

## Choose Your Words Wisely: What Verbal Hesitation Indicates About Eyewitness Accuracy

AYANNA K. THOMAS<sup>1\*</sup>, CAROLINE CHEN<sup>1</sup>, LEAMARIE T. GORDON<sup>1</sup> and THORA TENBRINK<sup>2</sup>

<sup>1</sup>Tufts University, Medford, USA

<sup>2</sup>Bangor University, Bangor, UK

*Summary:* Researchers have demonstrated qualitative differences in witness verbal reports made in the presence and absence of misinformation. The present study examined changes in linguistic markers present in verbal reports in the context of a repeated-retrieval misinformation study. After witnessing an event, an immediate retrieval group engaged in a free-recall test associated with the event. The delayed retrieval group completed a filler task. Following, all participants were presented with a post-event narrative that included neutral, consistent, and misleading details. Both groups then took two free-recall tests. We found that hesitations were more likely to accompany correctly remembered details if those details were altered in the narrative, than if there was consistency between the original event and narrative. We also found that retrieval prior to misinformation positively influenced the inclusion of hesitations in free-recall reports that immediately followed the narrative. Copyright © 2015 John Wiley & Sons, Ltd.

The common factor that unites all eyewitnesses is the obligation to report on the witnessed event. These verbal reports can take multiple forms. For example, an eyewitness may be asked an open-ended question that requires the retrieval of all relevant information. Eyewitnesses may also be asked questions about specific details, often receiving retrieval cues from investigators. Regardless of the style, all eyewitnesses tell and retell their story. In so doing, eyewitnesses not only have the explicit goal of being complete as well as accurate in their report, they also have the implicit goal of selecting the appropriate words and phrases to accurately represent the event. The choice of specific words and phrases not only communicates the details of the event but also communicates to the listener the eyewitness' confidence in his or her report. However, accompanying intentional choices, unprepared spoken language often includes false starts, repetitions, and hesitations that may not be intentional. Unintentional expressions and pauses, such as 'ums' and 'ahs', can be construed to represent uncertainty in one's recollection and may also reflect lexical inaccessibility. The present study focused on the inclusion of these unintentional items in eyewitness reports, in the context of a repeated-retrieval misinformation paradigm. We tested the hypothesis that unintentional items would be most pronounced in situations where participants had memorial or lexical uncertainty or inaccessibility (cf., Arnold, Wasow, Losongco, & Ginstrom, 2000; Hartsuiker & Notebaert, 2010).

Unintentional expressions account for approximately 6% of words uttered in unprepared spoken language, such as eyewitness reports (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 2001). Research suggests that expressions such as 'um' or 'uh' may represent doubt or uncertainty (Brennan & Williams, 1995; Smith & Clark, 1993). Research also suggests that these unintentional items are often associated with inaccurate or suggested information (Arciuli, Mallard, & Villar, 2010; Pickel, 1999;

Schooler, Gerhard, & Loftus, 1986) and are more likely to occur in conjunction with deception (Bruck, Ceci, & Hembrooke, 2002; Saykaly, Talwar, Lindsay, Bala, & Lee, 2013; Saykaly et al., 2013). For example, one study found that when participants reported details based on misleading information in a final test of memory, those details were often accompanied by verbal hesitation markers such as 'I think' or 'I believe' more often than when correct details were reported (Pickel, 1999). Similarly, Schooler et al. (1986) found that participants described misinformation-based details with fewer perceptual and/or contextual elements. These studies also demonstrated a robust misinformation effect in that participants were less able to remember correct details from the original event and more likely to produce misleading details on a final test after exposure to post-event misleading information (for a review, see Frenda, Nichols, & Loftus, 2011). This small body of research suggests that nonlinguistic verbal utterances may serve as an indicator of eyewitness report confidence, accuracy, and intention to deceive.

The research examining the quality and content of eyewitness verbal reports after exposure to post-event misleading information indicates that when participants include suggested information in their final reports, those details often are accompanied by more verbal hesitations than correctly reported details. Although an important demonstration, this previous research has focused on the evaluation of a single verbal report of a witnessed event. However, it is more likely that eyewitnesses will engage in several retrieval episodes prior to making a final memory report, or testifying in court. That is, prior to testifying, eyewitnesses will likely be questioned numerous times by investigators and may also recount events to friends and family. Research suggests that multiple reports directly influence nonlinguistic utterances (Saykaly et al., 2013). For example, when children ( $M_{\text{age}} = 7.58$ ) were interviewed over three consecutive days, nonlinguistic markers were shown to systematically decrease. The present study expands on previous research by examining the change in verbal hesitation markers over the course of multiple retrieval attempts within the context of the

\*Correspondence to: Ayanna K. Thomas, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA.  
E-mail: ayanna.thomas@tufts.edu

misinformation paradigm. In addition, the present study focuses on college-aged adults.

