Nonlinear effect amplification: Differential susceptibility of Verbal Overshadowing as a function of time to interference

John Protzko[1, 2]*

Sebastian Lundmark[3]

Jan Walleczek[4, 5]

Jonathan W. Schooler[2]

1 = Central Connecticut State University, USA

2 = University of California, Santa Barbara, USA

3 = University of Gothenburg, Sweden

4 = Phenoscience Laboratories, Germany

5 = Fetzer Franklin Fund, USA

* = corresponding author; protzko@gmail.com

Osf links: https://osf.io/892st/; https://osf.io/4v9jq/; https://osf.io/v3ugq/; https://osf.io/7tvbp

Acknowledgments: We would like to thank Drs. Chuck Brainerd, Valerie Reyna, John Wixted, and Christopher Wolfe, for their valuable insights and thoughts. John Protzko was supported by the Fetzer Franklin Fund during this work. This work was funded in part by the Fetzer Franklin Fund of the John E. Fetzer Memorial Trust.

The authors Declare no Conflicts of Interest.

Author Note: Data and materials, including analysis scripts, are available at https://osf.io/4v9jq/ for study 1, at https://osf.io/v3ugq/ for study 2. This study was preregistered prior to data collection at https://osf.io/7tvbp.

Abstract

Verbalizing visual memories can interfere with later accurate recall. Whereas changes in the magnitude of this Verbal Overshadowing effect as a function of delay have been reported, no study has systematically investigated multiple shorter non-immediate delays. Does VOE happen when verbalization occurs 5-minutes post-encoding? 10-minutes? 15-minutes? We show in a preregistered study involving 4,501 American adults randomly assigned to different timing paradigms, that the effect size of VOE at 5- or 10-minutes is nearly zero, with a stable and significant inhibitory effect from 15- to 20-minutes. We further investigate this nonlinearity in a second study of 3,174 individuals showing a distinct nonlinear 'effect amplification' sometime between 12- to 14-minutes. This apparent critical period after stimulus onset where susceptibility to verbal interference dramatically increases may help explain potential difficulties replicating VOE. More importantly, it suggests the possibility that the 12- to 14-minute period may represent a critical window for other interference paradigms as well.

Memory is fickle. We frequently forget things we did, we misremember what others said, we can walk into a room or open a refrigerator and have no idea what we were supposed to be doing. One of the reasons we fail to remember is, after we experience the to-be-remembered event, something interferes with the memory. This interference of new information on the encoding or retrieval of memories has a long history differentiating memory systems through finding sensitive periods where a memory is particularly susceptible to be interfered with (e.g. Atkinson & Shiffrin, 1968; Averbach & Coriell, 1961; Lechner et al., 1999; Muller & Pilzecker, 1900; Thorndike, 1913; Waugh & Norman, 1965).

Verbal Overshadowing is the reduction of memory performance and subsequent recognition that comes about by verbalizing the memory of a visual stimulus (Schooler & Engstler-Schooler, 1990). In the original Verbal Overshadowing paradigm, participants watch a 40-sec video of an unmasked man robbing a bank; then engage in a filler task for 20-minutes (taking Sudoku puzzles), then are asked to write down as much information as they can about the face (versus a control writing task). Writing a description of the face has been shown, both in online studies and in-laboratory studies, to reduce accuracy when picking out the bank robber from a lineup (e.g. Schooler & Engstler-Schooler, 1990). This verbal overshadowing of the visual memory has been extended beyond faces (e.g. Brandimonte et al., 1992; Chauvel et al., 2013; Fiore & Schooler, 2002; Melcher & Schooler, 1996, 2004; Perfect et al., 2002; Timperman & Miksza, 2017; Vanags et al., 2005) and has been the subject of successful large-scale preregistered replications (Alogna et al., 2014).

From a signal detection perspective, verbalization might impact recognition performance in two general ways: First, people may be more likely to falsely indicate that the perpetrator is not in the lineup. This is the criterion shift account (Clare & Lewandowsky, 2004), where

verbalizing causes people to become more conservative in the matching of their memory to the face. The second way people may be incorrect is by selecting the wrong stimulus (e.g. the wrong face; the discrimination account). To test whether the results are driven by a criterion shift or selecting the wrong face, researchers can exclude participants who say that the target stimulus they saw is not present. If a criterion shift is solely occurring, there would be no verbal overshadowing effect on choosing the wrong face after excluding those who said the target was not present (as in Wilson, Seale-Carlisle, & Mickes, 2018).

Previous literature has indicated the magnitude of the Verbal Overshadowing effect (VOE) depends considerably on its timing parameters: when there is a 20-minute delay between viewing a face and the subsequent verbalization, studies consistently find a significant VOE however, when verbalization occurs immediately after viewing the face VOE is dramatically attenuated, and in some cases eliminated entirely (e.g. Finger & Pezdek, 1999; Meissner & Bringham, 2001; Alogna et al., 2014; Wilson, Seale-Carlisle, & Mickes, 2018). Whereas the importance of timing parameters has been noted in the past, no study has systematically varied the timing parameters within the 0- to 20-minute interval. Thus, at present, we do not know whether the increase in the VOE as a function of delay between encoding and verbalization is linear or associated with a nonlinear 'effect amplification', such that the effect nonlinearly increases in magnitude from small and nonsignificant to the full effect within a narrowly delimited time window. Randomly assigning participants to different timing parameters may shed light on the possible functional form of the VOE. Furthermore, such an investigation may also provide insights into whether changes in VOE over time reflect a criterion shift (i.e., changes not present responses) and/or shifts in discrimination (i.e., changes the frequency of incorrectly selecting a distractor).

Overview of Current Studies

Confident that we could replicate the VOE using materials from the original VOE study and the Registered Replication Report (Alogna et al., 2014), we engaged in two preregistered studies where participants were randomly assigned to different delays before verbalizing their visual memories. Study 1 randomly assigned participants to 5- 10- 15- and 20-minute delays, and observed an apparent 'effect amplification' between the 10- and 15-minute delay. Study 2 narrowed in on this critical interval to determine precisely when the VOE amplification appeared, investigating the linearity of its emergence. To test whether the results were driven by a criterion shift, we analyzed the data both with all participants included and when restricting the data to whether participants said the perpetrator was present (e.g. testing the criterion shift by removing those who falsely said he was 'not present').

