*Brief Empirical Report*

# Lateral Eye Movements Do Not Increase False-Memory Rates: A Failed Direct-Replication Study

## Kevin van Schie[1] and Arne Leer[2]

[1]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, and
[2]Department of Clinical Psychology, Utrecht University

## Abstract

In this direct replication of Houben, Otgaar, Roelofs, and Merckelbach (*Clinical Psychological Science, 6*, 610–616, 2018), we tested whether making eye movements during memory recall increases susceptibility to creating false memories. Undergraduates ($N = 206$) watched a video of a car crash, after which they recalled the video with or without simultaneously making eye movements. Next, participants received misinformation about the video. Finally, during the critical test, they were questioned about video details. The results showed that making eye movements did not increase endorsement of misinformation (i.e., false memory), nor did it reduce (correct) memory details or memory vividness and emotionality. Random variation in sampling or measurement, low reliability of the test instrument, and observer-expectancy effects may explain discrepancies between study effects. Only multiple direct replications by different (independent) laboratories with standardized instruments will allow for assessing whether the effect is robust and largely independent of random variation and moderators.

Eye-movement desensitization and reprocessing (EMDR) is an evidence-based and first-choice treatment of post-traumatic stress disorder (PTSD) that is widely applied in Western countries (e.g., American Psychological Association, or APA, 2017; Bisson, Roberts, Andrew, Cooper, & Lewis, 2013). Still, a substantial number of patients do not show clinical improvement after EMDR treatment (e.g., Bradley, Greene, Russ, Dutra, & Westen, 2005). Not surprisingly, therefore, recent years have witnessed an increase in research on EMDR's mechanisms of change in an ultimate attempt to optimize the treatment (see van den Hout & Engelhard, 2012).

Specifically, much attention has been focused on unraveling the mechanism underlying EMDR's core element: Patients recall the hotspot of their traumatic memory while they make horizontal eye movements (EMs; Shapiro, 2017). In laboratory studies using an analogue of this procedure, participants typically report that the EM intervention renders their memories less vivid and less emotional (van den Hout & Engelhard, 2012). A meta-analysis showed that EMs are an effective and essential component of reductions in vividness and emotionality in analogue studies and (primarily) of reductions in subjective units of distress in full-protocol studies (Lee & Cuijpers, 2013).

Until recently, however, little was known about *how* EMs work. Analogue studies have consistently demonstrated that the EMs tax our finite working memory resources, which results in suboptimal memory recall, and that such taxation is causally involved in the reductions in memory vividness and emotionality reported immediately after the intervention (e.g., van den Hout, Muris, Salemink, & Kindt, 2001; van Veen, Engelhard, & van den Hout, 2016). In line with this working memory

**Corresponding Author:**
Kevin van Schie, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
E-mail: k.vanschie@essb.eur.nl

account, all kinds of tasks that are substantially taxing cause reductions in memory vividness and emotionality for negative and positive memories—for example, playing the computer game Tetris (Engelhard, van Uijen, & van den Hout, 2010), drawing complex figures (Gunter & Bodner, 2008), or doing mental arithmetic (Engelhard, van den Hout, & Smeets, 2011)—and fast EMs that are significantly taxing affect memory phenomenology to a larger extent than slow EMs, which are only minimally taxing (Maxfield, Melnyk, & Hayman, 2008; van Schie, van Veen, Engelhard, Klugkist, & van den Hout, 2016; van Veen et al., 2015).

It has further been shown that immediate changes in memory phenomenology are maintained over time: 24 hr (Leer, Engelhard, & van den Hout, 2014) and 1 week after the intervention (Gunter & Bodner, 2008). In addition, two studies have demonstrated that the EMs reduce memory accessibility, as evidenced by increased response latencies in a stimulus discrimination task (Leer et al., 2017; van den Hout, Bartelski, & Engelhard, 2013). Presumably, these changes reflect that the EMs result in the formation of a new memory trace or affect the original memory trace through memory reconsolidation (van den Hout & Engelhard, 2012).

