Misinformation effects in an online sample: Results of an experimental study with a five day retention interval

Olivia Sievwright Corresp., 1 , **Michael Philipp** 1 , **Aaron Drummond** 1 , **Katie Knapp** 2 , **Kirsty Ross** 1

1 School of Psychology, Massey University, Palmerston North, New Zealand

2 Work-Learn Institute, University of Waterloo, Waterloo, Canada

Corresponding Author: Olivia Sievwright
Email address: Olivia.Sievwright.1@uni.massey.ac.nz

Traditional face-to-face laboratory studies have contributed greatly to our understanding of how misinformation effects develop. However, an area of emerging concern that has been relatively under-researched is the impact of misinformation following exposure to traumatic events that are viewed online. Here we describe a novel method for investigating misinformation effects in an online context. Participants (N = 99) completed the study online. They first watched a 10-minute video of a fictional school shooting. Between five and ten days later, they were randomly assigned to receive misinformation or no misinformation about the video before completing a recognition test. Misinformed participants were less accurate at discriminating between misinformation and true statements than control participants. This effect was most strongly supported by ROC analyses (Cohen's d = 0.59, BF10 = 8.34). Misinformation effects can be established in an online experiment using candid violent viral-style video stimuli.

Misinformation Effects in an Online Sample: Results of an Experimental Study with a Five Day Retention Interval

Olivia M. Sievwright[1], Michael C. Philipp[1], Aaron Drummond[1], Katie Knapp[2], and Kirsty Ross[1]

[1] School of Psychology, Massey University, Palmerston North, New Zealand
[2] Work-Learn Institute, University of Waterloo, Waterloo, Canada

Corresponding Author:
Olivia Sievwright[1]
School of Psychology, Massey University, Private Bag 11-222, Palmerston North 4442, New Zealand
Email address: Olivia.Sievwright.1@uni.massey.ac.nz

Abstract

Traditional face-to-face laboratory studies have contributed greatly to our understanding of how misinformation effects develop. However, an area of emerging concern that has been relatively under-researched is the impact of misinformation following exposure to traumatic events that are viewed online. Here we describe a novel method for investigating misinformation effects in an online context. Participants (N = 99) completed the study online. They first watched a 10-minute video of a fictional school shooting. Between five and ten days later, they were randomly assigned to receive misinformation or no misinformation about the video before completing a recognition test. Misinformed participants were less accurate at discriminating between misinformation and true statements than control participants. This effect was most strongly supported by ROC analyses (Cohen's d = 0.59, BF10 = 8.34). Misinformation effects can be established in an online experiment using candid violent viral-style video stimuli.

Introduction

For decades, researchers have demonstrated the harmful effects of misleading post-event information on memory. Numerous laboratory studies have shown that participants exposed to post-event misinformation frequently endorse misinformation as part of their original memory of an event; this is known as the misinformation effect (Loftus, 2005). The application of these findings in legal cases has shed light on the detrimental effects of misinformation in the real world and has contributed to the overturn of wrongful convictions. Misinformation effects may also lead people to remember events as being more traumatic than initially reported (Porter et al., 2008). This study builds on extant misinformation research by testing a novel method for investigating misinformation effects in an online context.

Misinformation effects are traditionally studied in face-to-face laboratory settings using a three-stage procedure (Loftus et al., 1978). First, participants witness a target event, often through a video depicting a crime. Second, after a delay, participants are exposed to misleading information about the target event – this is often presented through a fictional eyewitness account, fictional news report, memory test, or through discussion with a co-witness. Finally, when tested on their memory for the target event (ranging 5 minutes to 2 weeks later), participants frequently report misinformation as part of their original recollection. Misinformation effects have led participants to report seeing a thief using a hammer instead of screwdriver (McCloskey & Zaragoza, 1985), or even recall non-existent footage of the car crash in which Diana, Princess of Wales was killed (Ost et al., 2002). This procedure has taught us a great deal about how false memories develop in laboratory conditions (for a review see Loftus, 2005).

