DOI: 10.1177/2167702617744326

Explaining memory amplification: Is it all about the test format?

Melanie K. T. Takarangi[1]

Jacinta M. Oulton[1]

&

Deryn Strange[2]

[1]School of Psychology, Flinders University, Australia (email: jacinta.oulton@flinders.edu.au; melanie.takarangi@flinders.edu.au)

[2]Department of Psychology, John Jay College of Criminal Justice, USA (email: dstrange@jjay.cuny.edu)

Corresponding Author:
Melanie Takarangi
School of Psychology
Flinders University
GPO Box 2100
Adelaide
South Australia, 5001
melanie.takarangi@flinders.edu.au

Abstract

Trauma-exposed people commonly exhibit a “memory amplification” effect, endorsing exposure to more traumatic events over time. Studies reporting this phenomenon have typically relied on checklists, where participants read event descriptions and indicate (yes/no) their exposure. We examined whether that approach is vulnerable to response biases and memory errors. In two experiments, participants viewed negative photos and completed an Old-New recognition test. In Experiment 1, participants completed either a photo recognition test or description test—comprised of written *descriptions of* negative photos. In Experiment 2, we measured analogue PTSD symptoms and participants completed the description test twice, 24-hours apart. The description-test condition performed worse on the memory test and were more biased to endorse negative photos compared to the photo-test condition. Further, this bias to endorse negative photos increased over time and was related to analogue PTSD symptoms. Overall, our findings suggest that test format plays a role in memory amplification.

Explaining memory amplification: Is it all about the test format?

Victims of trauma often endorse exposure to more traumatic events over time: the memory “amplification” effect (e.g., Southwick, Morgan, Nicolaou, & Charney, 1997). For example, when responding to a list of traumatic event descriptions, a veteran may indicate that s/he has not experienced “seeing human remains” but later claim to have had that experience. Importantly, memory amplification appears problematic: it is consistently associated with Posttraumatic Stress Disorder symptoms (e.g., Engelhard et al., 2008; Southwick et al., 1997). Thus, understanding what processes lead to amplification, and *why* amplification occurs is critical. Discrepancies in reports may arise because participants are poor at distinguishing experienced and non-experienced events, a memory distortion process. Indeed, misremembering negative photos over time is associated with worse outcomes (Oulton, Takarangi, & Strange, 2016). However, discrepancies could also represent changes in how people *interpret* unaltered memories (Engelhard & McNally, 2015); participants might reinterpret the sight of blood as “seeing human remains,” a shift in response bias. The typical test format to assess trauma exposure—sometimes vague event descriptions that might be considered subjective (e.g., “*witnessing violence*”; Engelhard et al., 2008)—may be especially vulnerable to reinterpretation. Of course, both distortion and reinterpretation processes may contribute to memory amplification. Here, we examined whether *test format* influences people’s ability to distinguish between experienced and non-experienced events, and/or their willingness to endorse them over time.

Memory amplification occurs across various samples, including Vietnam veterans (King et al., 2000; Southwick et al., 1997), American peacekeepers (Roemer, Litz, Orsillo, Ehlich, & Friedman, 1998) and 9/11 disaster restoration workers (Giosan, Malta, Jayasinghe, Spielman, & Difede, 2009). In one example (Engelhard, van den Hout, &

McNally, 2008), Dutch Army soldiers reported whether they had been exposed to war-related stressors (e.g., being shot at) at five months and 15 months post deployment. At 15 months, most (70%) soldiers reported exposure to at least one stressor they never endorsed earlier. Importantly, soldiers with more PTSD symptoms overall made more no-to-yes response changes. Other researchers have found small, but significant correlations between PTSD symptoms at follow-up and number of no-to-yes changes over time: 0.20, 95% CI [0.16, 0.24] (Giosan et al., 2009), 0.23 [0.06, 0.38] (Engelhard et al., 2008), 0.26 [0.22, 0.30] (King et al., 2000) and 0.32 [0.17, 0.60] (Southwick et al., 1997).

One problem with field studies is that trauma exposure cannot be corroborated. Thus, we do not know whether memory amplification reflects an increase in memory distortion or another process, such as people underestimating events initially and responding more accurately at follow-up, or reappraising the significance or meaning of an event. They might also exaggerate reports in response to demand characteristics or to enhance their self-image as a brave and courageous person (knowingly or unknowingly). Indeed, some studies suggest that reports of combat exposure among veterans are sometimes exaggerated or misrepresent their true involvement (e.g., Frueh et al., 2005). We developed a lab-based paradigm to address the problem of corroboration (Oulton et al., 2016). Participants viewed negative photos of graphic scenes (e.g., death and mutilation). Following tradition in cognitive research, participants then completed a standard recognition test—whereby participants identified photos as “old” (previously seen) or “new” (previously unseen)—on two occasions, one-week apart. Participants’ ability to distinguish between old and new photos (i.e., their sensitivity) decreased over time, but so did their tendency to endorse negative photos (i.e., response bias). However, mirroring the field research, among participants who remembered *more* photos over time, re-experiencing symptoms were associated with an increase in “old” responses (r=.28). This result suggests memory distortion at least partially accounts for amplification.

Importantly however, our memory test format differed from the field research, which has typically administered checklists: participants read *verbal* descriptions of specific stressors (e.g., “*seeing others killed or wounded*” and “*death of a close friend*”; Southwick et al., 1997) and indicate (yes/no) whether they have previously been exposed (e.g., Giosan et al., 2009) and/or the extent of exposure on Likert-type scales (e.g., Roemer et al., 1998). When shown response discrepancies, participants typically explain that they interpreted items—or key words on the checklist—differently the second time (Engelhard & McNally, 2015).

Could reinterpretation[1] contribute to memory amplification? Engelhard et al. (2008) suggest PTSD sufferers are prone to subjective reappraisal; they make sense of their current symptoms by attributing those symptoms to stressors during service, thus events previously considered irrelevant gain importance. Further, Englehard et al. argue that PTSD sufferers can be unwilling or unable to determine whether something actually happened to them and hence prone to gist-based retrieval strategies. For example, a veteran who is experiencing severe PTSD symptoms might be less willing to search their memory for events during service—due to the distress the process provokes—and therefore likely to respond “yes” to any checklist item that seems perceptually or conceptually familiar. We wondered whether these strategies would be especially likely when verbal descriptions are used, as in the field research. Thus, in Experiment 1, we examined whether people would be more likely to falsely recognize photos or descriptions of those photos; and whether they would be more biased to respond “old” to new photos or new photo descriptions. Specifically, to separate people’s ability to discriminate between old and new photos (i.e., sensitivity)—which

[1] Of course, reinterpretation could occur in the opposite direction too. After a period of time and/or reflection, people might begin to doubt that events they experienced really occurred and hence make memory omission errors. This explanation would explain why field studies also find some yes-to-no errors over time.

depend on a memory trace's strength—and people's tendency to respond to "old" to test items (i.e., response bias), we used a signal detection approach (e.g., Stainslaw & Todorov, 1999).