Such changes in memory may be desirable from a therapeutic point of view, but may come with unwanted side effects. Houben and colleagues (2018) hypothesized that the EM intervention results in increased susceptibility to accepting misinformation. They reasoned that the formation of false memories might arise when memory becomes vague (i.e., less detailed) as a result of the EM intervention and that people then become reliant on memory for central (gist) elements (see fuzzy trace theory; Brainerd, Reyna, & Ceci, 2008). Given that the EMs indeed fuel the formation of false memories, they may have significant consequences for the treatment of patients with PTSD using EMDR. Through suggestive pressure and EMs, patients may more easily recover false memories of trauma (Patihis & Pendergrast, 2018), which consequently may affect the credibility of (eyewitness) testimony and in extreme cases could result in innocent people being falsely accused. Indeed, EMDR specifically has been linked to potential false-memory cases (Patihis & Pendergrast, 2018; Shaw & Vredeveldt, 2019). In a broader sense, this is relevant to the ongoing discussion about the reliability of repressed memories (i.e., the "memory wars"; Patihis, Ho, Tingen, Lilienfeld, & Loftus, 2014).

Houben et al. (2018) showed 82 undergraduates a video of a car crash, after which the participants recalled what they saw either with or without simultaneous horizontal EMs. Subsequently, participants were provided with misinformation about the video in the form of an eyewitness narrative. Afterward, participants were provided with forced-choice interview questions about video details. Consistent with the hypotheses, EMs simultaneous with memory recall resulted in (a) reduced recall of actual video details and (b) increased recall of incorrect details that were provided in the eyewitness narrative as misinformation. However, there was no evidence that EMs decreased vividness and/or emotionality more than did keeping eyes stationary.

Given the potential impact of the findings, the current study's goal was to perform a direct replication of Houben et al. (2018). Thus far, there has been no strong tradition of (direct) replication research in social and behavioral sciences (e.g., Koole & Lakens, 2012; Makel, Plucker, & Hegarty, 2012), though recently, substantial efforts have been made (e.g., Open Science Collaboration, 2015). Replication ensures the self-correcting nature of psychological science, which is vital to scientific progress (e.g., Asendorpf et al., 2013; Nosek, Spies, & Motyl, 2012). It increases the reliability and generalizability of research findings and helps to determine a fairer estimation of the effect size, because reported effect sizes are about twice as large as unreported effect sizes (Franco, Malhotra, & Simonovits, 2016).

Because it was a direct replication, in the current study we used the same design, stimuli, and procedures as Houben et al. (2018), taken from the Open Science Framework. The sample size was set at 2.5 times the original sample (as recommended by Simonsohn, 2015) because sample estimations based on published (inflated) effect sizes may be misleading (Simonsohn, Nelson, & Simmons, 2014). We hypothesized that making EMs simultaneously with memory recall (a) reduces vividness and/or emotionality of the memory, (b) reduces recall of memory details, and (c) increases endorsement of misinformation.

## Method

### *Participants*

An ethnically diverse sample of first-year (62.6%), second-year (31.1%), and third-year (6.3%) undergraduate psychology students ($N$ = 206) from Erasmus University Rotterdam participated for course credit (162 women, 37 men; mean age = 21.06 years, $SD$ = 3.84, range = 17–55 years). Participants were equally allocated to one of two conditions on the basis of order of appearance. The ethical committee of the Faculty of Psychology and Neuroscience, Maastricht University, approved the original experiment and this approval was sufficient for conducting this direct replication.

### *Materials*

***Beck Depression Inventory–II.*** The Beck Depression Inventory–II (BDI-II; Beck, Steer, & Brown, 1996) is a reliable and valid self-report questionnaire that assesses

depressive mood via 21 items (Wang & Gorenstein, 2013). Each item measures a symptom that indicates depression and consists of four statements; one statement must be selected. All items are scored on a scale from 0 to 3, and higher total scores (range = 0–63) are indicative of higher severity of depressive mood.

**Video.** The video is a 3-min, 34-s graphic public-service announcement about the dangers of texting while driving (Strange & Takarangi, 2012). It depicts three female teenagers chatting in a car, one of whom is texting and driving. As a result, the driver crosses the centerline and collides with an oncoming vehicle. When the cars have come to a full stop, a third car crashes into them. The video depicts at least five fatalities, including a baby. The video continues to show the aftermath of the accident, including emergency vehicles and an air ambulance arriving. It ends with a close-up of the driver's face before she is transported to the hospital.

**Eye-movement task.** A gray dot was presented on a black background. In the recall + EM condition, the dot moved horizontally with a speed of 1 Hz (1 left-right-left movement per second) during four 24-s intervals separated by 10-s breaks. The recall + eyes stationary (ES) condition was identical to the recall + EM condition, except that the dot remained stationary in the middle of the screen. We used E-Prime software (Version 2.0; E-Prime, Psychology Software Tools, Pittsburgh, PA).