One area of emerging concern that has been relatively under-researched is the impact of misinformation following exposure to traumatic events that are viewed online. Advancements in Internet technology now offer us immediate access to coverage of real-world traumatic events (Peterson & Densley, 2017). Consequently, people are inadvertently exposing themselves to footage of such events during their regular Internet use. For example, within 24 hours of the

73   2019 Christchurch, New Zealand mosque attacks, the live-stream of the attack had been
74   uploaded to Facebook over 1.5 million times (Besley & Peters, 2019). This is particularly
75   concerning given accumulating evidence that traumatic content online can have harmful
76   psychological effects for viewers, including post-traumatic stress symptoms, depression, and
77   anxiety (e.g., Pfefferbaum et al., 2019; Redmond et al., 2019; Thompson et al., 2019).
78   Up-to-the-minute media coverage of unfolding events often includes inaccurate reports
79   based on incomplete information and mistaken details (Rapp & Salovich, 2018; Rich &
80   Zaragoza, 2016). For instance, during the Christchurch mosque attacks it was erroneously
81   reported that 'a good guy with a gun' stopped the shooter. In fact the man rushed at the shooter
82   with an electronic payment terminal (Cooke, 2019). Research suggests memory for highly
83   negative events is susceptible to distortion (Nahleen et al., 2020; Strange & Takarangi, 2012).
84   Exposure to inaccurate information facilitates misinformation effects for real-world traumatic
85   events witnessed through the media. People who view traumatic videos may be vectors of
86   distressing and inaccurate information to the general public. Thus, memory distortions from
87   misinformation effects may have clinical implications even in the absence of legal ones. For
88   instance, even if people who witness violent online videos are unlikely to be called as factual
89   event witnesses in a trial due to the ability to verify facts from the recorded footage, memory
90   distortions from misinformation effects may have implications for psychological distress and,
91   potentially, the development of clinical disorders such as PTSD. Therefore, it is imperative we
92   understand exactly how misinformation effects work in the context of traumatic online media
93   exposure.
94   Existing misinformation paradigms are not well-suited to investigating misinformation
95   effects in the context of media exposure to traumatic events. Typical misinformation studies use
96   staged videos of a burglary or car crash that clearly depict fictional events. These videos are
97   often not designed to evoke strong emotional responses. Additionally, the artificial laboratory
98   environment may lead to less intense emotional reactions than would typically occur when
99   witnessing a real-life traumatic event (Chae, 2010). Examining misinformation effects in an
100   online context minimises potential experimenter effects and increases ecological validity by
101   exposing people to traumatic media in the context people are most likely to encounter such
102   material—on their own computers.
103   Two important methodological decisions need to be made when measuring
104   misinformation effects: retention intervals and type of memory test used. Retention intervals are
105   the period between viewing the target event (encoding) and being exposed to misinformation
106   about the event, and the period between exposure to misinformation and memory retrieval.
107   Although longer retention intervals between encoding and misinformation exposure (i.e., at least
108   24 hours) produce larger misinformation effects (Frost et al., 2002; Paz-Alonso & Goodman,
109   2008), most studies use retention intervals of 30 minutes or less. Longer retention intervals
110   between encoding and misinformation exposure have theoretical and pragmatic advantages:
111   theoretically, increasing the retention interval should increase susceptibility to misinformation
112   because event memory decays over time (Conway, 2009; Sekeres et al., 2016). Longer retention