Recognition judgments are based on two distinct mechanisms (e.g., Mandler, 1980): recollection—the specific retrieval of previously studied items—and a sense of familiarity. Certain factors—including how memory is tested—can affect how much each mechanism contributes to a given judgment. Thus we had competing predictions for sensitivity and response bias. Turning first to sensitivity, because participants in the photo-test condition see studied and test items in the same format, they might experience more difficulty distinguishing old and new test items, compared to those who study photos but are tested on descriptions. Indeed, people falsely recognize "new" (never seen) items more frequently when they are conceptually or perceptually similar to studied items (e.g.,"*twister*" vs "*funnel*", Schacter, Verfaellie, & Anes, 1997; pictures of objects from the same category, Koustaal & Schacter, 1997). Thus, the overlap in perceptual information between study and the photo-test should lead to poorer sensitivity.

Similarly, participants might show a bias to respond "old" to photos more than descriptions, because when viewing the test photos they rely on feelings of familiarity to make their judgments. Importantly, familiarity might arise for "new" photos, because they share perceptual features with previously encoded pictures, for example, leading to source monitoring errors: new photos judged to be "old" (Johnson, Hashtroudi, & Lindsay, 1997). Conversely, participants may question whether descriptions are accurate depictions of encoded photos and therefore respond cautiously, requiring more evidence to label description test items as "old".

However, because photos contain perceptually rich and specific information, they might constrain people's ability to imagine new details, relative to descriptions.

Imagination lays the groundwork for false memory creation, increasing confidence an event occurred (Garry, Manning, Loftus, & Sherman, 1996) and producing clearer, more complete memories of false events (Scoboria, Wade, Lindsay, Azad, Strange, Ost & Hyman, 2017). Critically however, the mode of suggestion matters: false memories are more likely when participants receive the suggestion via a *description* compared to a *photo* (e.g., Garry & Wade, 2005). Therefore, participants may experience more difficulty distinguishing between old and new descriptions than between old and new photos.

In relation to response bias, participants may be more liberal with descriptions, because they could describe more than the specific photo to which they are meant to refer. Israel and Schacter (1997) found that when people encoded images instead of words they were less likely to respond "old" at test. They argued that participants demand a distinct recollection of a test item before saying "old" if they see photos at encoding. Therefore, participants might respond "Old" according to how easy it is to retrieve a certain memory that fits with the description, or the degree of perceptual information that memory entails.

In summary then, we had competing predictions; there are theoretical reasons to anticipate a memory advantage and liberal response bias towards both photos and photo descriptions. However, if photo descriptions lead to poorer sensitivity and liberal responding, our findings would suggest that test format plays a role in memory amplification. To test this idea, our participants viewed highly negative photos from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) and, 20 minutes later, completed an Old-New recognition test. Critically however, participants saw either a photo test—comprising seen and unseen negative IAPS photos—or a description test—comprised of written *descriptions of* seen and unseen negative IAPS photos. We compared participants' sensitivity and response bias on the test between these conditions.

Experiment 1

Method

Participants

We recruited participants using Amazon's Mechanical Turk: 228 participants[2] completed the Experiment (114 per condition). Participants were US residents, aged 19-74 (*M*=34.57, 95% CI [33.12, 36.02]). Over half (52.2%) were male and most identified their ethnicity as Caucasian (including White; 72.4%). Other participants identified as African American (including Black; 7.5%), Hispanic (6.6%), Asian American (6.6%), European (3.5%), or mixed ethnic origin (3.5%). We did not collect any other demographic or socioeconomic data.

Materials

Trauma analogue. We selected 64 photos—53 IAPS[3] photographs (Lang et al., 2008) and 11 additional photos (from Krans, Langner,Reinecke, & Pearson, 2013)—depicting negative scenes (e.g., violence and death). We divided photos into four sets of 16 target photos, matched for mean valence, *Fs* < 1 (Set 1: *M* = 1.93, 95% CI [1.79, 2.06], Set 2: *M* = 1.91 [1.69, 2.14], Set 3: *M* = 1.90 [1.69, 2.11], Set 4: *M* = 1.97 [1.79, 2.15]) and arousal (Set 1:*M* = 6.33 [6.12, 6.53], Set 2: *M* = 6.28 [5.85, 6.71], Set 3: *M* = 6.39 [6.03, 6.76], Set 4: *M* = 6.36 [6.03, 6.69]). Participants saw three sets of photos (48 photos total) at encoding. We counterbalanced sets across participants such that each combination appeared equally.

Positive Affect Negative Affect Schedule (PANAS). We measured positive and negative affect using the 20-item PANAS (Watson, Clark, & Tellegen, 1988). Participants

[2] We first ran a pilot study (*N*=92) to gain an estimate of Cohen's D for the possible effect of condition on response bias. Based on a precision analysis using this information (Cumming, 2013), we predetermined a sample size of at least 220 participants (110 per condition) to obtain a target margin of error (i.e., the half width of the target confidence interval) of 0.29.

[3] IAPS photo numbers: 2352, 2703, 2800, 2811, 3005, 3063, 3064, 3069, 3080, 3102, 3103, 3120, 3130, 3140, 3170, 3181, 3191, 3195, 3225, 3266, 3301, 3350, 3530, 6021, 6022, 6212, 6315, 6350, 6520, 6560, 6563, 6821, 9040, 9050, 9140, 9163, 9183, 9252, 9253, 9254, 9405, 9410, 9412, 9421, 9433, 9435, 9570, 9571, 9635, 9902, 9910, 9911, 9921

rate items (e.g., "excited" and "jittery") on a Likert scale (1=*Very slightly or not at all*, 5=*Extremely*) according to how they feel currently. The scales have excellent convergent and divergent correlations with more comprehensive mood measures and high test-retest reliability: .81 for NA and .79 for PA (Watson, Clark, & Tellegan, 1998). The Negative Affect subscale correlates highly with the Hopkins Symptom Checklist (HSCL) (r=.74) and the Positive Affect subscale has a modest negative correlation with the Beck Depression Inventory (r=-.34; Beck, Steer, & Brown, 1996). In the current study, Cronbach's alpha for Positive Affect was .93 and .91 before and after photo exposure, respectively. Cronbach's alpha for Negative Affect was .92 at both time points.

Recognition test. The recognition test consisted of either three sets of 16 photos or three sets of 16 *descriptions of photos* (e.g., "a young boy points a gun at the viewer"), depending on condition. To create photo descriptions, we showed pilot participants from Amazon Mechanical Turk a subset of 24 IAPS photos from our larger photo pool and, for each photo, asked them to "describe the event depicted in the photograph in one sentence." Using the specific details that were consistently described by participants, three research assistants created a fine-grained (i.e., accurate, informative, and descriptive details of the photo) description of each photo. For example, "*a man holds a knife to a woman's throat*" and "*two fire fighters pull an unconscious woman out of a smoke filled building*". We eliminated photos with overlapping descriptions or high similarity. For example, one photo's description was "*a mutilated, bloody, severed hand rests on a table*" while another's was "*a wounded right hand that has severe cuts and a compound fracture with bones sticking out.*" Due to this overlap, we only included one of these photos/descriptions in the final sets.