Study 1: Verbalizing at 5-, 10-, 15-, and 20-Minutes

The studies were administered online during 2020, using the materials from the online version of the Verbal Overshadowing Registered Replication Report (Alogna et al., 2014). We first engaged in extensive pilot testing at 20-minutes to make sure the procedures were sound and to test the possibility of using different filler tasks rather than having participants fill out Sudoku puzzles. We found that the effect at 20-minutes was weakened by replacing the Sudoku filler task with individual difference measures. As none of the individual difference measures moderated the strength of Verbal Overshadowing (see pilot results at https://osf.io/892st/, Protzko et al., 2022a), we conducted our studies with the original materials and Sudoku as the filler task.

During the pilot testing, we discovered the type of device participants used moderated the strength of the VOE (see again https://osf.io/892st/). Verbal Overshadowing at 20-minutes was

fully present for participants using desktops but was absent when taking the study on the smaller tablets or smartphones. Therefore, we restricted all participants to fill out Study 1 and 2 on a desktop computer (imposed by both the participant panel as well as a filter programmed into our studies).

Methods

Procedure and Materials

Participants were given a warning about the content of the study, agreed to the information sheet, and performed some checks to their system (see SOM for complete study flow and logic). Then participants read: "*This experiment consists of several tasks. Your first task is to pay attention to a short video. The video will appear on the next page of the survey. Please watch the video. The video has sound, so please ensure that your audio device is now on and at an audible volume. When the video ends, you will be able to continue the survey. When you are ready to view the video, please continue.*" On the next screen, all participants were shown a 43-second video of a bank robbery (used by Schooler & Schooler-Engstler, 1990). The screen with the video was set to automatically advance 45-seconds after the video screen had finished loading.

Next, each participant was randomly assigned to one of four timing delays, 5-, 10-, 15-, and 20-minutes. Participants in the 5-minute delay group, on the next screen, received instructions on how to solve Sudoku puzzles and were told to spend five minutes solving the Sudoku puzzles appearing on their screen. Participants in the 10-, 15-, and 20-minutes delays were given the same information and asked to solve Sudoku puzzles for 10-, 15-, or 20-minutes respectively. After solving the Sudoku puzzles, all participants were told to "*Please read and follow the instructions on the next page. You will have five minutes to complete this task. You*

need to work on this task for the full five minutes. Once five minutes has elapsed, you will be able to move on to the next task." And were then randomly assigned to either the control or the verbalizing task. Participants assigned to the control task were, on the next screen, asked to *"Please name 50 countries and their capital cities. For example: France, Paris. Please do not look this information up online – we are interested only in your own answers. Thank you."* And could then type as many countries and their respective capitals in open-text boxes numbered from 1 to 50 and appearing vertically on the same screen. Participants assigned to the verbalizing task were asked to "*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."* And where the participant could type their description in an open-text box appearing on their screen (Alogna et al., 2014). The screen with the control and verbalizing task were set to automatically advance five minutes after the screen had finished loading.

Next, to ensure the total time spent on the study was constant across delay times, participants completed another Sudoku puzzle filler task for 15, 10, or 5 minutes so that all participants, regardless of condition, would spend 25-minutes between seeing the video and seeing the lineup. The instructions on the screen with the second Sudoku filler task was identical to the first and the screens were again set to automatically advance after 15, 10, or 5 minutes depending on which time-delay the participant was assigned.

On the next screen, all participants were told "*That's the end of that task. Next you will see a lineup with 8 faces. Please identify the individual in the line up 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'." After clicking next, eight pictures appeared at the top of the participants' screens, numbered from 1 to 8 (see SOM). Below the eight pictures, participants were instructed to "*Please tick one of the following boxes to indicate your selection:*" and had nine radio buttons presented vertically and labeled 1 to 8 from top to bottom and with the ninth radio button labeled "Not present".

On the next screen, for exploratory purposes, participants rated their confidence in their choice and were on the last three screens asked to complete a Captcha to determine that the answers were not generated by automation software, to report whether they had participated seriously, and to report whether they completed the questionnaire on a computer, smartphone, tablet, or something else.

Throughout the procedure, participants who did not answer a question were told "Please answer the question" and could not continue unless they gave an answer. For the full questionnaire logic, see section S1 in the Supplementary Online Materials (SOM). Section 4 in the SOM detail how and when participants dropped out during the procedure.

In addition to the four delay groups, two groups of participants were randomly assigned to two pilot conditions for future studies. The first pilot condition assigned participants (N = 1,075) to a 20-minute verbalization delay but where participants were asked to verbalize or to do the control task using identical instructions as above but instead asked to do them for just a single minute rather than the standard five. This was done to see if the overall administration time could be reduced without the loss of a VOE. The second pilot condition assigned participants (N = 519) to solve the Sudoku task for 25 minutes and then answer the line-up question. This pilot was used to assess whether the control task (naming countries and capital cities) had an interference effect on memory.

Data Analysis

Previous investigations have analyzed correctness in many different ways, including Signal Detection Theory (e.g. Clare & Lewandowsky, 2004); ROC analysis (e.g. Wilson et al., 2018), whether the witness is correct (e.g. Schooler & Engstler-Schooler, 1990). As VO was the target of a multi-lab Registered Replication Report (e.g. Alogna et al., 2014) that successfully replicated VO at a 20-minute delay using the simpler whether the witness correctly identified the perpetrator, we chose that as the target of our investigation. This analysis involves only providing participants with a lineup where the subject is present, precluding some analyses (e.g. ROC, Wilson et al., 2018) but allowing participants to choose that the target is 'not present' which allows for others (e.g. Criterion Shifts). For all analyses we perform OLS regressions on accuracy with robust standard errors (e.g. Gomila, 2021). Using probit regression with robust standard errors does not change the results (see SOM).

Accuracy. Participants who identified the picture with the bank robber (i.e., picture six, the second picture on the bottom row in the line-up when counting from the left-most picture to the right-most picture) were coded as 1 in a 'correct' variable, and 0 otherwise.

Participants

Participants were drawn with the intention to have 4,400 completed interviews. The sample was a non-probability sample of American adults aged 18 or older drawn from the sample provider Dynata (see SOM for details). To conduct our power analysis, we used the results from the online study using the 20-minute delay from the Registered Replication Report (Alogna et al., 2014). To achieve 95% power, one needs 550 participants per control/verbalizing group. Therefore, we aimed to have 1,100 participants for each delay time, to ensure we could accurately power a 20-minute replication. This led us to aim for a sample size of 4,400 across

our four delay times. The final sample was 4,501 participants, and included participants who completed the study (see SOM S3 for the pattern of dropout). Data and materials, including analysis scripts, are available at https://osf.io/4v9jq/ (Protzko et al., 2022b).