**Filler task.** An online, nontimed version of the tile-matching game *Bejeweled* was used as a filler task. The game's goal is to score as many points as possible by swapping adjacent gems in a large field of differently colored gems to form horizontal or vertical chains of three or more gems of the same color. Formed chains disappear and gems fall from the top to fill in gaps in the field.

**Misinformation narrative.** Misinformation was provided in the form of a printed eyewitness narrative containing 10 true statements and 5 false statements (e.g., "The girl who was driving was texting with a boy called *John*" instead of ". . . *James*"). The 15 statements corresponded with the questions in the recognition test.

**Recognition test.** The recognition test contained 15 questions, each with two answer options (e.g., "To whom were the girls writing a text message? John/James"). All questions consisted of a true answer and a foil. For five questions, the foil contained false information (presented in the eyewitness narrative). For the remaining 10, the foil contained information that was not presented before and was thus incorrect. The experimenter administered the recognition test orally and the participant responded orally (Parker, Buckley, & Dagnall, 2009).

## Procedure

Throughout the experiment, all communication between experimenter and participants was in English. After giving informed consent, participants reported demographics (age, sex, gender, year of study, and ethnicity) and completed the BDI-II. Next, they watched the video while sound was played at a moderate volume via headphones. Directly after the video, participants rated vividness and emotionality of the observed event on a 100-mm pen-and-paper visual analog scale (VAS) that ranged from 0 (*not vivid at all/extremely negative*) to 10 (*extremely vivid/extremely positive*). All participants then performed the eye-movement task (assigned to either recall + EM or recall + ES) while sitting approximately 30 cm from the computer screen. Participants in the recall + EM condition moved their eyes horizontally by tracking the moving dot on the screen while thinking of the content of the video and any associated emotions. Participants in the recall + ES condition followed identical instructions but focused on a stationary dot. Next, vividness and emotionality VASs were again completed. Subsequently, participants played the filler task for 5 min. They then read the eyewitness narrative (with misinformation). Out of the participant's sight, the experimenter timed how long it took the participant to read the narrative. Then, participants played the filler task again for 5 min, followed by the recognition test, debriefing, and compensation.

## Results

### Randomization and manipulation checks

Using null-hypothesis significance testing (NHST) in JASP (Version 0.9.1; https://jasp-stats.org/), there was no evidence for differences in age, $t(204) = -0.72$, $p = .472$, $d = -0.100$, 95% confidence intervals (CI) = [−0.374, 0.173], sex/gender ratios, $\chi^2(2)s = 1.77$, $ps = .413$, or BDI-II scores, $t(204) = 0.74$, $p = .461$, $d = 0.103$, 95% CI = [−0.170, 0.376]. Participants took sufficient time reading the eyewitness narrative ($M = 98.55$, $SD = 27.10$, range = 53.00–281.00 s), $t(204) = 0.55$, $p = .582$, $d = 0.077$, 95% CI = [−0.196, 0.350].

### Vividness and emotionality

For vividness, there was a main effect of time, which showed a drop in ratings from preintervention ($M = 7.34$, 95% CI = [7.12, 7.56]) to postintervention ($M = 6.24$, 95% CI = [5.97, 6.50]), $F(1, 204) = 76.61$, $p < .001$, $\eta_p^2 = .273$ (see Table 1). There was no effect of Condition, $F(1, 204) = 0.10$, $p = .748$, $\eta_p^2 = .001$, nor was there an interaction effect, $F(1, 204) = 0.99$, $p = .321$, $\eta_p^2 = .005$.

**Table 1.** Vividness and Emotionality Scores Before and After the Intervention and Recognition Test Results

| Measure | Condition | |
| --- | --- | --- |
| | Recall + EM ($n$ = 103) | Recall + ES ($n$ = 103) |
| Vividness rating | | |
| Preintervention | 7.44 [7.12, 7.76] | 7.24 [6.94, 7.54] |
| Postintervention | 6.21 [5.81, 6.61] | 6.27 [5.92, 6.61] |
| Emotionality rating | | |
| Preintervention | 2.60 [2.24, 2.97] | 2.25 [1.93, 2.57] |
| Postintervention | 3.54 [3.21, 3.88] | 2.82 [2.48, 3.16] |
| Recognition test | | |
| Misinformation answers | 0.90 [0.73, 1.08] | 0.85 [0.67, 1.02] |
| Correct answers | 13.20 [12.95, 13.45] | 13.42 [13.18, 13.66] |

Note: Values are means with 95% confidence intervals (CI) in square brackets.