113   intervals also mimic real-world time intervals between witnessing an event and memory
114   retrieval. An online approach should help to minimise practical issues around longer retention
115   intervals because multiple participant visits to the laboratory are not required.
116   The type of memory test used is important to detecting misinformation effects.
117   Misinformation studies employ different types of memory tests to quantify memory accuracy and
118   misinformation endorsement, including forced-choice recognition, cued recall, free recall,
119   remember/know memory judgements, and source monitoring tests. One of the most common
120   memory tests used in misinformation research is a two-alternative forced-choice recognition test,
121   a test where participants' answers are categorically coded as correct or incorrect. In these tests,
122   participants are asked about some aspect of the target event and are given two possible response
123   options to choose from. For example, participants might be asked to decide whether the thief's t-
124   shirt was blue or green. Although a comprehensive analysis of the advantages and disadvantages
125   of the various memory tests is beyond the scope of the present manuscript, there is debate about
126   what type of test is optimal, particularly in the context of eyewitness evidence, and the primary
127   disadvantage of such categorical measures of memory is that they are relatively insensitive to
128   small shifts in discriminability that fail to reach participants' criteria to change their categorical
129   response (e.g., Wixted & Mickes, 2018).
130   In this study, we employ a different technique which may be able to detect small
131   discrepancies in people's ability to distinguish between memory for the original event and
132   memory for post-event misinformation. Rather than asking participants to decide between two
133   competing items, we ask participants to make a categorical judgement about how confident they
134   are that a particular item was present in the target video. For example, participants might be
135   asked how certain they are that the thief's t-shirt was blue on a scale from 1 (certain this is false)
136   to 6 (certain this is true). These ratings allow us to plot a receiver operating characteristic (ROC)
137   curve, which provides a clearer indication of precisely how memory accuracy differs at varying
138   confidence levels (Stanislaw & Todorov, 1999; Wixted & Mickes, 2018). The advantage of
139   using this type of recognition test is that it may detect subtle changes in memory strength and
140   bias when memory change does not meet the threshold for reporting misinformation as true
141   (Stanislaw & Todorov, 1999). ROC curves enable us to detect reduced certainty that
142   misinformation is false even when accuracy rates are not explicitly affected. For example,
143   misinforming participants that the thief's t-shirt was blue may cause participants to be less
144   certain that the thief's t-shirt was actually green. Although they may correctly report 'false' for
145   the blue t-shirt, the effect of misinformation can be seen in the reduced certainty ratings.
146   **Present Study**
147   This study tests the effectiveness of a novel experimental paradigm for investigating
148   misinformation effects in the context of media exposure to trauma using an ostensibly realistic,
149   online, violent viral-style video as the target event, and a one-week retention interval. We also
150   used more sensitive estimates of misinformation effects by measuring participants' confidence in
151   their memory reports to determine discriminability (d') and points on a ROC curve. To our

152 knowledge, this is the first study to measure misinformation effects after one week using such a
153 procedure.
154   Memory vividness and emotionality ratings have been examined in recent misinformation
155 and trauma therapy studies (e.g., Calvillo & Emami, 2019; Houben et al., 2018; van Schie &
156 Leer, 2019). To contribute to future research, a secondary aim of this study was to establish the
157 minimum therapeutically meaningful change in ratings of memory emotionality and vividness
158 across sessions. Although not elaborated on here, the findings from these analyses will help to
159 inform a baseline of memory vividness and emotionality for a future study.
160
161 ## Materials & Methods
162 All procedures reported here involving human participants were in accordance with the ethical
163 standards of the Massey University Human Ethics Committee (Notification number:
164 4000021787). Participants were informed of the purpose, procedures, and requirements of the
165 study prior to participating. Informed consent was implied through participation in the study and
166 all participants were debriefed as to the purpose of the study at the conclusion of each study.
167 Preregistration information for this study can be accessed at: https://aspredicted.org/ma4st.pdf.
168 The materials and data for this study are openly available on the Open Science Framework
169 (OSF): https://osf.io/r48bm/?view_only=8137681d7e044d06a6bef57d4604b32f
170
171 **Design**
172 This study consisted of two sessions spaced approximately one week apart (see Figure 1).
173 Session 1 was identical for everyone. In Session 2, half of participants were randomly assigned
174 to the misinformation condition and half were assigned to the no-misinformation control
175 condition. Memory vividness and emotionality ratings were compared across a 2(Condition:
176 misinformation or control) × 2(Time: session 1 and session 2) mixed factorial design. We carried
177 out an informal pilot test of the study material with a small group of students prior to collecting
178 data. Feedback from the pilot study indicated the study was clear and plausible.
179
180 **Participants**
181 Participation was restricted to those who were at least 18 years old. Potential participants were
182 asked not to participate if they had a current or previous diagnosis of post-traumatic stress
183 disorder, depression, or anxiety, or if they had been exposed to a traumatic or violent life event,
184 particularly gun-related violence. We aimed to recruit 100 participants to complete both
185 experimental sessions. This sample size was a compromise between collecting a larger sample
186 size than in previous similar misinformation studies (e.g., Houben et al., 2018) and resource
187 constraints.
188 One hundred and twenty-nine participants took part in Session 1 of the experiment. Data from 12
189 participants were excluded for not viewing the entire 10-minute video and data from a further 12
190 participants were excluded for failing both attention checks. Session 2 was made available to the
191 remaining 105 participants 5-10 days after they completed Session 1. Ninety-nine participants