Results

Verbal Overshadowing at 5-, 10-, 15-, and 20-minutes

Verbalizing had a nonsignificant effect on the accuracy of memory when it occurred only 5-minutes ($b_{\text{5-minute delay}}$ (1207) = 0.00, SE = 0.03, p = 0.89) or 10-minutes after the stimulus was presented ($b_{\text{10-minute delay}}$ (1117) = -0.01, SE = 0.03, p = 0.78; see Table S1). However, in stark contrast, at both 15- and 20-minutes after encoding the memory, verbalizing the person's face caused a significant decrease in accuracy for the face ($b_{\text{15-minute delay}}$ (1114) = -0.07, SE = 0.03, p = 0.006; $b_{\text{20-minute delay}}$ (1055) = -0.06, SE = 0.03, p = 0.031; see Table S1). Notably, there was no difference in magnitude of the VOE at these latter two timing delays (χ^2 (1, N = 2,173) = 0.13, p = 0.71).

Table 1 details the number and percent of the correct, inaccurate, and not present identifications within each condition and across time-delays.

Table 1. Descriptive statistics of Study 1.

	Time-delay			
	5-minutes	10-minutes	15-minutes	20-minutes
Control condition				
Accurate identification	172 (29%)	149 (28%)	166 (30%)	150 (29.8%)
Inaccurate identification	291 (49.1%)	276 (51.9%)	262 (47.4%)	237 (47%)
Inaccurate not present	130 (21.9%)	107 (20.1%)	125 (22.6%)	117 (23.2%)
Total	593	532	553	504
Verbalize condition				
Accurate identification	181 (29.4%)	160 (27.3%)	128 (22.7%)	132 (23.9%)
Inaccurate identification	281 (45.6%)	275 (46.9%)	269 (47.8%)	207 (37.4%)

Inaccurate not present	154 (25%)	152 (25.9%)	166 (29.5%)	214 (38.7%)
Total	616	587	563	553

Notes. Percent participants across identifications within time-delay in parentheses.

Illustrating the effect of the delay, the pattern appearing in Figure 1 suggests a nonlinear 'effect amplification' of Verbal Overshadowing paradigm, such that the effect is not statistically significant when verbalization occurs 5- and 10-minutes post encoding, but is comparably observed when verbalization takes place 15- and 20-minutes post-encoding.

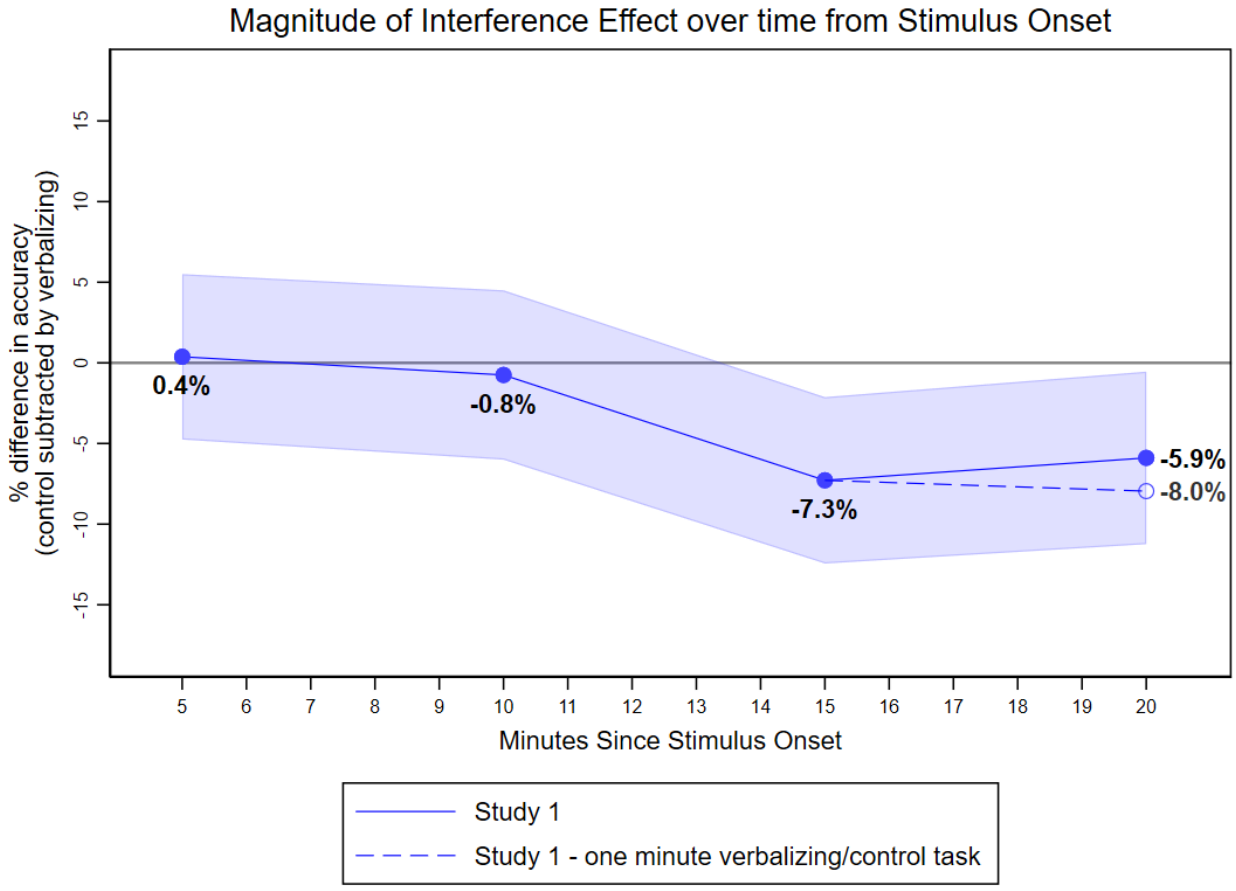

Figure 1. Effect of verbalizing a visual stimulus on subsequent recall, by time in between seeing the stimulus and engaging in the verbalization task. Shaded area represents the 95%CI.

Pilot conditions for future studies

The group that solved Sudoku for the full 25-minutes before answering the line-up question were equally accurate compared to those who completed the standard 20-minute Verbal

Overshadowing control task ($b_{control}$(1021) = -0.00, SE = 0.03, p = 0.97) and were also more accurate in identifying the face than those who verbalized ($b_{verbalize}$(1070) = -0.06, SE = 0.03, p = 0.027; see SOM, Table S3, column 1).. Thus, we are not concerned about naming capital cities having an interference on memory at 20-minutes, but future studies could test the control task at different timings and with greater power.