The test comprised one set of "old" items (i.e., previously presented photos or descriptions of photos) and two sets of "new" items (previously unseen photos or descriptions of photos). One set of new items were neutral IAPS photos, or their descriptions—included to ensure participants paid attention—and the other set was one of our previously unseen target negative photo/description sets. Before completing the test, participants read instructions stating that they were about to see some photos (or descriptions of photos) and that their task was to identify each as OLD or NEW. We asked the description condition to "select OLD if the description is of a photo you saw earlier in the session" and "select NEW if the description is of a photo you did not see earlier in the session." We asked photo test participants to "select OLD if you saw the same photo earlier in the session" and "select NEW if you did not see the same photo earlier in the session." Participants indicated their confidence in each decision (1=*not at all confident*, 10=*extremely confident*).

There were twelve versions of the test for both conditions, counterbalanced so that every test item appeared equally often as 'new' and 'old' across participants. For all tests photos and descriptions appeared in a random order.

Procedure

This research was approved by the Flinders University Social and Behavioural Research Ethics Committee and the City University of New York's University Integrated Institutional Review Board and conducted in accordance with the provisions of the World Medical Association Declaration of Helsinki. We administered the experiment as an online survey using Qualtrics software. We told participants that the experiment was about juror decision-making and would allow us to evaluate how different graphic visual evidence affects mock juror decision-making; in particular, whether particularly graphic material exacerbates the difficulty for jurors in remaining objective and attending to their task. We informed participants in the study advertisement and information form that participation involved viewing potentially distressing and graphic images.

We randomly assigned participants to receive either a photo test or descriptions test. Participants first completed the PANAS. Next, we instructed participants that they were about to be shown some photos depicting events, one at a time. Participants then viewed 48 target photographs in a randomized order. Each photograph appeared for 2.5 seconds. To ensure participants attended to the photos, we asked participants to rate each photo on emotional arousal (i.e., "*How emotionally arousing was the picture?*") immediately after the photo's presentation on a Likert scale (1=*not at all*, 7=*highly*). After the encoding phase, participants completed the PANAS again and responded to the question "*how closely did you pay attention to the photos presented?"* (1=*not at all*, 7=*extremely closely).* Participants also rated how disgusting, distressing and unpleasant they found the photos (1=*not at all*, 7=*extremely*). We also included an attention check following encoding (see Oppenheimer, Meyvis & Davidenko, 2009). Specifically, we asked participants to select, using a checklist, the sensory impressions they were aware of experiencing while viewing the photos. However, embedded within the text was an instruction to ignore the original question and simply type a specific word into the "other" free-test box. Participants who failed this check (i.e., did not type the word) were immediately brought to the end of the survey, and received a partial payment (90 cents). Data for non-completers were not analyzed.

Next, participants worked on unrelated Sudoku puzzles for 20 min. We based this delay on our previous memory research using IAPS photos (Oulton, Takarangi, & Strange, 2016), and similar time delays previously used in misinformation studies (e.g., Vornik, Sharman, & Garry, 2003; Wright, Self, & Justice, 2000). We then asked participants how often they found themselves thinking about the photographs since viewing them (1=*almost never*, 5=*extremely frequently*). Participants completed the recognition test. Finally, we asked participants whether they closed their eyes or looked away from the photos when first viewing them. We explicitly informed participants that their answer to this question would not affect payment. We debriefed participants at the end of the survey and paid them $4.50 for their time. Our data for Experiment 1 can be found on the Open Science Framework (OSF) at https://osf.io/6u4wh.

Results

Emotional Impact of Photos

We first examined the photos' effect on mood. We compared positive and negative affect scores before and after viewing the photos, using 2 (Photo Test, Description Test) x 2 (Time 1, Time 2) mixed ANOVAs. There were no main effects of test condition (*F*s<1). As expected, the analyses revealed significant main effects of time for positive ($F(1, 226)=114.97$, $p<.001$, $\eta_p^2= .34$, 95% *CI* [.24, .42]) and negative ($F(1, 226)=211.15$, $p<.001$, $\eta_p^2=.48$ [.39, .56]) affect. Positive affect significantly reduced (Time 1: $M=30.12$ [28.90, 31.35], Time 2: $M=25.33$ [24.19, 26.48], $d= 0.53$[4]), while negative affect significantly increased following photo exposure (Time 1: $M=12.58$ [11.94, 13.21], Time 2: $M=19.80$, [18.73, 20.88], $d=1.07$. These results confirm that the photos negatively affected participants' mood.

Photo ratings

Participants rated the photos as very unpleasant ($M=6.25$, *95% CI* [6.10, 6.41]), distressing ($M=5.72$ [5.53, 5.92]) and disgusting ($M=5.99$ [5.82, 6.16]). Participants reported paying very close attention to the photos ($M=6.73$ [6.64, 6.82]) and thinking about them moderately often during the delay period ($M=2.50$ [2.34, 2.66]; 1= *Almost never*, 5= *Extremely frequently*). The conditions did not significantly differ on any of these ratings,

[4] We used ESCI software (Cumming, 2012) to calculate confidence intervals for effect sizes. However, because CI calculation is not available for paired sample t-tests with degrees of freedom greater than 200, we do not report this information.

ps>.05. Thinking about the photos was not significantly related to response bias (r=-.05, p=.46) or sensitivity (r=-.11, p=.10).

Memory Accuracy[5]

Next, we turned to our specific hypotheses regarding memory accuracy. Recall that we were primarily interested in participants' ability to distinguish between seen and unseen photos, and their *bias* to respond "old" at test. In particular, we wondered whether test format would affect participants' tendency to respond "old" to test items. To separate participants' ability to distinguish between old and new test items (i.e., sensitivity) from their response bias, we used a signal detection method (Stainslaw & Todorov, 1999). We classified old photos/descriptions as signal events and new, negative photos/descriptions as noise events: correctly identifying an old photo or description as "old" was coded as a hit, and incorrectly identifying a new negative photo or description as "old" was coded as a false alarm. We calculated signal detection measures d' and c, where d' denotes the ability to discriminate between old and new test items and c denotes response bias. Note that $c < 0$ indicates a response bias toward saying old, and $c > 0$ is a response bias toward saying new to test items. A d' value of 0 indicates an inability to distinguish old test items from new test items, whereas larger values indicate a greater ability to distinguish old test items from new test items.

We had competing predictions about sensitivity. Either the photo test condition would be more accurate compared to the description test condition—due to participants demanding a distinct recollection of a test item and greater constraint on their imagination—or they would be less accurate, due to overlap in perceptual information between study and test items. As shown in Figure 1a, our data support the former

[5] We examined age and gender as potential moderators for the effect of condition on sensitivity and response bias. Results were non-significant for all analyses, all *ps*>.05. Excluding participants who closed their eyes or looked away from the photos (26.8%) did not change the results of the analyses for sensitivity or response bias.

prediction: participants were significantly better at distinguishing between old and new photos, compared to descriptions of old and new photos, $t(216.55)=17.84$, $p < .001$, $d=2.37$, *95% CI* [2.02, 2.70]. These findings are consistent with the assumption that photos, relative to descriptions, constrain imagination therefore reducing opportunity for false memories. Further, our data suggest that when distinctive perceptual information is available at test, people are more accurate at remembering what they have and have not seen.