For emotionality, there was a main effect of time; compared with ratings before the intervention ($M$ = 2.43, 95% CI = [2.18, 2.67]), those after the intervention ($M$ = 3.18, 95% CI = [2.94, 3.42]) were less negative, $F(1, 204)$ = 66.23, $p$ < .001, $\eta_p^2$ = .245. There was also an effect of condition; recall + EM scores ($M$ = 3.07, 95% CI = [2.75, 3.39]) were less negative than recall + ES scores ($M$ = 2.53, 95% CI = [2.23, 2.84]), $F(1, 204)$ = 5.67, $p$ = .018, $\eta_p^2$ = .027. The Time × Condition interaction approached significance, $F(1, 204)$ = 3.88, $p$ = .050, $\eta_p^2$ = .019.

### *Correct answers and misinformation answers*

Contrary to our hypothesis, there was no evidence that participants in the recall + EM condition more readily endorsed misinformation answers ($M$ = 0.90, 95% CI = [0.73, 1.08]) than participants in the recall + ES condition ($M$ = 0.85, 95% CI = [0.67, 1.02]), $t(204)$ = 0.46, $p$ = .650, $d$ = 0.063, 95% CI = [–0.210, 0.336]. Moreover, there was no evidence that participants in the recall + EM condition ($M$ = 13.20, 95% CI = [12.95, 13.45]) had fewer correct answers than participants in the recall + ES condition ($M$ = 13.42, 95% CI = [13.18, 13.66]), $t(204)$ = 1.19, $p$ = .235, $d$ = −0.166, 95% CI = [–0.439, 0.108]. Similar to the results of the original study, the main effects of endorsed misinformation answers, $F(1, 202)$ = 0.59, $p$ = .443, $\eta_p^2$ = .003, and correct answers, $F(1, 202)$ = 0.37, $p$ = .544, $\eta_p^2$ = .002, were not qualified by BDI scores.

Like Houben et al. (2018), we also analyzed our data on the endorsement of misinformation using Bayesian hypothesis testing (BHT) in JASP. We performed a sequential analysis with a robustness check, which shows how the evidential strength develops from when data are collected until the intended sample size is reached. An advantage of BHT over NHST is that it also allows for quantifying evidence in favor of $H_0$ by means

of the Bayes factor (BF; Dienes, 2016). The BF expresses the likelihood of the data under $H_0$ relative to $H_1$ (or vice versa). For instance, $BF_{01}$ = 5 means that the data are five times more probable under the null hypothesis compared with the alternative hypothesis. The black line (JASP default prior of $r$ = 1/√2) in Figure 1 illustrates the evidential trajectory for endorsement of misinformation answers in favor of $H_0$ over $H_1$ and shows that the evidence for $H_0$ increases with the number of data points until the intended sample size is reached, $BF_{01}$ = 5.98. Moreover, a robustness check with wide (dotted) and ultrawide (gray) priors shows no difference in the interpretation of the evidential trajectory. A comparable analysis for correct answers revealed an evidential trajectory that is somewhat capricious in nature, but eventually the evidence is in favor of $H_0$ (black line) when the final sample size is reached, $BF_{01}$ = 3.39, regardless of other priors used (see Fig. 2).

### *Correlation between memory change and endorsement of misinformation*

Although the working mechanism hypothesized to be crucial for endorsement of misinformation—decreases in vividness and/or emotionality—was not assessed in Houben et al. (2018), we did test it. There was no evidence for a correlation between endorsing misinformation and decreases in vividness, $r$ = .04, $p$ = .572, or emotionality, $r$ = −.01, $p$ = .854. Moreover, Bayesian equivalents of these procedures showed strong evidence in favor of the null hypothesis for vividness, $BF_{01}$ = 9.79, and emotionality, $BF_{01}$ = 11.28.