In addition, the size of the VOE was equally strong for those randomly assigned to fill out the description or naming capital cities for 5-minutes compared to only one minute ($b_{\text{verbalized*1-minute task}}$ (2128) = -0.02, SE = 0.04, p = 0.59; see SOM, Table S3, column 2). The same effect of verbalizing over delay was the case when removing people who said the perpetrator was 'not present. That is, at shorter delays, 5-minutes ($b_{\text{5-minute delay}}$ (923) = 0.02, SE = 0.03, p = 0.53) and 10-minutes ($b_{\text{10-minute delay}}$ (858) = 0.02, SE = 0.03, p = 0.60) there were no significant decrease in the accuracy of memory, whereas at the 15-minute delay the overshadowing effect had been amplified to a significant effect ($b_{\text{15-minute delay}}$ (823) = -0.07, SE = 0.03, p = 0.05). At the 20-minute delay, we observed a fascinating pattern based on whether the participants had been randomly assigned to fill in the Verbalizing/Control task for 5-minutes or 1-minute. Engaging in the Verbalizing/Control task for 5-minutes, there was no effect of verbalizing on accuracy when removing people who said the target was 'not present' ($b_{\text{20-minute delay-5-min}}$ (724) = 0.00, SE = 0.04, p = .96)—with a marginal effect when engaging in the Verbalizing/Control task for 1-minute ($b_{\text{20-minute delay-1-min}}$ (773) = -0.06, SE = 0.03, p = 0.072; see Table S2). However, these two VOE estimates when removing 'not present' responders did not statistically significantly differ (χ^2(1, N = 1,501) = 1.65, p = .20).

Exploratory Analyses

The main result of our investigation was to test the magnitude of the VOE when participants are randomly assigned to different delays after seeing the to-be-encoded information. Only target-present lineups were used to replicate the externally validated procedures from a multilab RRR (Alogna et al., 2014). Using only target-present lineups, however, precludes a formal signal detection analysis which would involve also including lineups where the target was not present (thus, all choices of perpetrators would be noise). One way to begin to explore such criterion shifts in our study is to analyze the results of verbalizing on whether they chose some other person in the lineup (instead of selecting the target was 'not present'. Thus, removing people who responded 'not present' can help investigate the criterion shift, although outside of a formal signal detection analysis.

The same effect of verbalizing over delay was the case when removing people who said the perpetrator was 'not present'. That is, at shorter delays, 5-minutes ($b_{\text{5-minute delay}}$ (923) = 0.02, SE = 0.03, p = 0.53) and 10-minutes ($b_{\text{10-minute delay}}$ (858) = 0.02, SE = 0.03, p = 0.60) there were no significant decrease in the accuracy of memory, whereas at the 15-minute delay the overshadowing effect had been amplified to a significant effect ($b_{\text{15-minute delay}}$ (823) = -0.07, SE = 0.03, p = 0.05). At the 20-minute delay we observed a fascinating pattern based on whether the participants had been randomly assigned to fill in the Verbalizing/Control task for 5-minutes or 1-minute. Engaging in the Verbalizing/Control task for 5-minutes, there was no effect of verbalizing on accuracy when removing people who said the target was 'not present' ($b_{\text{20-minute delay-5-min}}$ (724) = 0.00, SE = 0.04, p = .96)—with a marginal effect when engaging in the Verbalizing/Control task for 1-minute ($b_{\text{20-minute delay-1-min}}$ (773) = -0.06, SE = 0.03, p = 0.072; see Table S2). There was no difference between the VOE between 15- to 20-minutes when the Verbalizing/Control task was done for 1-minute ($\chi^2(1, N = 1{,}600) = 0.00$, p = 0.96), and neither

between 15- to 20-minutes when the Verbalizing/Control task was done for 5-minutes ($\chi^2(1, N = 1{,}551) = 1.87, p = .17$)

Discussion

We found the VOE at 5- and 10-minutes after viewing the face was nonsignificant, with a strong, significant, and constant effect at 15- and 20-minutes after seeing the face. This suggests that the presence and strength of a VOE is sensitive to the unique timings between seeing the stimulus, encoding it, and the administration of the interfering verbalizing task. This apparent nonlinear 'effect amplification' of Verbal Overshadowing between 10 and 15 minutes after viewing the face has not been shown systematically before.

The nonlinear amplification function is reminiscent of a cubic relationship, with a stable null effect (5- to 10-minutes), a nonlinear amplification (12- to 14-minutes), and then a stable significant effect (15- to 20-minutes). The data from Study 1, however, are not complete enough to test a cubic relationship, as there are not enough degrees of freedom with just four datapoints. To explore what this strong 'effect amplification ' may look like, we ran a second study where participants were randomly assigned to delay times between the time where visual memory appears to become sensitive to interference effects, that is, between 10- and 15-minutes. This allowed us to pinpoint the timing at which Verbal Overshadowing is amplified and assess whether the amplification is nonlinear as assessed by a cubic analysis.

As we did not find a negative effect of naming capital cities as a control group at 20-minutes (compared to just doing Sudoku puzzles for another few minutes), we continued to use it as a control condition in Study 2. Furthermore, as the magnitude of the Verbal Overshadowing was not different with 1- instead of 5-minutes of verbalizing, we used the 1-minute verbalizing task in Study 2.

Furthermore, the effect of verbalizing on accuracy seems to be entirely driven by a criterion shift at delays of 5 and 10 minutes, and, oddly, at 20-minutes but only when the Verbalizing/Control task was performed for 5-minutes. When the Verbalizing/Control task was performed for 1-minute, there was still a residual discrimination effect from Verbalizing on accuracy (see Table 2 for descriptives). This could be chalked up to a possible sampling error, if it wasn't for the fact that at 15-minutes delay there was evidence for a discrimination effect. This further highlights both the nonlinearity of the Verbal Overshadowing effect in its relationship to the amount of time that passes between encoding and verbalization. For all of these reasons, we attempt to investigate what is happening between 10- and 15-minutes in a second study.

Table 2. Descriptive statistics of the 20-minutes delay with 1-minute task and 5-minute task.