We also had competing predictions about response bias. Either photo-test participants would be more biased to respond “old”, because of feelings of familiarity that arise due to shared perceptual features with previously encoded pictures, or description-test participants would show a greater bias because it is easier to retrieve a memory that fits with a description. As shown in Figure 1b, the description test condition was significantly more biased to respond “old” to the test items compared to the photo test condition, $t(210.03) = 11.21$, $p < .001$. The effect size was very large according to Cohen’s benchmarks, $d=1.49$, [1.19, 1.78].

Memory Confidence

Along with memory accuracy and bias, we examined whether test format affected participants’ confidence at test. We compared mean confidence scores for Old and New negative test items, using a 2 (Photo Test, Description Test) x 2 (Old photos, New photos) mixed ANOVA. Participants were significantly more confident when identifying old photos ($M=8.88$, [8.70, 9.06]) compared to new photos ($M=7.55$, [7.31, 7.79], $d=0.82$), a significant main effect of photo type, $F(1, 226)=307.49$, $p <.001$, $\eta_p{}^2=.58$, *95% CI* [.50, .64]. The photo test condition were significantly more confident ($M=8.88$, [8.71, 9.04]) overall when identifying photos compared to the description test condition ($M=7.55$ [7.33, 7.76], $d=0.87$), a main effect of condition, $F(1, 226)=60.10$, $p<.001$, $\eta_p{}^2=.21$ [.12, .30].

There was also a significant interaction between condition and photo type, $F(1, 226)=13.50$, $p<.001$, $\eta_p^2=.06$, *95% CI* [.01, .12]. Follow-up independent samples t-tests revealed that the photo test condition was significantly more confident when identifying old photos (M= 9.43, [9.25, 9.61]) compared to the description test condition (M=8.33, [8.05, 8.60]), $t(196.73)=6.60$, $p<.001$, d=0.87, [.60, 1.15]. Similarly, the photo test condition was significantly more confident when identifying new photos (M=8.38, [8.10, 8.66]) compared to the description test condition (M=6.72, [6.39, 7.05]), $t(226)=7.65$, $p<.001$, d=1.01 [.74, 1.29]. However, the difference in confidence between conditions was greatest for new photos. These results fit with our memory accuracy data, suggesting that participants found it particularly difficult when deciding whether a New photo description was old or new.

Conclusions

Taken together, test format influenced participants' ability to distinguish between old and new test items and their willingness to say a test item was old. Specifically, descriptions led to a reduction in memory accuracy and confidence, as well as an enhanced response bias towards saying "old" to test items, relative to those given a photo test. These findings have important implications: the field research on memory amplification presents participants with descriptions of traumatic stressors and this test format may encourage participants to endorse trauma exposure. Moreover, because memory amplification arises over time, we wondered, do people become *more* biased to respond yes after a longer delay, when trace memory for the original has faded even further? If so, the data would support an explanation for memory amplification that is specific to the test format used in the field data.

We therefore ran a second experiment assessing analogue PTSD symptoms with *all* participants completed the description test on *two* occasions (as in the field research), 24 hours apart. This amendment allowed us to determine if and how participants' response bias and sensitivity at test changed over time and whether these changes were related to analogue PTSD symptoms. We expected sensitivity would worsen over time, due to memory decay. We had competing predictions for response bias. On the one hand, as memory traces inevitably weaken over time, people might become more conservative at test, due to decreased confidence in their ability to correctly identify photos. We found this pattern of data previously, using only a photo test (Oulton et al., 2016). On the other hand, in the field research people tend to endorse more trauma exposure over time, perhaps due to a motivation to justify one's rising distress and reinterpretation of event descriptions, for example (Engelhard et al., 2000). Thus, we might replicate that pattern, with participants becoming even more willing to respond "old" to test items at the second test.

Experiment 2

Method

Participants

We recruited participants using Amazon's Mechanical Turk: 111 completed the experiment (16 participants (12.6%) failed to complete Session 2). We removed three participants who completed the second test more than 24 hours after receiving the test link. Our analyses therefore focus on the remaining 108 participants. Participants were US residents, aged 22-78 (M=36.19, 95% CI [33.95, 38.44]). Most (56.5%) were male and Caucasian (including White; 77.8%). Others identified as Asian American (6.5%), African American (including Black; 5.6%), Hispanic (4.6%), European (1.9%), or mixed ethnic origin (3.7%). We did not collect any other demographic or socioeconomic data.

Materials and Procedure

The experiment consisted of two separate online sessions, approximately 24 hours apart. For Session 1, the materials and procedure were identical to Experiment 1. However, all participants received the description test and participants completed an adapted version[6] of the PTSD Checklist for DSM-5 (PCL-5; Weathers et al., 2010) at the end of the survey (see also Carleton, Sikorski & Asmundson, 2010; Nixon et al., 2007). Participants rated how bothered they were by items describing PTSD symptoms *since viewing the photos* (e.g., repeated, disturbing and unwanted memories of the photos; 0=*not at all*, 4=*extremely*). We used the PCL for DSM-5 to calculate a total symptom severity score, along with the four symptom cluster severity scores: re-experiencing (Cluster B), avoidance (Cluster C), negative alterations in cognition and mood (Cluster D) and arousal (Cluster E). The test-retest reliability of the PCL-5 is high (Keane et al., 2014). The scale has excellent convergent validity with the PCL-C and other generalized anxiety disorder scales (Bovin et al., 2015).

Twenty-four hours after the first session, participants received an email with a link to Session 2. Within Session 2, participants first completed the PANAS, followed by the description recognition test. Consistent with the field research, all test items were exactly the same on both occasions (albeit presented in a random order at both time points). However, we slightly modified the wording of the test instructions for the second test to ensure participants would respond “old” only if they remembered the photo (as opposed to the photo’s description). That is, we wanted to avoid participants responding “old” to test descriptions, simply because they remembered seeing the same description in the test the day before. We asked participants to select old if “the description is of a PHOTO you saw yesterday in session 1” and select new if “the description is of a PHOTO you did not see yesterday in session 1”. After the test, participants completed the full, 20-item version of

[6] For the first administration of the PCL-5 we excluded 6 items that are meaningless for a 20 min delay period (i.e., “repeated disturbing dreams of the stressful experience”, “avoiding external reminders of the stressful experience”, “loss of interest in activities that you used to enjoy”, “feeling distant or cut off from other people”, “taking too many risks or doing things that could cause you harm” “trouble falling or staying asleep”). Therefore, the revised scale consisted of 14 items in total. Participants completed the full (20-item) version of the scale 24 hours later.

the PCL-5 in relation to the photos a second time. At the end of the study participants were paid $6.00 for their time. Our data for Experiment 2 can be found on the Open Science Framework (OSF) at https://osf.io/6u4wh.