### Discussion

The aim of this direct-replication study was to test whether making EMs during memory recall increases a
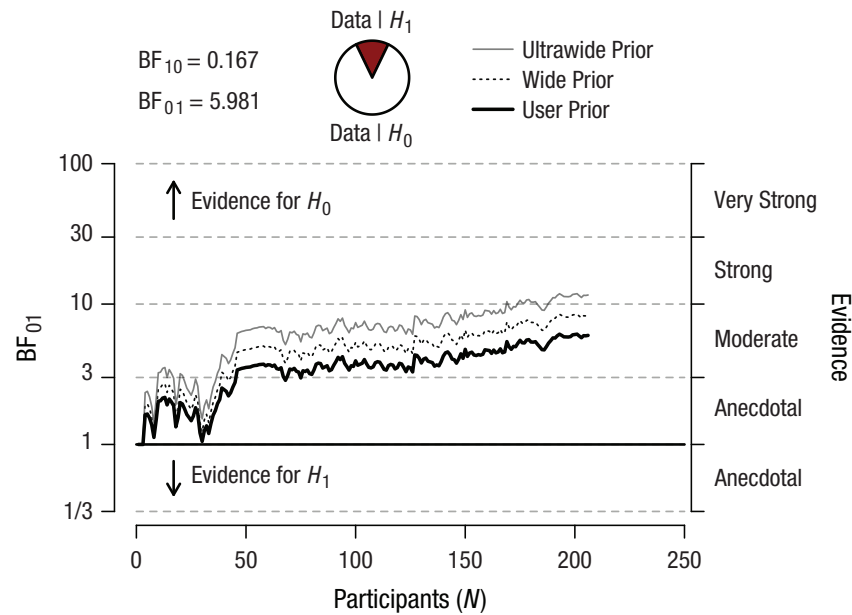
**Fig. 1.** The evidential trajectory for sequential analysis of endorsed misinformation answers with robustness checks for different priors (user, wide, and ultrawide). The graph shows the value of $BF_{01}$ as a function of the number of participants in the sample for three Cauchy priors ($r$) of different widths. For the wide prior, $r = 1$, and for the ultrawide prior, $r = \sqrt{2}$; the user prior corresponds to the JASP default value, which in this case is $r = 1/\sqrt{2}$. The arrows indicate ranges in which the $BF_{01}$ indicates there is evidence for $H_0$ (i.e., values > 1) and in which the $BF_{01}$ indicates there is evidence for $H_1$ (i.e., values < 1). The pie chart shows the relative likelihood of the data under each of the hypotheses. $BF_{01}$ = Bayes factor in favor of the null hypothesis ($H_0$) over the alternative hypothesis ($H_1$). $BF_{10}$ = Bayes factor in favor of $H_1$ over $H_0$.

person's susceptibility to endorsing misinformation. This is a question of significance, because this is a key element of EMDR, a treatment often used for PTSD (e.g., APA, 2017). Thus, a therapeutically beneficial intervention may have adverse effects. However, we found that memory recall with simultaneous EM (vs. eyes stationary) did not increase endorsement of misinformation and thus false memory. Moreover, it also did not reduce (correct) memory details or self-reported memory vividness and emotionality.

Evidently, our results contradict Houben et al.'s (2018) findings. Making EMs simultaneously with memory recall did not result in higher false-memory rates compared with keeping eyes stationary. Conceivably, this is the result of our finding that EMs did not reduce vividness and emotionality more so than merely recalling the memory. Although differential changes were numerically slightly larger in Houben et al., they reported no statistically significant differences between the conditions, either. Absence of differential decreases in vividness and/or emotionality is not uncommon, but might be more related to studies testing memory for novel materials (e.g., pictures) than to studies testing memory for autobiographical events (e.g., Leer, Engelhard, Dibbets, & van den Hout, 2013; van Schie,

Engelhard, & van den Hout, 2015; but see Leer et al., 2017). Moreover, experiential evidence relates the size of the decrease to the duration of the intervention (e.g., Leer et al., 2014), but novel (yet unconsolidated) materials might benefit from shorter instead of longer EM interventions (Leer et al., 2017). However, Houben et al. hypothesized specifically that when a memory is less vivid, a person would be more prone to accepting misinformation. No such relationship was present in the current study or even in the original study,[1] which casts doubt on whether this truly is the mechanism of action, provided that endorsement of misinformation is affected by the EMs at all.