	Time-delay	
	20-minutes + 1-minutes task	20-minutes + 5-minute task
Control condition		
Accurate identification	174 (30.5%)	150 (29.8%)
Inaccurate identification	262 (46.0%)	237 (47.0%)
Inaccurate not present	134 (23.5%)	117 (23.2%)
Total	570	504
Verbalize condition		
Accurate identification	114 (22.6%)	132 (23.9%)
Inaccurate identification	225 (44.6%)	207 (37.4%)
Inaccurate not present	166 (32.9%)	214 (38.7%)
Total	505	553

Notes. Percent participants across identifications within time-delay in parentheses.

Study 2: Verbal Overshadowing at 11-, 12.5-, and 14-minutes

Methods

Design, Procedure, and Materials

Study 2 used identical materials as in Study 1, except a changed timing between seeing the video and engaging in either the verbalizing or control task (for the full questionnaire logic, see section S2 in SOM). As we observed no verbal overshadowing effect at 5- or 10-minutes in Study 1 and a stable VOE between 15- to 20-minutes, we chose three times between 10- and 15-minutes. We chose 12.5 minutes to be exactly between the two, and 11 and 14 to be 1.5 minutes both before and afterward.

Participants

Participants were drawn with the intention to have 3,000 completed interviews. The sample was a non-probability sample of American adults aged 18 or older, again drawn from the sample provider Dynata. We did not conduct a formal power analysis, however, because one cannot power out for a null effect (as in Verbal Overshadowing at 10-minutes). We simply kept the sample size the same per condition as we had in Study 1 and aimed to pinpoint the effect size magnitudes along the predicted path shown in Figure 1. The analyses only include participants who completed the study (see SOM S3 for the pattern of dropout). Data and materials, including analysis scripts, are available at https://osf.io/v3ugq/ (Protzko et al., 2022c). This study was preregistered prior to data collection at https://osf.io/7tvbp (Protzko et al., 2022d).

Results

Wholly in-line with the sensitive period between 10- and 15-minutes discovered in Study 1, the results of Study 2 showed an increased deleterious effect on memory at the 11-minute delay, albeit not yet statistically significantly so ($b(1242) = -0.03$, SE = 0.03, $p = 0.23$). At 12.5-minutes, a marginally significant deleterious effect was discovered ($b(866) = -0.05$, SE = 0.03, p

= 0.091), and at a 14-minutes delay the deleterious effect on memory was statistically significant (b(1060) = -0.08, SE = 0.03, p = 0.005).

Table 3 details the number and percent of the correct, inaccurate, and not present identifications within each condition and across time-delays.

Table 3. Descriptive statistics of Study 2.

	Time-delay		
	11-minutes	12.5-minutes	14-minutes
Control condition			
Accurate identification	202 (32.3%)	147 (35.4%)	183 (34.8%)
Inaccurate identification	289 (46.2%)	176 (42.4%)	246 (46.8%)
Inaccurate not present	135 (21.6%)	92 (22.2%)	97 (18.4%)
Total	626	415	526
Verbalize condition			
Accurate identification	180 (29.1%)	136 (30%)	144 (26.9%)
Inaccurate identification	264 (42.7%)	176 (38.9%)	235 (43.8%)
Inaccurate not present	174 (28.2%)	141 (31.1%)	157 (29.3%)
Total	618	453	536

Notes. Percent participants across identifications within time-delay in parentheses.

Illustrating this sudden onset of the interference effect that verbalization had on memory, the solid line between the 11-, 12.5-, and 14-minute delays in Figure 2 shows a close replication of the predicted line between 10- and 15-minutes delays in Figure 1. At 11-minutes, the Verbal Overshadowing effect was increasing, from -0.8% at 10-minutes (Study 1) to -3.1% at 11-minutes ($b_{\text{11-minute delay}}$ (1242) = -3.1, SE = 0.03, p = .23). The verbal overshadowing effect continued to grow at 12.5-minutes to a 5.4% reduction in accuracy ($b_{\text{12.5-minute delay}}$ (866) = -5.4, SE = 0.03, p = .091). At 14-minutes, the size of the verbal overshadowing effect on accuracy had

grown to to an 8% reduction in accuracy ($b_{14\text{-minute delay}}$ (1060) = -8.0, SE = 0.03, p = .005), which is the same magnitude as seen at 15-minutes (see Study 1 and Figure 2).

