscript. Maryanne Garry and Chris Meissner both reviewed the protocol for RRR2 to ensure that it matched the parameters of Study 1 from S&E-S. Chris Meissner, Kim Wade, and Robert Michael provided feedback on a preliminary draft of the manuscript, and Meissner also provided the data he had compiled for the Meissner and Brigham (2001) meta-analysis for reanalysis.

Thanks to Brian Nosek and Jeffrey Spies for their assistance with the registration process at Open Science Framework and for making it possible for us to use OSF as the home for all materials for Registered Replication Reports at *Perspectives*. Finally, and most importantly, thanks to Jonathan Schooler for his cooperation in developing the protocol and for his input and assistance throughout the process.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Funding for participant payments was provided to individual labs by the Association for Psychological Science via a grant from the Center for Open Science.

Notes

1. This effect size estimate was based on a reanalysis of the data from the Meissner and Brigham (2001) meta-analysis using the same effect size measure used in this RRR. The data, a forest plot, and the R code used to conduct this analysis are available at https://osf.io/ybeur/.
2. Traditional measures of the "file drawer" problem did not reveal substantial publication bias in the verbal overshadowing literature (Meissner & Brigham, 2001), but the power-based analysis likely is more sensitive in measuring the existence of publication bias in the face of studies with small samples.
3. τ is essentially the standard deviation of the total heterogeneity. In this case, τ is 1.07%. It is a measure of the distribution of the true effects. I^2 is an estimate of the proportion of the heterogeneity that goes beyond what would be expected by chance. It is the total heterogeneity divided by the total variability. H^2 is the total variability divided by the sampling variability. The closer it is to 1, the more that the variability across effect size estimates is consistent with sampling variability rather than meaningful heterogeneity. Q is a null-hypothesis test of whether there is meaningful heterogeneity.
4. S&E-S reported that errors consisted of 59% false alarms in the verbal description condition and 60% false alarms in the control condition. Based on the raw numbers provided in Jonathan Schooler's dissertation data, the actual percentages were 59.3% and 62.5%. In Figure 4, we used the raw numbers rather than the percentages reported in S&E-S.

References

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology*, *20*, 396–413.

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 739–755.

Fallshore, M., & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1608–1623.

Finger, K., & Pezdek, K. (1999). The effect of the cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology*, *84*, 340–348.

Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, *19*, 151–156.

Lehrer, J. (2010, December 13). The truth wears off: Is there something wrong with the scientific method? *The New Yorker*, 52–57.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, *15*, 603–616.

Schooler, J. W. (2011). Unpublished results hide the decline effect. *Nature*, 470, 437.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*, 36–71.