Another reason for the discrepancy in results may simply be random variation in sampling or measurement. For instance, the original study's between-subjects design and relatively small sample size (compared with that of the current study) may have unintentionally contributed to spuriously large differences between the two conditions (i.e., creating a false-positive result). Moreover, the reliability of the forced-choice interview questions is problematic; the Kuder–Richardson 20 internal consistency score in the current study was .31 (95% CI = [.15, .45]), whereas .7 or .8 is deemed "acceptable" for general research purposes (Henson, 2001).
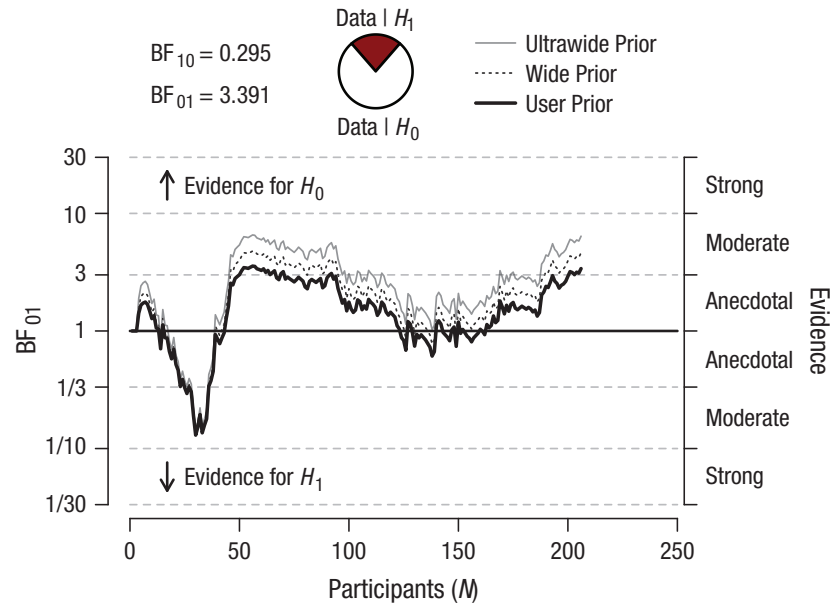
**Fig. 2.** The evidential trajectory for sequential analysis of correct answers with robustness checks for different priors (user, wide, and ultrawide). The graph shows the value of $BF_{01}$ as a function of the number of participants in the sample for three Cauchy priors $(r)$ of different widths. For the wide prior, $r = 1$, and for the ultrawide prior, $r = \sqrt{2}$; the user prior corresponds to the JASP default value, which in this case is $r = 1/\sqrt{2}$. The arrows indicate ranges in which the $BF_{01}$ indicates there is evidence for $H_0$ (i.e., values > 1) and in which the $BF_{01}$ indicates there is evidence for $H_1$ (i.e., values < 1). The pie chart shows the relative likelihood of the data under each of the hypotheses. $BF_{01}$ = Bayes factor in favor of the null hypothesis ($H_0$) over the alternative hypothesis ($H_1$). $BF_{10}$ = Bayes factor in favor of $H_1$ over $H_0$.

Low reliability of the test instrument might result in floor effects for endorsed misinformation and consequently may have compromised finding a difference between conditions in the current sample.

It is also possible that the current study represents a false-negative finding. Given the tenets of NHST, it is inevitable that a number of replication attempts are bound to be unsuccessful simply as a result of chance (Fisher, 2006; Neyman & Pearson, 1928). Thus, ruling out a false negative with absolute certainty is impossible. Assuming that the current study is not a false negative and represents a true effect, a crucial question is: To what extent do these nonsignificant replication results actually support the null hypothesis? Using NHST, such a claim is difficult to substantiate (but see Schuirmann, 1987), but the additional BHT showed that there is indeed evidence in favor of the null hypothesis that increased with more collected data, especially for endorsement of misinformation.

What else may then explain the difference in outcomes between the original study and the present study? Observer-expectancy effects may have played a role in both studies (Rosenthal, 1966). Investigators in either study may have subconsciously influenced the

participants in the experiment, causing them to respond in correspondence with their expectations. Although this is only one of the many biases that could have influenced the results (see Sackett, 1979), there are ways to circumvent or minimize the effects of biases (in future experiments); for example, by using double-blind testing, by using methods that do not rely on external pressure but rely more on automatic processes such as the Deese/Roediger-McDermott false-memory paradigm (Deese, 1959; Roediger & McDermott, 1995), or by standardizing procedures via computerized data-acquisition techniques. The recognition test that was presented orally by the experimenter seems an especially ideal candidate for standardization in order to keep risk for bias at bay.