In the present study, we used a repeated-retrieval misinformation paradigm and free-recall tests during each retrieval phase. In the context of a repeated-retrieval design, the present study can examine the changes in linguistic markers such as hesitations, across repeated-retrieval attempts, and between retrieval attempts that precede and follow the presentation of misinformation. Our primary goal was to examine the verbal markers that accompany correctly and incorrectly reported details in a repeated-retrieval misinformation paradigm. A second goal was to investigate changes in verbal markers across retrieval attempts. Verbal reports from two groups of participants were compared. Both groups were presented with a staged crime. Following presentation, one group, the *immediate retrieval* group, was asked to provide a verbal report of the event. A second group, the *delayed retrieval* group, engaged in an unrelated task. Both groups were then presented with an audio narrative of the event that included details consistent and inconsistent with the originally witnessed event. After reading the post-event narrative, both groups were asked to provide two free-recall reports of the event. The final two recall reports were spaced by 10 minutes. The only difference between the two groups was the free-recall report that occurred prior to misinformation presented in the *immediate* group. The two groups will be referred to as *immediate* and *delayed*, from this point forward.

Prior research suggests that verbal hesitation markers may be an indicator of uncertainty and lexical inaccessibility (Brennan & Williams, 1995; Hartsuiker & Notebaert, 2010; Smith & Clark, 1993). We hypothesized that exposure to misleading details in the post-event narrative would cause temporary inaccessibility of original event details. As such, even when these correct details are reported, they will be accompanied by more hesitation markers as compared with non-misinformation conditions. Further, we hypothesized that testing that precedes the post-event narrative should increase learning of new misleading details, which in turn will result in temporary inaccessibility of original event details (e.g., Gordon & Thomas, 2014; Gordon, Thomas, & Bulevich, 2015; Thomas, Bulevich, & Chan, 2010). We hypothesize that temporary inaccessibility or increased retrieval effort will be indicated by an increase in hesitation markers associated with correctly retrieved details. Thus, immediate retrieval followed by the presentation of misinformation should result in more accompanying hesitation markers than misinformation presentation that does not follow immediate retrieval. This prediction emerges from the finding that retrieval prior to the presentation of a post-event narrative increased learning of details in the narrative and also increased response latencies on a final test of memory on misleading trials (e.g., Thomas *et al.*, 2010). Increased response latencies suggested that participants had to spend more time deciding between multiple possible final answers. Thus, the prediction regarding hesitation markers would provide convergent evidence that retrieval prior to the presentation of a post-event narrative would result in improved learning of misleading details, and decreased access to original event details. These hesitation markers would indicate the decision participants would make between the

well-established memories for the original event and the test-potentiated misleading details.

## METHOD

### Participants

Forty-two undergraduate students ( $n = 21$  in each testing condition) from Tufts University in Massachusetts participated in this experiment for credit toward a course requirement.

### Design

We employed a 2 (Retrieval Group: immediate vs. delayed)  $\times$  3 (Retrieval Attempt: B1, A1, and A2)<sup>1</sup>  $\times$  3 (Item Type: consistent, control, and misleading) mixed variables design. Retrieval Group was manipulated as a between-participants variable. Retrieval Attempt and Item Type were manipulated as within-participants variables. Importantly, the number of retrieval attempts made was dependent on retrieval group. Participants in the *immediate* group made all three attempts. Participants in the *delayed* group only made retrieval attempts after the presentation of the post-event narrative (A1 and A2).

### Materials and procedure

After giving informed consent, participants were presented with a 24-minute video clip from the 1955 film *Rififi* (Vuattoux & Dassin, 1955). *Rififi* is a crime film shot using black and white film stock and contains a long scene in which four thieves commit a burglary. Participants were presented with the burglary scene as the witnessed original event. This scene was chosen because it portrayed a crime, and it served as a considerably richer 'witnessed event' than stimuli used in early misinformation studies such as those used by Pickel (1999). After presentation of the burglary scene, participants engaged in a 5-minute distractor task in which they completed a paper and pencil Sudoku puzzle. Following this brief retention interval, the *immediate retrieval* group made a verbal report of the previously witnessed scene, whereas the *delayed retrieval* group made a verbal report of an unrelated event.