Results

Emotional Impact of Photos

As in Experiment 1, positive affect significantly reduced (Time 1: M=29.54 [27.72, 31.35], Time 2: M=24.76 [23.27, 26.25]; $t(107)$= 8.61, p<.001, d=0.55, *95% CI* [0.40, 0.69]) and negative affect significantly increased following photo exposure [7](Time 1: M=12.05 [11.22, 12.88], Time 2: M=19.19, [17.62, 20.77]), $t(107)$=-10.89, p<.001, d=1.08 [0.84, 1.32]. Before the second test, mean positive affect was significantly lower (27.06 [25.07, 29.05]) than positive affect prior to encoding, $t(107)$=5.44, p<.001. However, the effect size was small, d=0.25 [0.15, 0.34]. This finding may reflect anticipation of viewing negative material or a carryover effect after viewing the negative stimuli. Mean negative affect before the second test (M=12.31 [11.41, 13.22]) did not significantly differ from mean negative affect before encoding, $t(107)$=-.74, p=.46. Mean PCL-5 scores (in relation to the photos) were significantly lower 24 hours after photo presentation (M=9.07, [6.94, 11.19]) compared to 20 minutes after photo presentation (M=12.97 [10.82, 15.13]), $t(106)$=4.23, p<.001, d=0.35 [0.18, 0.52].

Photo ratings

As in Experiment 1, participants found the photos very unpleasant (M=6.35, *95% CI* [6.15, 6.55]), distressing (M=5.65 [5.33, 5.96]) and disgusting (M=5.91 [5.65, 6.16]) and reported paying very close attention to the photos (M=6.77 [6.68, 6.86]). Participants reported thinking about the photos moderately often, when asked 20 minutes after photo exposure (M=2.32, 95% *CI* [2.07, 2.56], 1= *Almost never*, 5= *Extremely frequently*)

[7] Cronbach's alpha for positive affect was .92 and .87 before and after photo exposure, respectively. For negative affect, cronbach's alpha was .91 both before and after photo exposure.

Memory Accuracy[8]

Next, we examined how participants' response bias and sensitivity changed over time. A paired samples t-test revealed that participants' sensitivity significantly worsened over time, $t(107)=6.09$, $p <.001$, $d=0.46$, *95% CI* [0.30, 0.62]). As shown in Figure 2a, participants became worse at distinguishing old from new items over time. In addition, a paired sample t-test revealed that participants became significantly *more biased* to respond "old"—as measured by criterion *c*—to test items over the 24 hour period, $t(107)=6.69$, $p<.001$, $d=0.47$ [0.32, 0.63] (Figure 2b). Thus, replicating the field research, overall participants endorsed exposure to more photos at the second assessment compared to the first assessment (i.e., their memory amplified).

Previous field studies have been unable to use a signal detection approach, because they do not verify what events participants actually experienced. Therefore, we examined reporting discrepancies across time using two other approaches from the memory amplification field literature. First, following Southwick et al. (1997), we created variables indicating whether each photo was: (1) identified as "old" at both times, (2) identified as "old" initially but later identified as "new", (3) identified as "new" initially but later identified as "old", and (4) identified as "new" at both time points. Almost all participants (98.1%) changed at least one of their test responses over time; the mean number of response changes was 5.43, 95% CI [4.80, 6.05]. A total of 94.4% of participants changed at least one of their responses from "new" to "old", while 66.7% changed at least one response from "old" to "new". These percentages were slightly elevated compared to the field studies, particularly for new-old (i.e., no-yes) changes. It should be noted, however, that there is great variability across studies for both percentage of no-yes changes (e.g.,

[8] We also examined age and gender as potential moderators for the effect of time on response bias and sensitivity. Results were non-significant for all analyses, all *p*s>.05. Excluding participants who closed their eyes or looked away from the photos at encoding (16.7%) did not change the pattern of results for sensitivity or response bias.

70% (Southwick et al., 1997), 40.7% (Krinsley et al.,), 76% (King et al.) and 70% (Engelhard et al.,) and yes-no changes observed (e.g., 46% (Southwick et al., 1997), 45.1% (Krinsley et al.), 58.3% (King et al.), 80% Engelhard et al.). These discrepancies may reflect differences in how exposure items are worded, or differences in the degree of trauma exposure across samples. The mean number of New-Old changes was 3.91 [3.33, 4.48] and the mean number of Old-New changes was 1.52 [1.23, 1.81].

Second, following Giosan et al. (2009), we calculated an overall memory change score by subtracting the proportion of negative photos (both old and new) endorsed as "old" at Time 2 from the number of photos identified as "old" at Time 1. Thus, if people remembered more negative photos over time (i.e., memory amplification), their change score would be a negative value; if they remembered fewer photos at Time 2 (''forgetting'') it would be a positive value. The mean memory change score was -2.39 [-3.05, -1.73]. That is, on average participants endorsed two additional photos at Time 2. Most participants (73.1%) endorsed more photos over time, 15.7% of participants decreased the number of photos endorsed and 11.1% endorsed the same number of photos at both time points.

Analogue PTSD Symptoms and Memory Performance

Next, we examined the relationship between analogue PTSD symptoms and memory amplification. Consistent with the field research, we expected that PTSD symptoms would be positively associated with endorsing more photo exposure over time. Again, we operationalized this change in three ways. We first examined participants' bias to respond "old" to test items: we calculated a change in response bias (or memory amplification) score by subtracting ***c*** scores at T2 from ***c*** scores at T1. Positive values represented becoming more biased to respond "new" over time, and negative values represented becoming more biased to respond "old" over time (i.e., memory amplification).

Seventy-eight participants (72.2%) became more biased to respond “old” over time, 9 (8.3%) participants’ bias did not change and 21 (19.4%) participants became more biased to respond “new”. Consistent with the field research, there was a small, significant negative relationship between overall PTSD symptoms at T2[9] and change in response bias, $r = -.20$, 95% CI [-.38, -.01], $N = 107$, $p = .035$ (We also calculated the relationship for the PCL subscales: re-experiencing: $r=-.21$ [-.38, -.02], $p=.032$; avoidance: $r=-.16$ [-.34, .03], $p=.11$; cognition and mood change: $r=-.10$ [-.28, .09], $p=.33$; arousal: $r= -.25$ [-.42, -.06], $p= .01$). Specifically, the more participants experienced analogue PTSD symptoms in relation to the photos, the more biased they became to respond “old” to test items over time. Participants’ rating of how frequently they thought about the photos over the first 20 minute delay period was not significantly related to change in response bias ($r=-.04$ [-.22, .16], $p=.70$). However, these ratings were tentatively related to change in sensitivity, $r=.19$ [.00, .37], $p=.055$. This finding may reflect a rehearsal effect, such that participants who thought about the photos more were less likely to forget the photos they had previously viewed.