Alternatively, many different moderators could be argued to explain the discrepancies between studies—for example, motivational differences between participants, differences in sample composition between the original and current study (e.g., more female participants, a wider age range), or use of the revised BDI. However, a list of such moderators would be infinite, and thus any claims about moderation in the current study would be largely speculative at this point. Moreover, even perfectly

matched populations may produce contradictory results that one may theoretically attribute to a moderator. The only way to assess whether an effect is robust (and largely independent of biases or moderators) is to have multiple (independent) laboratories perform the same direct replication (Simons, 2014).

In conclusion, the current experiment was a first direct replication of Houben et al. (2018), but failed to find the original study's effect that making EMs during memory recall increases a person's susceptibility to endorsing misinformation. This suggests that the original study may be a false positive and that treatment of PTSD via EMDR does not come with the adverse effect of increased false memory formation. This does not mean that EMDR is free of memory distortions per se. After all, recalling a memory (as is done in EMDR) is inherently a reconstructive process, and distortions in the form of misinformation can slip in at any time (Loftus, 2005). At this point, one study showed the presence of an effect of EMs on false memory and one study showed the absence of such an effect. Reliably and validly investigating whether (and how) making EMs during recall robustly increases a person's susceptibility to false memory formation will require a multilab, direct-replication attempt and the use of standardized test instruments.

## Action Editor

Scott O. Lilienfeld served as action editor for this article.

## Author Contributions

K. van Schie and A. Leer developed the study concept. K. van Schie was responsible for data collection and data analysis. K. van Schie and A. Leer interpreted the data, drafted the manuscript, and approved the final version of the article.

## Acknowledgments

## Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

## Open Practices

All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/j479p/ and https://osf.io/4UBJ8/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/2167702619859335. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at https://www.psychologicalscience.org/publications/badges.

## Note

1. The data from Houben et al. (2018) are available from the Open Science Framework. Post hoc correlations show no relationship between endorsement of misinformation and drops in vividness, $r = -.08$, $p = .487$; $BF_{01} = 5.72$, or emotionality, $r = -.10$, $p = .351$; $BF_{01} = 4.73$.

## References

American Psychological Association. (2017). *Clinical practice guideline for the treatment of PTSD*. Retrieved from https://www.apa.org/about/offices/directorates/guidelines/ptsd.pdf

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., . . . Perugini, M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108–119. doi:10.1002/per.1919

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory–II*. San Antonio, TX: Psychological Corporation.

Bisson, J. I., Roberts, N. P., Andrew, M., Cooper, R., & Lewis, C. (2013). Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. *Cochrane Database of Systematic Reviews*, *12*. Article CD003388. doi:10.1002/14651858.CD003388.pub3

Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *American Journal of Psychiatry*, *162*, 214–227. doi:10.1176/appi.ajp.162.2.214

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: a review of data and theory. *Psychological Bulletin*, *134*, 343–382. doi:10.1037/0033-2909.134.3.343

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89.

Engelhard, I. M., van den Hout, M. A., & Smeets, M. A. (2011). Taxing working memory reduces vividness and emotional intensity of images about the Queen's Day tragedy. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*, 32–37.

Engelhard, I. M., van Uijen, S. L., & van den Hout, M. A. (2010). The impact of taxing working memory on negative and positive memories. *European Journal of Psychotraumatology*, *1*, 5623, 1–8. doi:10.3402/ejpt.v1i0.5623

Fisher, R. A. (2006). *Statistical methods for research workers*. Guildford, England: Genesis.

Franco, A., Malhotra, N., & Simonovits, G. (2016). Under-reporting in psychology experiments: Evidence from a study registry. *Social Psychological and Personality Science*, *7*, 8–12. doi:10.1177/1948550615598377

Gunter, R. W., & Bodner, G. E. (2008). How eye movements affect unpleasant memories: Support for a working-memory

account. *Behaviour Research and Therapy*, *46*, 913–931. doi:10.1016/j.brat.2008.04.006

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177–189.

Houben, S. T., Otgaar, H., Roelofs, J., & Merckelbach, H. (2018). Lateral eye movements increase false memory rates. *Clinical Psychological Science*, *6*, 610–616.

Koole, S. L., & Lakens, D. (2012). Rewarding replications a sure and simple way to improve psychological science. *Perspectives on Psychological Science*, *7*, 608–614.

Lee, C. W., & Cuijpers, P. (2013). A meta-analysis of the contribution of eye movements in processing emotional memories. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*, 231–239.