Prior to making their first verbal report, participants in the *immediate* group were first given pictures of the four characters in the previously presented scene. Each character was given an arbitrary label (Persons A, B, C, and D) so that participants could refer to those labels in the free-recall report of the witnessed event. This was necessary because the previously presented scene included no audio dialog and thus no reference to names. As such, we wanted participants across all recall groups to use the same designations when referring to specific characters. Participants were allowed to refer to these pictures throughout their recall attempts. Participants were then read the following instructions:

<sup>1</sup> B1 refers to the retrieval attempt made before the presentation of the audio narrative. A1 refers to the first retrieval attempt made after the presentation of the audio narrative. A2 refers to the second retrieval attempt made after the presentation of the audio narrative.

*Please provide a detailed description of the video you just witnessed. Be as detailed as possible and report everything that comes to mind. For example, you could describe the people in a given scene. Descriptions should include what people were wearing, what they were doing, and their surroundings. Make sure to speak slowly. Keep the microphone at least a few inches from you to help with the sound quality of your recording.*

Participants then made a verbal report of the previously witnessed scene.

Participants in the *delayed* group were instructed to recall a recent event in detail. The recent event could have been the writing experience of a research paper or the contents of their favorite class in college. As in the *immediate* group, participants were encouraged to give as many details as possible and to speak slowly. Participants were told to speak for at least 5 minutes. Average length of verbal reports made directly after viewing the video did not differ between the two groups. That is, the *immediate retrieval* group, who made a verbal report about the previously presented video, spent on average 5.67 minutes reporting on the witnessed events. Participants in the *delayed* group, who reported on a control event, spent on average 5.27 minutes reporting.

Following this first report, all participants were given a second Sudoku puzzle to complete. After 5 minutes, the puzzle was removed, and participants were presented with an audio recording that described the previously viewed scene from Riffi. The audio narrative was 4.30 minutes in length and was constructed to include details consistent and inconsistent with the video. In addition, control details were included as baseline. Consistent, inconsistent or misleading, and control details were counterbalanced using a Latin square procedure. As an example, participants in one counterbalance may have been given a narrative that included the following sentence: 'They are taken to an *elevator*.' *Elevator* was the critical detail and was consistent with what was originally viewed in the video. In a second counterbalance, participants may have been given the following: 'They are taken to the *stairs*.' *Stairs* serves as a misleading detail. Finally, in a control of this item, participants in a third counterbalance were given the following: 'They are taken to their *apartment*.' The *apartment* serves as a control because the specific detail of how the individuals were taken to the apartment was omitted. Eighteen critical items were included in the narrative and distributed equally across item type, such that six served as misleading, six as consistent, and six as control.

The critical items were selected based on pilot testing. In a separate pilot experiment, 20 participants watched the video clip and were given a 70-item cued recall test to assess encoding of specific central and peripheral details. Items were selected based on performance on this test. Items chosen were ones in which the average recall performance for that item did not fall below .40 and not above .70. That is, we chose items that participants were likely to correctly recall on a cued recall test about 50% of the time.

Following the narrative, participants were instructed to complete a brief vocabulary test. This test took approximately 5 minutes to complete. Afterwards, all participants engaged in a free-recall test of the originally witnessed event.

All participants were given instructions to recall the originally witnessed video. Instructions used for the first recall round associated with the *immediate retrieval* group were used again for both groups. Participants in the *immediate* and *delayed* groups did not differ in average length of report ( $M_{\text{immediate}} = 5.67$ ;  $M_{\text{delayed}} = 6.02$ ). A 5-minute retention interval filled with a Sudoku puzzle followed. Then, participants were told to give a final report of the originally witnessed event. Again, the same instructions used for the *immediate* group were used. Again, average length of report did not differ between the two groups ( $M_{\text{immediate}} = 5.52$ ;  $M_{\text{delayed}} = 5.49$ ). Upon completion of the final report, participants were debriefed and thanked for their time.

By the end of the study, participants in the *immediate retrieval* group produced three verbal reports of the originally witnessed event, with one report occurring before the presentation of the post-event narrative. Participants in the *delayed retrieval* group produced two verbal reports of the witnessed event, both occurring after presentation of the post-event narrative.