We next examined the relationship between analogue PTSD symptoms and response changes. Based on the field research (e.g., Engelhard et al., 2008; Southwick et al., 1997), we anticipated that the number of New to Old changes would be associated with total PCL scores. Indeed, we observed a significant positive correlation with a comparable effect size to prior field research, $r= .33$, $p<.001$ (re-experiencing: $r=.31$, $p=.001$; avoidance: $r=.17$, $p=.08$, cognition change: $r=.28$, $p=.004$; arousal: $r=.34$, $p<.001$). Further, there were no significant relationships between PTSD symptoms and the 3 other types of response change, all $ps>.05$. Thus, analogue PTSD symptoms were *uniquely associated* with later remembering photos as “old” that were previously remembered as “new”.

[9] One participant failed to complete the PCL at Time 2. Thus, their data is not included in the correlational analyses.

Finally, we also found that memory change scores (Giosan et al., 2009) were significantly correlated with analogue PTSD symptoms at time 2: that is, an increase in the number of endorsed photos at Time 2 was associated with more severe analogue PTSD symptoms at follow up, r=-.27 [-.44, -.09], N=107, p=.005. Echoing this finding, how distressing participants' rated the photos also predicted memory amplification, r = -.21 [-.38, -.02], p= .026. Interestingly however, there was no significant relationship between analogue PTSD symptoms and change in sensitivity, r=.10 [-.09, .28], p=.32. Thus, although symptoms appear to play a role in determining participants' tendency to endorse photo exposure, they do not appear to be related to participant's ability to distinguish between what they have and have not seen before. Specifically, the correlations between analogue symptoms and change in response bias (Z=2.26, p=.02), and memory change scores (Z=2.86, p<.01), but not new-to-old changes (Z=1.73, p=.08), were significantly different to the correlation between analogue symptoms and change in sensitivity (Steiger, 1980).

Memory Confidence

Finally, we analyzed how participants' confidence changed over time. Despite sensitivity significantly worsening over time, paired sample t-tests revealed that mean confidence did not significantly differ over time for either old (Time 1: M= 8.50 [8.26, 8.73], Time 2: M= 8.51 [8.29, 8.72]; t(107)=-0.15, p=.88, d=0.01, *95% CI* [-0.10, 0.12]) or new test items (Time 1: M= 7.06 [6.78, 7.34], Time 2: M= 6.94 [6.64, 7.25]; t(107)=1.23, p=.22, d=0.08 [-0.05, 0.20).

General Discussion

Our aim was to determine whether test format plays a role in memory amplification. In Experiment 1, we showed that using a verbal test format—akin to what is used in the field research—led to poorer memory sensitivity, and a greater bias to say "old", compared to a photo test. Our study is the first to show that, when participants are presented with *descriptions* of highly negative photos, they are more likely to endorse never seen photos, than when they are re-exposed to negative photos in the same modality as encoding. This finding is important, because trauma exposure assessments are typically descriptions. In Experiment 2—analyzing memory consistency according to the variety of methods that have been employed in the literature to date—we reliably demonstrated that participants were inconsistent in their responses over time, and endorsed more photo exposure at Time 2. This pattern fits with the field research on memory amplification (e.g., Engelhard et al., 2008; King et al., 2000; Southwick et al., 1997), and has the additional benefit that we were able to verify photo exposure.

The results from Experiment 1 suggest event descriptions give reign to imagination, and generally lack the perceptual information inherent to a traumatic event. Thus, when people judge whether an event description happened to them, they must determine how similar that description is to their actual experience, rather than judging a verbatim stimulus. Of course, one could argue that people sometimes incorrectly endorsed a description they thought was of a photo they actually saw, inflating our error / bias rate in this condition. However, we were careful to only use descriptions of discrete and discriminable (compared to subjective) photographs; after pilot testing, we removed any descriptions that could apply to more than one photo. Further, some of the field research *has* used subjective descriptions (e.g., "extreme threat to your personal safety", "witnessing violence") that could plausibly apply to multiple events, likely inflating their memory amplification rate (e.g., Engelhard et al., 2008; Southwick et al., 1997).

The finding that memory amplification itself may be malleable has methodological, practical and theoretical implications. For example, the size of the memory amplification effect observed in previous research may—at least partially—reflect the method by which memory is tested. Indeed, inconsistencies in people's responses to questions about their experience might not always represent distortion of memory per se but may be a feature of the test format. In particular, descriptions of trauma used in exposure checklists may be especially vulnerable to reinterpretation and response biases, and these biases may magnify over time, as we showed in Experiment 2. To illustrate, if a veteran initially says they don't remember seeing friends killed or injured and later responds that they do remember, this change could represent the creation of a new event memory—as field studies often assume—*or* it could reflect how the victim approached the test. They may be relying more on gist memory and vague impressions of similarity between survey items and their past experience, be more motivated to search their memory for what happened to them during service or be re-interpreting the meaning of the word "friend".

Practically speaking, even if reporting discrepancies do not always represent changes in a person's *memory* of what happened, they still have implications for real world trauma reporting and prognosis. For example, researchers and clinicians often consider checklists to be objective indices of true trauma exposure and PTSD to be a direct outcome of that exposure. Our findings strongly suggest that assumption is simply not accurate. Instead, our results align with a memory-based model of PTSD (Rubin, Berntsen, & Johansen, 2008) whereby PTSD symptoms derive from one's current *memory* of the negative event and not the event per se. Importantly—like all memories—the trauma memory is subject to alteration and distortion.

In Experiment 2, we found that memory amplification was consistently related to analogue PTSD symptoms; a finding that parallels the field and laboratory research. There are several possible reasons for this relationship. One is that people experiencing more severe symptomology at T2 are more motivated to recall negative photos that help them justify and make sense of these symptoms (Engelhard et al., 2008), and to reinterpret events to fit with their symptom profile. Another, related, possibility is that people with PTSD symptoms are unable or unwilling to determine exactly what they witnessed, relying on heuristic processing (i.e., rapid, non-strategic decisions) at test. For example, distressed people might frequently mistake imagined details about the trauma with what they actually witnessed, because they base their decisions on the qualitative characteristics (e.g., degree of perceptual detail) of activated memories. Similarly, people with PTSD may retrieve less specific-representations, and rely more on the gist of what they experience (Engelhard et al., 2008). Hence, vague impressions of similarity between checklist items and their experience could result in false recognition and a response bias toward endorsing trauma. In the context of our paradigm, participants with many intrusive memories about the photos may have been less inclined to search their memory for specific representations of the photos shown (to avoid the negative feelings/symptoms associated with thinking about the images) and, therefore, responded "old" to any photos that were similar to the general gist of the photos shown at encoding. Indeed, we found the re-experiencing symptoms of PTSD were most strongly associated with a change in bias toward responding "old" over time, supporting this proposal. Of course, it is also possible a third variable—a liberal response threshold or general tendency for over-reporting—better explains the relationship between PTSD symptoms and memory amplification. Albeit a challenging task, future research might test this possibility by including a third variable that assesses an emotional response that is unrelated to our key variables.

Taken together then, our findings suggest that test format contributes to memory amplification, but the strength of this effect may *depend* on the presence and strength of PTSD symptoms. In particular, people with more severe PTSD symptoms may be more likely to show a pronounced amplification effect when test format is verbal—due, for example, to a need to justify distress or a reliance on heuristic processing—compared to participants who are relatively asymptomatic and less likely to exhibit these tendencies.