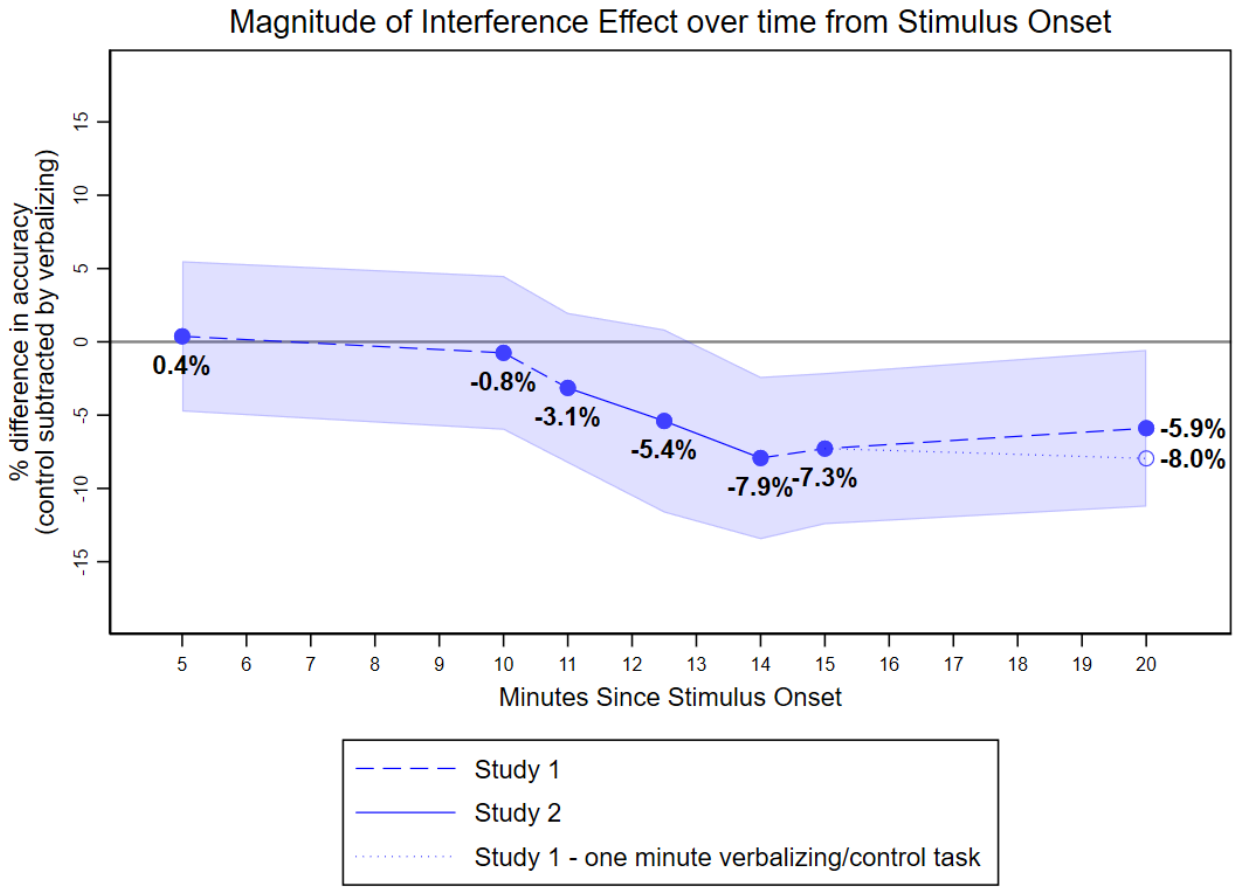

Figure 2. Interference effect of verbal descriptor task compared to control task (Study 1 and Study 2).

Furthermore, in line with a sudden onset of a period where visual memory is susceptible to interference by verbalizing, an exploratory test showed that the relationship between the seven delays was statistically significantly better modeled following a cubic function than a linear function where the deleterious interference effect grew at a steady pace (Likelihood-ratio test: χ^2 (2, N = 7) = 16.69, p = .0002).

Criterion Shifts

In Study 1, there was evidence for both a criterion shift and a discrimination function at 15-minutes delay. At 11-minutes the effect of verbalizing was due to a criterion shift, represented by the change from -3.1% interference when including all responses, to only a negative -0.6% difference when excluding participants who said the perpetrator was not present ($b(933) = -0.006$, SE = 0.03, $p = 0.85$; see Table S4). Similarly, at 12.5-minutes, the criterion shift still explained the slight difference between those who verbalized and those who named capitals ($b(633) = -0.02$, SE = 0.04, $p = 0.63$). At 14-minutes, the criterion shift appeared to, at least directionally, not be the only cause of the increased deleterious effect ($b(806) = -0.05$, SE = 0.03, $p = 0.18$), and at 15-minutes, a criterion shift was no longer the sole explanation for the deleterious effect of verbalizing on memory ($b(823) = -0.07$, SE = 0.03, $p = 0.05$) (see Study 1).

To explore this further, we plot the ratio of the Verbal Overshadowing effect attributable to a criterion shift. We do so by estimating the percent of the overall VOE that is attributable to a discrimination function or a criterion shift. Plotting this as a function of the delay again show the strong nonlinear amplification and that the amplification seemed due to a growing discrimination effect between 10-15 minutes (see Figure 3). At 11-minutes, the entire effect was due to criterion shift, whereas criterion shift explained only 64.4% and 41.2% at 12.5- and 14-minutes respectively. At 15-minutes, even less (10.2%) of the effect was explained by a criterion shift, leaving 89.8% due to discrimination. At 20-minutes (Study 1), the criterion shift again explained 100% of the VO effect. However, for the 20-minutes pilot with 1-minute verbalizing/control task, only 21% was explained by criterion shift, suggesting that between 15- and 20-minutes, the sudden onset of the discrimination effect seem to either dissipate or maybe remain constant.

Whether the non-linear amplification of the Verbal Overshadowing effect between 10-15 minutes is due to a sudden onset of a discrimination effect or due to some other mechanisms such as criterion shift remain to be properly confirmed and studied in future data collections.

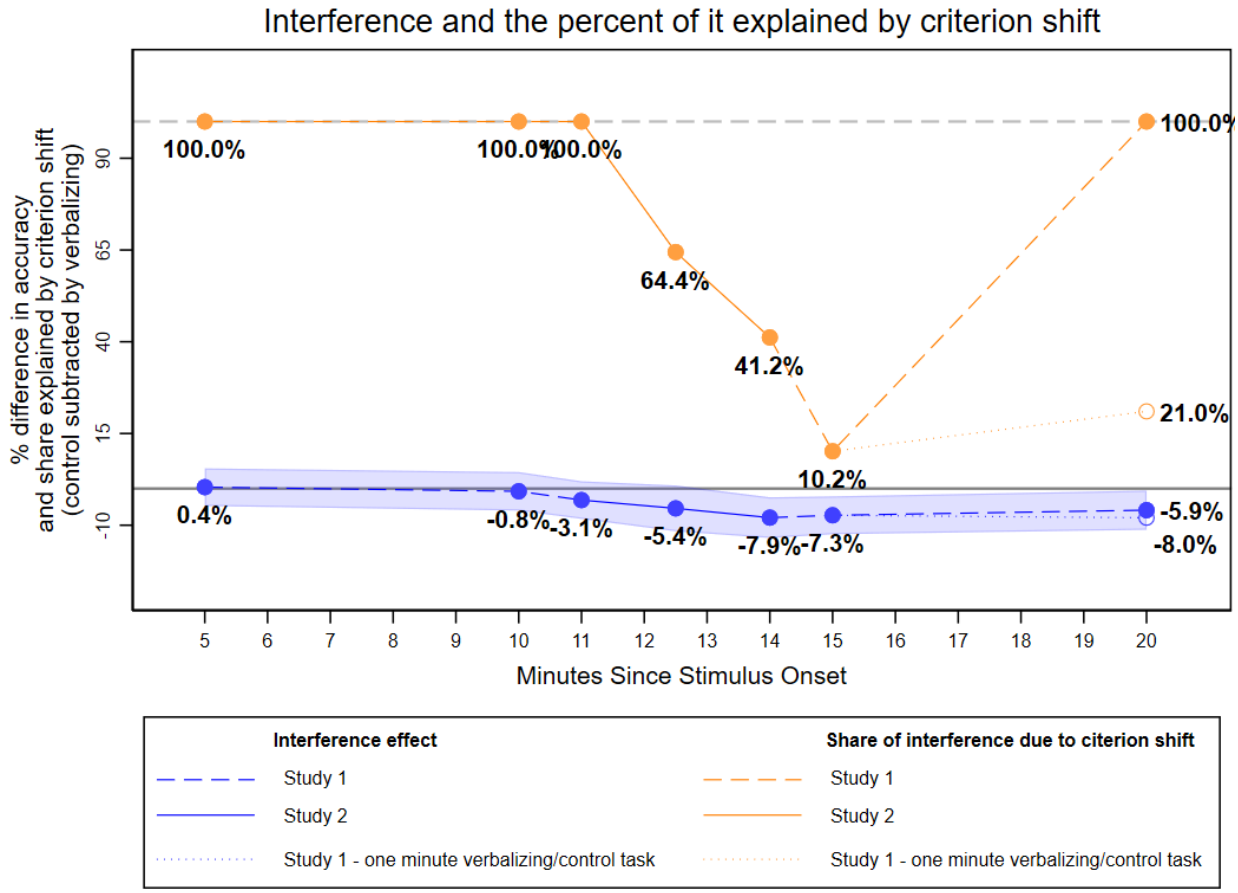

Figure 3. Share of interference effect of verbal descriptor task compared to control task due to criterion shift (Study 1 and Study 2).