## ANALYSES

### Transcription

Orthographically correct transcriptions of participants' verbal reports were typed using the transcription software, F5 (<http://www.audiotranskription.de/english/f5>). Participants' reports were transcribed to analyze discourse markers in speech. All speech and verbal utterances were transcribed and double checked, by two separate transcribers. Transcription included hesitation markers, punctuation, nonlinguistic events, pauses, false starts, and repetitions. Punctuation markers were used in the conventional way. Question marks were transcribed for questions, exclamation marks for unusual cases of exclamation, and so on. Any nonlinguistic events were transcribed within parentheses (i.e., human noise for coughing, clearing of throat, or yawning). Any clearly discernible pauses of several seconds were noted in brackets (i.e., (5) for 5 seconds pause). Minimal hesitations that seemed to disrupt the flow and fluency of natural speech were similarly noted (i.e., (-) for a quarter second hesitation and (-) for a half second hesitation). Segmentation of speech, where the participants restarted a phrase or repeated part of a phrase, was noted by a dash and a line break. Discourse markers included in analysis were hesitation markers and pauses. The inclusion of these varied hesitation types was based on early research examining word predictability in the context of hesitations (see Henderson, Goldman-Eisler, & Skarbek, 1965; Tannenbaum, Williams, & Hillier, 1965) and further inspired by the methodology of cognitive discourse analysis (Tenbrink, 2015) that aims at identifying linguistic markers that reflect cognitive processes of various kinds.

### Verbal hesitations and pauses: aggregation

Verbal transcriptions were subjected to a specifically tailored program that coded for specific key items. The program was written to code for the experimentally determined 24 critical details. Eighteen critical details were presented in the video.



An additional six critical details (i.e., misleading details) were presented in the audio synopsis. If a critical detail was uttered, the program then coded whether hesitations (e.g., um and uh) or pauses ranging from .25 to 3 seconds occurred up to 10 items before or after the critical detail. In this study, the only hesitations uttered were *um* and *uh*. Pauses did not last longer than 3 seconds. Previous research suggests that unintentional items that occur up to 10 verbal positions before a detail of interest may be relevant to the preparation of speech for that detail (Frieda Goldman-Eisler, 1958; F. Goldman-Eisler, 1961a, 1961b; Tannenbaum *et al.*, 1965). This research served as rationale for our choice of 10 verbal positions. In this context, a verbal position is defined as any linguistic expression or a pause longer than .25 seconds. Nonlinguistic items, such as sneezes and coughs, were not included in verbal position counts.

Initial analyses found that verbal hesitations and pauses were randomly distributed across the 10 position points under examination. Verbal hesitations and pauses were tallied by two independent raters. Once tallied, these verbal markers were subjected to preliminary analyses to determine whether a specific kind of item was more common than another. No differences across these items were found. As such, verbal hesitation markers ('um', 'uh', and pauses) were aggregated for a given critical detail, and the average number of hesitations associated with each critical detail class (consistent, control, and misleading) was computed for each participant in the context of each recall attempt.

## RESULTS

### Verbal hesitations and pauses: statistical analysis

The primary focus of the following analyses was on average verbal hesitations associated with the production of correct details, or details presented in the original video. We compared average hesitations across verbal reports, and as a function of misinformation presentation. Recall that participants in the *immediate retrieval* group produced three verbal reports related to the original event. We begin by analyzing hesitations across these three verbal reports only in this testing group.

As can be seen in Figure 1, verbal hesitations varied as a function of both number of retrieval attempts and type of

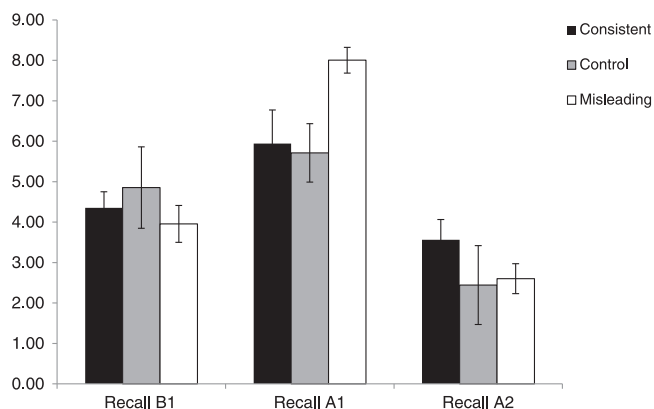


Figure 1. Average hesitations associated with accurate detail retrieval as a function of recall output for the *immediate retrieval* group (standard error plotted)