Of course, our experiments have limitations. Using a trauma analogue means that our conclusions may lack generalizability. Although participants rated the photos as very unpleasant and distressing, viewing negative photos clearly does not provoke the same degree of fear as a real trauma. Relatedly, unlike traumatic events in the field, there was obviously no threat to life or safety. Further, unfortunately our data do not allow us to determine the direction of the relationship between PTSD and amplification; we cannot determine whether people develop symptoms in response to an amplified memory, or whether memory amplification causes more symptoms. Finally, because we did not use a remember/know judgment, or directly assess whether participants were using gist-based retrieval, we cannot infer whether participants were using recollection or familiarity (or gist) memory, to guide their decisions. Thus, we are limited in the conclusions we can draw about the precise mechanism driving the effects of test format we observed.

Considering our key findings, and limitations in the inferences that we can make from these findings, we would suggest several areas for future research attention. First, we think it important to directly assess the specific strategies participants adopt when endorsing (or not) trauma events. Thus, future research using an analogue paradigm to investigate memory amplification could use remember-know judgments to examine whether participants respond “old” because they genuinely remember an image or because it seems familiar or matches the general gist of what they have seen before. It may also be important to add an incentive for recognition accuracy, or examine reaction time, to evaluate whether motivation or effort play a role in memory amplification. Second, we need to know more about the characteristics or processes that could increase the familiarity of not-experienced items. For example, a measure of imagination—about the photos themselves and/or participants' trait ability and tendency to imagine or fantasize—would help us to determine whether individual differences in degree of imagination explain the effects of test format we observed. Also, analyzing the content of any intrusive thoughts participants experience over the delay period may help to elucidate whether people imagine information beyond the content of photos. Third, because our experiments do not allow us to separate out a response bias determined by some other less relevant cause, future research could 'simulate' the memory amplification effect by modifying participants' report criteria for responding old or new—such as by varying the proportion of old items on the test to increase the likelihood participants respond old (e.g., Rhodes & Jacoby, 2007). This method would allow us to test whether other factors—such as whether someone is currently distressed, or has a PTSD diagnosis—increase susceptibility to such a response bias manipulation. Finally, as a methodological point, including comparison positive and/or neutral stimuli would determine whether our findings are specific to stimuli that elicit negative arousal or whether they reflect a general memory phenomenon.

Taken together, our data show that the test format typically adopted in memory amplification research is vulnerable to response bias and memory errors. Critically, these biases and errors *enhance* as memory traces weaken over time and people with PTSD symptoms are particularly susceptible to this enhancement. Thus, our findings are the first to suggest that test format plays a role in memory amplification.

References

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, *78*(2), 490-8.

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2015). Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in Veterans. *Psychological Assessment. Advance online publication.* doi: 10.1037/pas0000254

Carleton, R. N., Sikorski, J., & Asmundson, G. J. (2010). Terrifying movie stimuli: A new design for investigating precursors for posttraumatic stress. *Psychological Trauma: Theory, Research, Practice, and Policy, 2,* 206-217. doi: 10.1037/a0020100

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Engelhard, I. M., & McNally, R. J. (2015). Metacognitive appraisal of memory inconsistency for traumatic events in Dutch veterans. *Memory*, *23*, 972-980. doi: 10.1080/09658211.2014.942669

Engelhard, I. M., van den Hout, M. A., & McNally, R. J. (2008) Memory consistency for traumatic events in Dutch soldiers deployed to Iraq. *Memory, 16,* 3-9. doi: 10.1080/09658210701334022

Frueh, B. C., Elhai, J. D., Grubaugh, A. L., Monnier, J., Kashdan, T. B., Sauvageot, J. A., Hamner, M. B., Burkett, B.G., & Arana, G. W. (2005). Documented combat exposure of US veterans seeking treatment for combat-related post-traumatic stress disorder. *The British Journal of Psychiatry, 186*(6), 467-472. doi: 10.1192/bjp.186.6.467

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*, 208-214. doi: 10.3758/BF03212420

Garry, M., & Wade, K. A. (2005). Actually, a picture is worth less than 45 words: Narratives produce more false memories than photographs do. *Psychonomic bulletin & review*, *12*, 359-366. doi: 10.3758/BF03196385

Giosan, C., Malta, L., Jayasinghe, N., Spielman, L., & Difede, J. (2009). Relationships between memory inconsistency for traumatic events following 9/11 and PTSD in disaster restoration workers. *Journal of Anxiety Disorders, 23,* 557–561. doi: 10.1016/j.janxdis.2008.11.004

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577-581. doi: 10.3758/BF03214352

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. doi: 0.1037/0033-2909.114.1.3

Keane, T. M., Rubin, A., Lachowicz, M., Brief, D., Enggasser, J. L., Roy, M., Hermos, J., Helmuth, E., & Rosenbloom, D. (2014). Temporal stability of DSM-5 posttraumatic stress disorder criteria in a problem-drinking sample. *Psychological Assessment, 26,* 1138-1145. doi: 10.1037/a0037133

King, D. W., King, L. A., Erickson, D. J., Huang, M. T., Sharkansky, E. J., & Wolfe, J. (2000). Posttraumatic Stress Disorder and retrospectively reported stressor exposure: A longitudinal prediction model. *Journal of Abnormal Psychology, 109,* 624-633. doi: 10.1037/0021-843X.109.4.624

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*, 555-583. doi: 10.1006/jmla.1997.2529

Lang, P.J., Bradley, M. M., & Cuthbert, B.N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report No. A-8.* Florida: University of Florida Gainesville.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review*, *87*, 252-271. doi: 10.1037/0033-295X.87.3.252

Nixon, R. D. V., Nehmy, T., & Seymour, M. (2007). The effect of cognitive load and hyperarousal on negative intrusive memories. *Behaviour Research and Therapy, 45,* 2652-2663. doi: 10.1016/j.brat.2007.06.010

Oulton, J. M., Takarangi, M. K. T., & Strange, D. (2016). Memory amplification for trauma: Investigating the role of analogue PTSD symptoms in the laboratory. *Journal of anxiety disorders*, *42,* 60-70. doi: 10.1016/j.janxdis.2016.06.001

Roemer, L., Litz, B. T., Orsillo, S. M., Ehlich, P. J., & Friedman, M. J. (1998). Increases in retrospective accounts of war-zone exposure over time: The role of PTSD symptom severity. *Journal of Traumatic Stress, 11,* 597-605. doi: 10.1016/S0005-7967(98)00013-8

Rubin, D. C., Bernsten, D., & Bohni, M. K. (2008). A memory-based model of posttraumatic stress disorder: evaluating basic assumptions underlying the PTSD diagnosis. *Psychological Review, 115,* 985-1011. doi: 10.1037/a0013397

Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: conceptual and perceptual false recognition. *Neuropsychology*, *11*, 331. doi: 10.1037/0894-4105.11.3.331