Leer, A., Engelhard, I. M., Dibbets, P., & van den Hout, M. A. (2013). Dual-tasking attenuates the return of fear after extinction. *Journal of Experimental Psychopathology*, *4*, 325–340. doi:10.5127/jep.029412

Leer, A., Engelhard, I. M., Lennert, B., Struyf, D., Vervliet, B., & Hermans, D. (2017). Eye movement during recall reduces objective memory performance: An extended replication. *Behaviour Research and Therapy*, *92*, 94–105.

Leer, A., Engelhard, I. M., & van den Hout, M. A. (2014). How eye movements in EMDR work: Changes in memory vividness and emotionality. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*, 396–401.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361–366.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*, 537–542.

Maxfield, L., Melnyk, W. T., & Hayman, C. A. G. (2008). A working memory explanation for the effects of eye movements in EMDR. *Journal of EMDR Practice and Research*, *2*, 247–261.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika*, 175–240.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615–631.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, Article aac4716. doi:10.1126/science.aac4716

Parker, A., Buckley, S., & Dagnall, N. (2009). Reduced misinformation effect following saccadic bilateral eye movements. *Brain and Cognition*, *69*, 89–97.

Patihis, L., Ho, L. Y., Tingen, I. W., Lilienfeld, S. O., & Loftus, E. F. (2014). Are the "memory wars" over? A scientist-practitioner gap in beliefs about repressed memory. *Psychological Science*, *25*, 519–530.

Patihis, L., & Pendergrast, M. H. (2018). Reports of recovered memories of abuse in therapy in a large age-representative U.S. national sample: Therapy type and decade comparisons. *Clinical Psychological Science*, *7*, 3–21. doi:10.1177/2167702618773315

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. East Norwalk, CT: Appleton-Century-Crofts.

Sackett, D. L. (1979). Bias in analytic research. In M. A. Ibrahim (Ed.), *The case-control study consensus and controversy* (pp. 51–63). Amsterdam, The Netherlands: Elsevier.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*, 657–680.

Shapiro, F. (2017). *Eye movement desensitization and reprocessing (EMDR) therapy: Basic principles, protocols and procedures* (3rd ed.). New York, NY: Guilford Press.

Shaw, J., & Vredeveldt, A. (2019). The recovered memory debate continues in Europe: Evidence from the United Kingdom, the Netherlands, France, and Germany. *Clinical Psychological Science*, *7*, 27–28.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*, 559–569.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p*-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*, 666–681.

Strange, D., & Takarangi, M. K. (2012). False memories for missing aspects of traumatic events. *Acta Psychologica*, *141*, 322–326.

van den Hout, M. A., Bartelski, N., & Engelhard, I. M. (2013). On EMDR: Eye movements during retrieval reduce subjective vividness and objective memory accessibility during future recall. *Cognition & Emotion*, *27*, 177–183.

van den Hout, M. A., & Engelhard, I. M. (2012). How does EMDR work? *Journal of Experimental Psychopathology*, *3*, 724–738.

van den Hout, M., Muris, P., Salemink, E., & Kindt, M. (2001). Autobiographical memories become less vivid and emotional after eye movements. *British Journal of Clinical Psychology*, *40*, 121–130.

van Schie, K., Engelhard, I. M., & van den Hout, M. A. (2015). Taxing working memory during retrieval of emotional memories does not reduce memory accessibility when cued with reminders. *Frontiers in Psychiatry*, *6*, Article 00016. doi:10.3389/fpsyt.2015.00016

van Schie, K., van Veen, S. C., Engelhard, I. M., Klugkist, I., & van den Hout, M. A. (2016). Blurring emotional memories using eye movements: Individual differences and speed of eye movements. *European Journal of Psychotraumatology, 7*. doi:10.3402/ejpt.v7.29476

van Veen, S. C., Engelhard, I. M., & van den Hout, M. A. (2016). The effects of eye movements on emotional memories: Using an objective measure of cognitive load. *European Journal of Psychotraumatology*, *7*. doi:10.3402/ejpt.v7.30122

van Veen, S. C., van Schie, K., Wijngaards-de Meij, L. D., Littel, M., Engelhard, I. M., & van den Hout, M. A. (2015). Speed

matters: Relationship between speed of eye movements and modification of aversive autobiographical memories. *Frontiers in Psychiatry*, 6, Article 00045. doi:10.3389/fpsyt.2015.00045

Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria*, 35, 416–431.