critical item. In general, by the time participants in the *immediate retrieval* group provided their third report (second after the audio synopsis), hesitations in association with retrieval of correct critical details were infrequent. As predicted, hesitations spiked in the second recall, and that spike was most apparent on misleading trials. A 3 (Recall Output: B1, A1, and A2)  $\times$  3 (Item Type: consistent, control, and misleading) within-subjects analysis of variance (ANOVA) on average hesitations associated with *correctly retrieved details* yielded a main effect of Recall Output,  $F(2, 40) = 16.35, p < .001, \eta_p^2 = .45$ . Pairwise comparisons employing a Bonferroni correction confirmed that participants produced more hesitations in the second recall output ( $M = 6.55$ ) as compared with the first ( $M = 4.38$ ),  $t(20) = 3.13, d = .83$ . In addition, fewer hesitations were produced in the third output ( $M = 2.86$ ) as compared with the second,  $t(20) = 4.93, d = 1.13$ . Finally, the third output resulted in fewer hesitations than the first,  $t(20) = 3.13, d = .99$ . We also found a significant interaction between item type and recall output,  $F(4, 80) = 3.39, p < .05, \eta_p^2 = .15$ . This interaction was driven by average hesitations associated with production of critical details after exposure to conflicting information in the narrative. As Figure 1 illustrates, there were more hesitations across all classes of items after the presentation of the post-event narrative; however, the average number of hesitations was particularly high on misleading trials ( $M = 8.00$ ). That is, after reading misleading details, participants were more likely to hesitate with a pause or unintentional verbal expression when they reported the correct detail, as compared with when they read consistent or control details. Average hesitations associated with correct detail production on misleading item trials dramatically decreased by the third recall output ( $M = 2.60$ ).

Average verbal hesitations were also examined as a function of group (Figure 2). This series of analyses were conducted to examine how retrieval prior to the presentation of the post-event narrative would affect pauses and other hesitations in verbal reports made after the narrative. We hypothesized that participants in the *immediate retrieval* group would incorporate more hesitations in verbal reports made directly following the narrative than participants in the *delayed retrieval* group. A 2 (Group: immediate and delayed)  $\times$  2 (Recall Output: A1 and A2)  $\times$  3 (Item Type: consistent, control, and misleading) mixed-design ANOVA on average hesitations was performed to test this hypothesis. We found a significant interaction among these three independent variables,  $F(2, 80) = 4.49, p < .05, \eta_p^2 = .10$ . To decompose this interaction, we conducted two separate 2 (Group: immediate and delayed)  $\times$  3 (Item Type: consistent, control, and misleading) mixed-design ANOVAs on average hesitations for each recall output. The first ANOVA was conducted on average hesitations associated with recall output A1, or the first report made *after* presentation of the narrative. We found a significant interaction between Group and Item Type,  $F(2, 80) = 4.78, p < .05, \eta_p^2 = .11$ . Pairwise comparisons employing a Bonferroni correction found that these two groups differed only on misleading item trials,  $t(40) = 4.62, d = 1.47$ . Participants in the *immediate retrieval* group produced significantly more hesitations than participants in the *delayed retrieval* group. The second ANOVA on hesitation

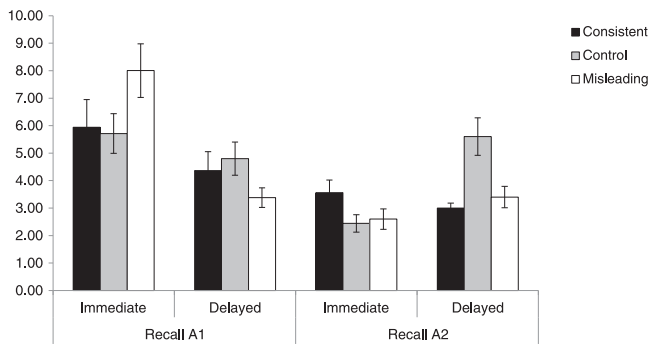


Figure 2. Average hesitations associated with correct detail retrieval as a function of retrieval group and item type (standard error plotted)

markers associated with the final recall output also found a significant interaction between Group and Item Type,  $F(2, 80) = 10.41, p < .001, \eta_p^2 = .21$ . However, this interaction was driven by hesitation markers associated with control trials. Pairwise comparisons employing Bonferroni corrections found that participants in the *delayed* group ( $M = 5.6$ ) produced significantly more hesitations on control trials as compared with participants in the *immediate* group ( $M = 2.44$ ),  $t(40) = 4.19, d = 1.30$ .<sup>2</sup>

**Video memory accuracy**

The examination of memory accuracy centered on verbal production of targeted critical details. Transcriptions were coded for the 18 critical details represented in the video. Those details were categorized into three groups (consistent, control, and misleading). This categorization is only meaningful after the post-event narrative, which included details consistent and inconsistent with the video. Thus, verbal reports made prior to the presentation of the narrative should not result in systematic differences in detail production as a function of item type. Table 1 demonstrates this particular finding. That is, prior to receiving the audio narrative, participants were as likely to produce details from consistent, control, and misleading categories,  $F < 1$ . However, following the synopsis, this pattern changed.