Discussion

Study 2 showed that the VOE is not only sensitive to the timing of when verbalization occurs after seeing the stimulus, but the effect follows a nonlinear 'effect amplification' somewhere between 12- to 14-minutes. The exact timing of this amplification is unclear as assumptions of linearity would necessarily not hold for a nonlinear 'effect amplification' process. However, that the effect size of Verbal Overshadowing between 5- to 10-minutes after seeing the stimulus is minimal and nonsignificant yet nonlinearly amplifies within 12-14

minutes, represents a novel discovery in the dynamics of interference and memory. All timings investigated during both studies one and 2 can be seen in Figure 4.

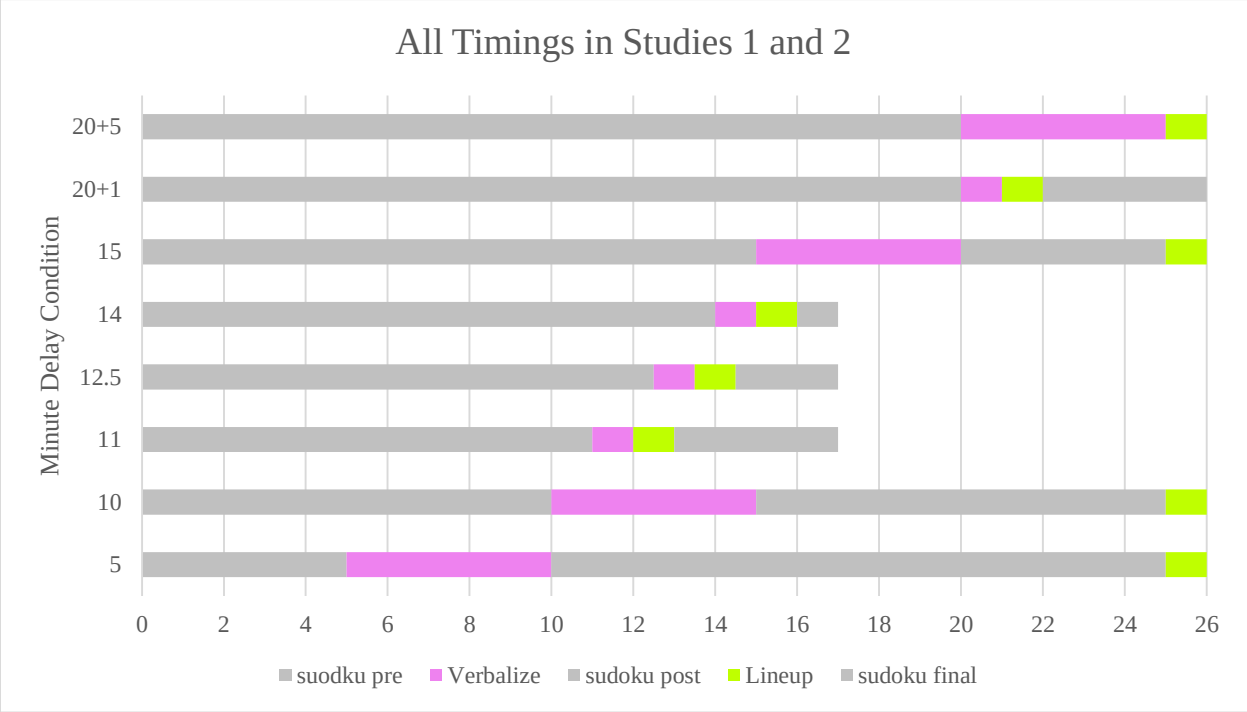

Figure 4: Timing of the filler task (grey bars), Verbalizing or engaging in the Control task (Violent bars), and performing the lineup task (lime bars) across studies one and two. Study one used a 5-, 10-, 15-, and 20-minute delay (with the additional 5 vs. 1 min Verbalizing/Control test at 20-minutes). Study two tested an 11-, 12.5-, and 14-minute delay with the 1-minute Verbalizing/Control task form Study one.

In addition, the magnitude of the Verbal Overshadowing effect was nearly identical at 14-minutes in Study 2 and 15-minutes in Study 1. This is important as the 14-minute delay had the lineup directly afterwards, while the 15-minute delay had a 5-minute gap between Verbalizing and lineup (both to be consistent with overall data collection time with the other conditions ran in their respective studies. This suggests the effect of delay time on Verbal Overshadowing is not altered by the gap between verbalizing and the lineup.

General Discussion

We set out to identify when within the 20-minute window after seeing a face the VOE is observed. Study 1 found a stable, near-0, null effect from 5- to 10-minutes after seeing the stimulus, then a stable VOE from 15- to 20-minutes. This prompted us to investigate between 10- and 15-minutes, where our results mapped on to the predicted effect sizes (Study 2; Figure 2). Between 14- and 20-minutes, the VOE was entirely stable, showing no increase in magnitude.

This is the first systematic investigation randomly assigning people to timing delays in between 0- to 20-minutes. The nonlinear VOE amplifying somewhere around 12-14 minutes has not been observed before. This nonlinear 'effect amplification' has important implications for metascience, memory, and legal proceedings. That the VOE is stronger at longer delays suggests that studies replicating the VOE at short delays in the 5- to 10-minute timing delay may fail to produce significant effects, whereas studies with post-encoding verbalization durations longer than 15 minutes may fail to find any further impact of greater post-encoding delays. For memory research more generally, this brings up interesting questions about what is special and generalizable about the 12-14 minute window. Such rapid onset of an effect amplification could be the result of multiple separate processes that have their own timing parameters. For legal proceedings, the work here suggests there is an exceptionally limited band of delay, somewhere 5- to 10-minutes after witnessing an event, where a witness can give a verbal description of a suspect and *not* have it affect their later memory. Verbalizations outside of this 5- to 10-minute window may be more likely to corrupt memory and lineup accuracy.