Scoboria, A., Wade, K. A., Lindsay, D. S., Azad, T., Strange, D., Ost, J., & Hyman, I. E. (2017). A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies. *Memory, 25*, 146-163. doi: 10.1080/09658211.2016.1260747

Southwick, S. M., Morgan, C. A., Nicolaou, A. L., & Charney, D. S. (1997). Consistency of memory for combat-related traumatic events in veterans of Operation Desert Storm. *American Journal of Psychiatry, 154*, 173-177.doi: 10.1176/ajp.154.2.173

Stainslaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behaviour Research Methods, Instruments, & Computers, 31*, 137-149. doi: 10.3758/BF03207704

Steiger, J. H. (1980). Tests for comparing elements of a correlational matrix. *Psychological Bulletin, 87*, 245-251. doi: 10.1037/0033-2909.87.2.245

Vornik, L., Sharman, S., & Garry, M. (2003). The power of the spoken word: Sociolinguistic cues influence the misinformation effect. *Memory, 11*, 101-109. doi: 10.1080/741938170

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scale. *Journal of Personality and Social Psychology, 54,* 1063-1070. doi: 10.1037/0022-3514.54.6.1063

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2010). The PTSD Checklist for DSM–5 (PCL-5). Scale available from the National Center for PTSD website: http://www.ptsd.va.gov

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology*, *91*(2), 189-202. doi: 10.1348/000712600161781

Authors' Contribution

MKTT developed the study design. JMO collected and analyzed the data. MKTT and JMO drafted the manuscript, with critical revisions from DS. All authors approved submission of the final version.

Acknowledgements

This research was supported by the Australian Research Council (Grant DP140102661).

a) b)

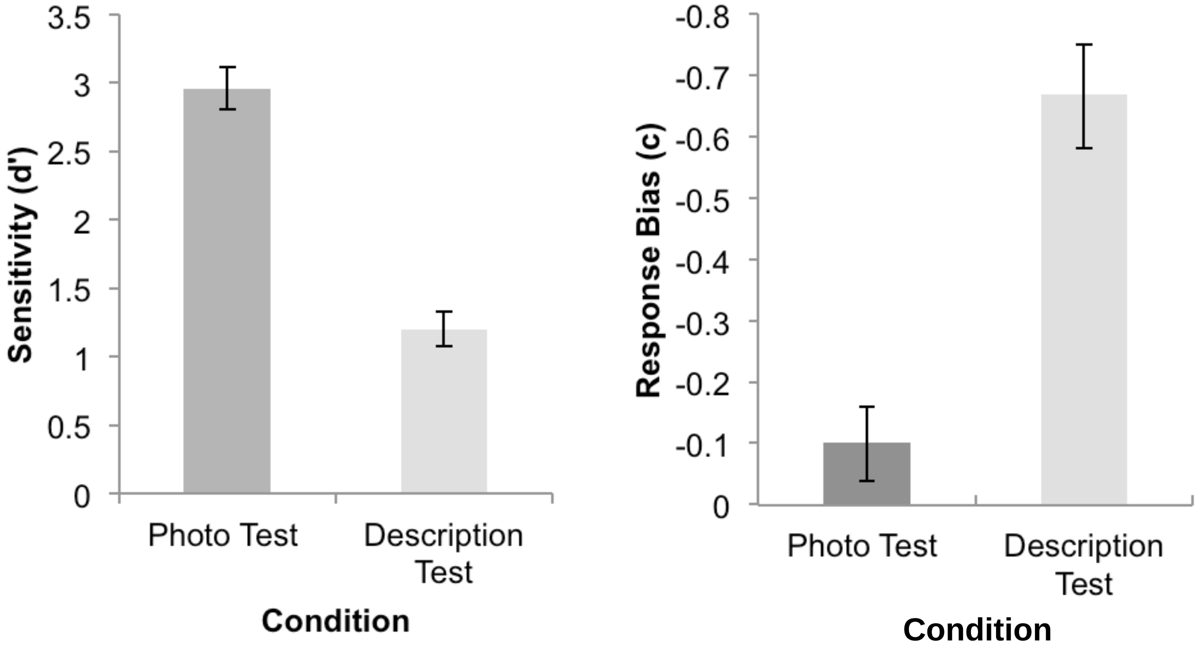

Figure 1. Mean sensitivity (a) and mean response bias (b) by condition. Error bars represent 95% confidence intervals. Increasing negative response bias values represent an increasing bias towards responding “OLD” to test items.

a) b)

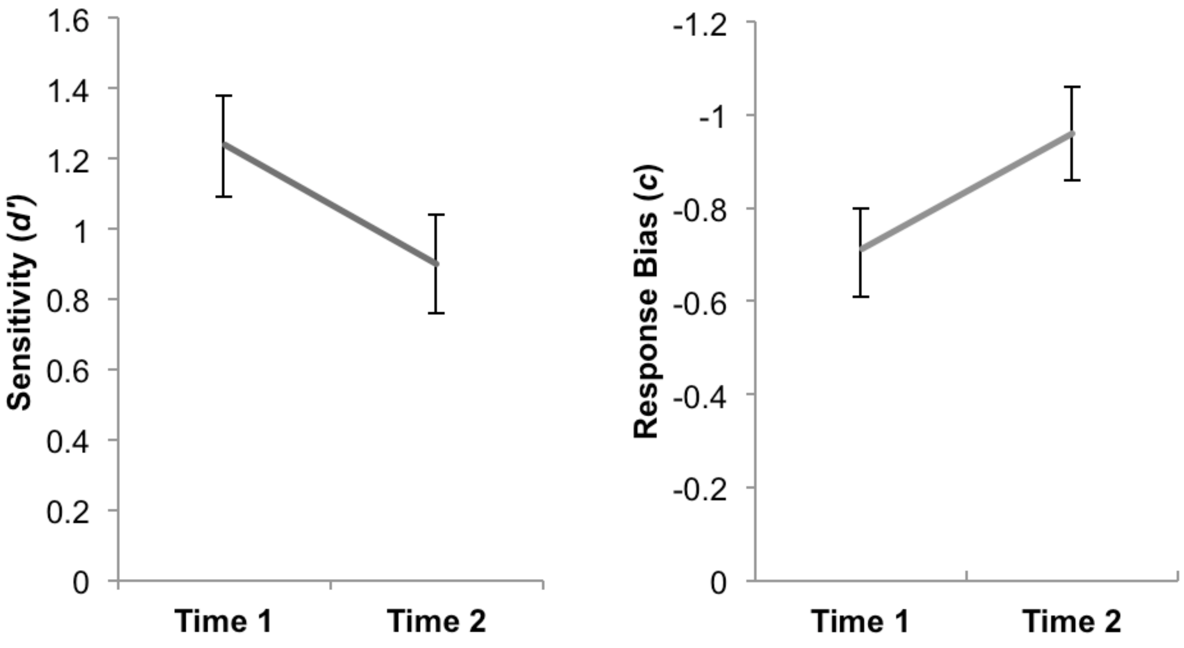

Figure 2. Mean sensitivity (a) and mean response bias (b) at Time 1and Time 2. Error bars represent 95% confidence intervals.