To analyze memory for critical details following the synopsis, we conducted a 2 (Group: immediate and delayed)  $\times$  2 (Recall Output: A1 and A2)  $\times$  3 (Item Type: consistent, control, and misleading) mixed-design ANOVA on

Table 1. Correct critical detail responding as a function of recall attempt and item type (standard error in parentheses)

	Consistent	Control	Misleading
<b>Immediate retrieval group</b>			
B1 retrieval before narrative	.39 (.04)	.42 (.04)	.39 (.04)
A1 retrieval after narrative	.51 (.05)	.62 (.06)	.41 (.06)
A2 retrieval after narrative	.56 (.06)	.62 (.06)	.42 (.06)
<b>Delayed retrieval group</b>			
A1 retrieval after narrative	.51 (.05)	.38 (.04)	.34 (.03)
A2 retrieval after narrative	.50 (.05)	.39 (.04)	.27 (.03)

average correct critical detail production in verbal reports that followed the presentation of the post-event narrative. We found a main effect of Item Type,  $F(2, 80) = 11.71, p < .001, \eta_p^2 = .23$ . After exposure to the audio narrative that included misleading details, participants in both the *immediate* and *delayed retrieval* groups were less likely to produce correct information, or information from the video, on misleading trials ( $M = .36$ ) as compared with control ( $M = .48$ ) and consistent ( $M = .52$ ) trials. Pairwise comparisons between control and misleading, and consistent and misleading trials were both significant after a Bonferroni correction, [ $t(41) = 3.87, d = .56; t(41) = 4.15, d = .74$ ]. This finding confirms the standard misinformation effect. In addition, Recall Output significantly interacted with Group,  $F(1, 40) = 4.09, p < .05, \eta_p^2 = .10$ . As can be seen in Table 1, recall accuracy increased between the second and third attempts for participants in the *immediate* group but declined for participants in the *delayed* group. Finally, the interaction between Item Type and Group was marginally significant,  $F(2, 80) = 2.58, p = .08, \eta_p^2 = .06$ .

**Recalling narrative critical details (misinformation production)**

Recall that six misleading details were presented in the audio narrative. A 2 (Group: immediate and delayed)  $\times$  2 (Recall Output: A1 and A2) mixed-design ANOVA performed on average production of these misleading details found a main effect of Recall Output,  $F(1, 40) = 9.18, p < .005, \eta_p^2 = .19$ . Participants were more likely to produce misleading details on the final recall report ( $M = .17$ ) as compared with the recall report made immediately after presentation of the audio synopsis ( $M = .08$ ). No other effects were significant. Thus, repeated post-narrative recall increased the likelihood that suggested details would be included in verbal reports.

**GENERAL DISCUSSION**

Various forms of hesitation are common to almost all speakers across all languages. Hesitations interrupt the normal flow of speech and can indicate to the listener that the speaker may be uncertain or unprepared. In the present study, we demonstrated that repeated retrieval of a witnessed event reduced these interruptions. However, this conclusion is qualified by a number of important factors. Inclusion of hesitations was influenced not only by the number of times a report was made but also by *when* the first report was made, and by reexposure to information related to the original event. Participants who produced three verbal reports demonstrated a dramatic reduction in hesitations markers associated with correctly produced critical details. However, hesitation markers were affected by the post-event narrative. Original event detail production after the presentation of misleading details were associated with longer pauses and more verbal hesitations than original event detail production that occurred after the presentation of consistent or neutral

<sup>2</sup> We had also planned to examine hesitation markers associated with the production of misleading details in Recall Outputs 2 and 3; however, production of misleading items was low in both recall outputs, and thus, the number of missing values was too high to conduct this analysis.

information in the narrative. That is, when participants were able to correctly access original event details, they did so with more difficulty after the presentation of misleading information than after the presentation of neutral or consistent information. This finding is not surprising in the broader context of the misinformation literature. It has long been known that the presentation of misleading information negatively impacts memory for original event details (cf., Loftus, 2003). In addition to replicating this standard misinformation effect, the present study extends misinformation disruption to linguistic markers associated with free-recall reports. We argue that misleading details interfered with access to original event details. Difficulty accessing original details was manifested behaviorally through verbal and nonverbal hesitations.

Further, we found that number of recall reports had dramatic effects on verbal and nonverbal hesitations that accompanied critical details. Participants who produced a verbal report before the narrative included significantly more hesitation markers in the verbal reports that immediately followed the narrative, than participants who made their first report after the narrative. This finding is particularly striking, because it suggests that the immediate verbal report resulted in reduced accessibility to original event details and in more uncertainty surrounding details on misleading trials. Thus, verbal hesitations may indicate an internal struggle to either access the correct detail or decide between two relevant details. Importantly, this internal struggle was more apparent in the *immediate* group as compared with the *delayed* group. Additionally, changes in hesitations across post-narrative reports suggest that the internal struggle between competing pieces of information decreased with each report. By the time participants made the last verbal report, verbal and nonverbal hesitations were at their lowest points. This finding is consistent with previous research demonstrating that repeated interview reduced uncertainty in reports in young children (Saykaly *et al.*, 2013).