Limitations

This work only used one stimulus, one interference task, and one control task. While this was done to test the dynamic effects of the Verbal Overshadowing paradigm over time (using the stimuli from the original Verbal Overshadowing study; Schooler & Engstler-Schooler, 1990) and

its worldwide replication (Alogna et al., 2014), it is possible a different time for the nonlinear 'effect amplification ' could occur with different materials (target and/or test stimuli), a different control task (instead of naming capitals) or a different filler task (instead of Suduko puzzles). Future work can look at the robustness of not only the presence but the timing of the nonlinear turning on of Verbal Overshadowing.

Furthermore, this study focused on using an independently replicated procedure (used in Alogna et al., 2014). Because of that, we did not include conditions where the target was absent, and thus any selection of a target would be considered 'noise'. Although comparison of performance averages that do versus do not include "not present" selections enables us to address whether criterion, discrimination or both are taking it place (see Mickes & Wixted, 2015, for more discussion), the absence of a perpetrator absent line up limits our ability to use signal detection to generate formal quantitative estimates of shifts in criterion and d' .

Future Directions

We found both evidence for a criterion shift and evidence of additional interference reflecting reduced discrimination. Excluding participants who were impacted by a criterion shift (i.e.; who selected the "not present" option) weakened, but did not eliminate Verbal Overshadowing 15- to 20-minutes after the stimulus had been seen. The observation of evidence for both a shift in criterion and discrimination criterion shift occurring is in line with the suggestion of Clare & Lewandowsky (2004), who argued that while a Criterion Shift can account better for the VOE, the full effect is likely a multifaceted combination of Criterion Shift and Interference. The relative impact of these mechanisms may depend on a variety of factors, associated with differences in participants, stimuli, instructions, and/or time delay. Future research might profitably explore these factors and assess the mechanisms underpinning them.

The nonlinearity we are discussing is on the treatment effect. There is no a priori reason such a treatment effect *should* be nonlinear across delay times. This is especially the case as our base-rates for accuracy are neither near ceiling at the short delay of 5-min, nor near floor at the longest delay of 20-min. There seems to be something unique happening in the 11-15-min window to the treatment effect. One possibility for this unique timing is some yet to be identified middle-term memory turns on in the 11-15 minute window. Another possibility is this is a confluence of multiple interacting memory processes that change at different rates. For example, the effects of verbalization may have one unique time course while the effects of the control task could have another. This diverging responses may shift the timing of the nonlinear effect amplification window. Another example of overlapping memory processes is the differing time courses of a criterion shift versus a discrimination effect. Criterion shifts appear to be operating on one time course, and discrimination effects on another. The unexpected nonlinearity over time appears to be primarily driven by false identifications, but not entirely. Thus, their overlap may be giving rise to the nonlinear effect amplification seen here. Future research is needed to flesh out the source of this nonlinearity in treatment effects from the Verbal overshadowing paradigm.

There may be great value in investigating whether such a nonlinear 'effect amplification ' exists in other paradigms in this 12- to 14-minute window. While short-term memory is differentially susceptible to interference effects within one minute of seeing a stimulus, it is unclear as yet why we would see a sensitive period 12- to 14-minutes after seeing a visual stimulus to verbal interference. Robust demonstrations of this critical period using different memory paradigms, stimulus sets, and rigorous control might elucidate the nature and generality of this window.

Conclusion

Verbal Overshadowing, the memory interference that can arise from verbalizing visual memories, is a dynamic phenomenon with a unique time-course. The effect is nonlinear with respect to delay between encoding and verbalization, and, in the present paradigm, shows a sharp effect amplification between 12- to 14-minutes. Understanding the source of this newly discovered nonlinear effect amplification time-course may offer new insights into understanding verbal overshadowing, and perhaps memory more generally.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, *2*(4), 89-195.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *The Bell System Technical Journal*, *40*(1), 309-328.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition*, 20: 449–455.

Chauvel, G., Maquestiaux, F., Ruthruff, E., Didierjean, A., & Hartley, A. A. (2013). Novice motor performance: better not to verbalize. *Psychonomic bulletin & review*, 20(1), 177-183.

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(3), 396-413.

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 739.

Finger, K., & Pezdek, K. (1999). The effect of cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology*, *84*(3), 340.

Fiore, S. M. & Schooler, J. W. (2002). How did you get here from there? Verbal overshadowing of spatial mental models. *Applied Cognitive Psychology*, 16: 897–910.

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, *150*(4), 700-710.

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 years of consolidation—remembering Müller and Pilzecker. *Learning & Memory*, *6*(2), 77-87.

Meissner, C. A., & Brigham, J. C. (2001). A meta- analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 15(6), 603-616.

Melcher, J. M., & Schooler, J. W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of memory and language*, 35(2), 231-245.

Melcher, J. M., & Schooler, J. W. (2004). Perceptual and conceptual training mediate the verbal overshadowing effect in an unfamiliar domain. *Memory & cognition*, 32(4), 618-631.

Mickes, L., & Wixted, J. T. (2015). On the applied implications of the "verbal overshadowing effect". *Perspectives on Psychological Science*, 10(3), 400-403.

Muller, G. E., & Pilzecker, A. Experimentelle Beitrage zur Lehre vom Gedachtnis. *Z. Psychol*., 1900, 1, 1-300.

Perfect, T. J., Hunt, L. J., & Harris, C. M. (2002). Verbal overshadowing in voice recognition. *Applied Cognitive Psychology*, 16: 973–980.

Schooler, J. W. (2014). Turning the lens of science on itself: Verbal overshadowing, replication, and metascience. *Perspectives on Psychological Science, 9*(5), 579-584.

Thorndike, E. L. Educational psychology: H. *The psychology of learning*. New York: Teachers College, Columbia University, 1913.

Timperman, E., & Miksza, P. (2017). Verbalization and musical memory in string players. *Musicae Scientiae*, 1-19. doi: 10.1177/1029864917727332.

Vanags T., Carroll, M., & Perfect, T. J. (2005). Verbal overshadowing: A sound theory in voice recognition? *Applied Cognitive Psychology*, 19: 1127–1144.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89–104.

Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. (2018). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General*, *147*(1), 113.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of personality and social psychology*, *111*(4), 493.