### FLUCTUATIONS IN CERTAINTY AND ACCURACY ACROSS MULTIPLE REPORTS

The present study suggests that retrieval practice that occurs before and after the presentation of new but related information may increase uncertainty in free-recall reports. Uncertainty is greatest immediately following presentation of new information in the form of the post-event narrative. This finding is consistent with previous research that demonstrated a disruption in the relationship between confidence and accuracy in the misinformation and retrieval-enhanced suggestibility (RES) paradigm. For example, Thomas *et al.* (2010) found that participants who took a cued recall test before the presentation of a post-event narrative that included details inconsistent with the original event were overconfident in produced suggested details, and under-confident in correct details associated with misleading trials. The paradigm used by Thomas *et al.* is similar to the present one. That is, participants were exposed to an event, took an immediate test, and then were given a post-event narrative. A final test occurred shortly after the presentation of the post-event narrative. The first and second verbal

reports correspond to the tests employed in Thomas *et al.* The primary difference between the two studies is the kind of test used. Whereas Thomas *et al.* used identical cued recall tests, the present study employed free-recall reports. However, regardless of type of test, immediate retrieval followed by misleading post-event information increased uncertainty even when participants were correct on trials that included misleading details. In Thomas *et al.*, uncertainty was measured by confidence in memory judgments. In the present study, uncertainty was indexed by hesitations and unintentional expressions.

Increased uncertainty on misleading trials may be an indirect result of test-potentiated learning. In the context of a misinformation paradigm, after testing, participants learn misinformation better and sometimes have difficulty discounting the incorrect information on final memory tests (Gordon & Thomas, 2014). Gordon *et al.* (2015) argued that test-potentiated learning in the RES paradigm may occur because initial test questions serve as endogenous cues and direct participants' study efforts during post-retrieval learning. They also suggested that exposure to information in the narrative that conflicts with original memory may guide attention and processing resources. Importantly, in the present study, participants who took the initial free-recall test of video details before exposure to the post-event narrative were *not* more likely to produce misinformation on the final test of memory. However, the present experiment may not have been sufficiently powered to find this RES effect. In fact, RES was found in the context of a free-recall task but only with a much larger sample of participants (see Wilford, Chan, & Tuhn, 2014).

The present study also demonstrated that increased uncertainty in the *immediate retrieval* group was temporary. That is, these participants demonstrated a decrease in hesitation markers that accompanied misleading trial details as compared with the *delayed retrieval* group on the final free report. In addition, *immediate retrieval* participants demonstrated greater accuracy on the final test of memory than participants whose first report occurred after the presentation of the post-event narrative. This finding suggests that immediate retrieval may have important positive consequences for original event memory. Akin to the *testing effect* (e.g., Zaromb & Roediger, 2010), in the context of the present study, testing immediately following new learning resulted in reliable long-term retention. Unlike standard testing effect studies, the present experiment did not compare a retrieval practice group to a study practice group. Rather, we compared an interleaved study-testing group to a delayed testing group. In the context of both verbal learning (Tulving & Watkins, 1974) and event memory (Gordon *et al.*, 2015), retrieval practice prior to new learning has been shown to potentiate learning of post-event details. In the present study, this potentiation effect was demonstrated on the test immediately following the post-event narrative by an increase in verbal hesitations that accompanied the production of original event details on misleading trials. However, this *forward effect* of testing did not come at the expense of standard *backward effects*. That is, interleaved study and testing resulted in better memory for the original event than delayed testing. This benefit was apparent on the final test of memory.



## CONCLUSIONS

Verbal hesitations in unprepared speech convey important information to the listener. These markers indicate when speakers are having difficulty constructing a message and planning speech (Arnold et al., 2000; Clark & Fox Tree, 2002; Hartsuiker & Notebaert, 2010; Krahmer & Swerts, 2005). The present research also suggests that these markers may index a decision process between two competing pieces of information. Although unintentional, eyewitnesses may produce nonlinguistic items when internally deciding between the original and misleading post-event information. Not surprisingly, this decision process changes with repeated retrieval. Participants become more confident in the verbal output, and thus, verbal hesitations decline. With repeated retrieval comes certainty in response. This should come as no surprise to those of us who regularly give presentations. A common practice in public speaking is rehearsal. Eyewitnesses engage in a similar practice before testifying. Practice serves to reduce and potentially eliminate unintended items. Finally, in a naturalistic setting where eyewitnesses are often encouraged to give unprompted reports (see U.S. Department of Justice, 2003), the unprompted initial responses may serve to develop a strong memorial representation of the original event. Participants in the present study who produced a report before new learning demonstrated better final memory than those whose first report was delayed. Thus, even when incorrect or suggested details are presented at a later point, the unprompted initial responses may serve to facilitate the development of a memory representation of the original event, and ultimately good discrimination between original and post-event details.

## ACKNOWLEDGEMENTS

The authors would like to thank Aron Kaufman and Mark Westerfield for their efforts in transcribing verbal reports and coding data. Funding from the DFG (German Science Foundation) for the SFB/TR 8 Spatial Cognition, project I6-[NavTalk], is gratefully acknowledged.

## REFERENCES

- Arciuli, J., Mallard, D., & Villar, G. (2010). "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics*, 31(3), 397–411. doi: 10.1017/S0142716410000044
- Arnold, J., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 32–68.
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2), 123–147. doi: 10.1177/00238309010440020101
- Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34(3), 383–398. doi: 10.1006/jmla.1995.1017
- Bruck, M., Ceci, S. J., & Hembrooke, H. (2002). The nature of children's true and false narratives. *Developmental Review*, 22(3), 520–554. doi: 10.1016/S0273-2297(02)00006-0
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111. doi: 10.1016/S0010-0277(02)00017-3
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, 29(2), 320–326. doi: 10.3758/BF03194926
- Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science*, 20(1), 20–23. doi: 10.1177/0963721410396620
- Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *The Quarterly Journal of Experimental Psychology*, 10, 96–106. doi: 10.1080/17470215808416261
- Goldman-Eisler, F. (1961a). Continuity of speech utterance, its determinants and its significance. *Language and Speech*, 4(4), 220–231.
- Goldman-Eisler, F. (1961b). The distribution of pause durations in speech. *Language and Speech*, 4(4), 232–237.
- Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition*, 42(2), 186–197. doi: 10.3758/s13421-013-0361-2
- Gordon, L. T., Thomas, A. K., & Bulevich, J. B. (2015). Looking for answers in all the wrong places: How testing facilitates learning of misinformation. *Journal of Memory and Language*, 83, 140–151.
- Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, 57(3), 169–177. doi: 10.1027/1618-3169/a000021
- Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1965). The common value of pausing time in spontaneous speech. *The Quarterly Journal of Experimental Psychology*, 17(4), 343–345. doi: 10.1080/17470216508416454
- Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, 48(1), 29–53. doi: 10.1177/00238309050480010201
- Loftus, E. F. (2003). Make-believe memories. *American Psychologist*, 58(11), 867–873. doi: 10.1037/0003-066X.58.11.867
- Pickel, K. L. (1999). Distinguishing eyewitness descriptions of perceived objects from descriptions of imagined objects. *Applied Cognitive Psychology*, 13(5), 399–413. doi: 10.1002/(SICI)1099-0720(199910)13:5<399::AID-ACP603>3.0.CO;2-3
- Saykaly, C., Talwar, V., Lindsay, R. C. L., Bala, N. C., & Lee, K. (2013). The influence of multiple interviews on the verbal markers of children's deception. *Law and Human Behavior*, 37(3), 187–196. doi: 10.1037/lhb0000023
- Saykaly, C., Talwar, V., Lindsay, R. C. L., Bala, N. C., Lee, K., Bertrand, M., & Nugent, S. M. (2013). Adults' ability to detect deception of stressful and non-stressful stories of children. *Psychology, Crime & Law*, 19(10), 865–879. doi: 10.1080/1068316X.2012.700311
- Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology Learning Memory and Cognition*, 12(2), 171–181. doi: 10.1037/0278-7393.12.2.171
- Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32(1), 25–38. doi: 10.1006/jmla.1993.1002
- Tannenbaum, P. H., Williams, F., & Hillier, C. S. (1965). Word predictability in the environments of hesitations. *Journal of Verbal Learning & Verbal Behavior*, 4(2), 134–140. doi: 10.1016/S0022-5371(65)80097-4
- Tenbrink, T. (2015). Cognitive discourse analysis: Accessing cognitive representations and processes through language data. *Language and Cognition*, 7(1), 98–137.
- Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language*, 63(2), 149–157. doi: 10.1016/j.jml.2010.04.004
- Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning & Verbal Behavior*, 13(2), 181–193. doi: 10.1016/S0022-5371(74)80043-5
- Vuattoux, R. G., & Dassin, J. (1955). *Du rififi chez les hommes*. Pathé: France.
- Wilford, M. M., Chan, J. C. K., & Tuhm, S. J. (2014). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied*, 20(1), 81–93. doi: 10.1037/xap0000001
- Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995–1008. doi: 10.3758/MC.38.8.995