192   completed Session 2. No participants expressed suspicion that incorrect details presented in the
193   misinformation narrative were intentional or part of the study.
194   The final sample ($N = 99$) included 53 people (33 male, 20 female) in the misinformation
195   condition and 46 (33 male, 11 female, 2 non-binary) in the control condition. The mean age of
196   participants did not differ between the misinformation condition ($M = 28.3$, SD $= 9.32$) and the
197   control condition ($M = 26.6$, SD $= 8.31$), $t(97) = 0.93$, $p = 0.36$. The mean retention interval
198   between Session 1 and Session 2 also did not differ between the misinformation condition ($M =$
199   5.49 days, SD $= 1.15$) and control condition ($M = 5.35$ days, SD $= 1.06$), $t(97) = 0.64$, $p = 0.53$.
200   **Measures/Materials**
201   **Trauma film.** In Session 1, participants viewed a 10-minute video comprised of excerpts
202   from the film 'Zero Day' (Coccio, 2002) to temporarily induce mild feelings of distress similar
203   to those experienced in reaction to witnessing a traumatic event. The film depicts a school
204   shooting and the moments leading up to the shooting using ostensible home-video footage and
205   security camera recordings. The video clip begins with two perpetrators introducing themselves
206   to the camera and planning their attack on the school. The perpetrators are then shown entering
207   the school and killing and tormenting multiple students via security camera recordings. The
208   video ends with the perpetrators taking their own lives as an emergency dispatch officer pleads
209   with them over the phone. An image of playing cards was inserted at the end of the video for
210   attention check purposes, described later.
211   **Vividness rating scale.** Memory vividness was measured using a Visual Analogue Scale
212   (VAS) during Session 1 and one week later in Session 2. In Session 1, participants rated how
213   vivid (clear) their memory of the trauma film was on a scale from 0 (not vivid at all) to 10
214   (extremely vivid). At Session 2, participants rated memory vividness using the same scale. They
215   also rated their perceived change in vividness from Session 1 using a 5-point scale with the
216   following response options: (1) a lot more vivid, (2) a little more vivid, (3) the same, (4) a little
217   less vivid, (5) a lot less vivid. We used this to determine the smallest effect size of interest
218   regarding the change in memory vividness over time (Anvari & Lakens, 2019).
219   **Emotionality rating scale.** Memory emotionality was also measured using a VAS in
220   Session 1 and 2. Participants rated how emotional their memory for the trauma film was on a
221   scale from 0 (extremely negative) to 10 (extremely positive). At Session 2, participants rated
222   their perceived change in memory emotionality from Session 1 to Session 2 on a 5-point scale,
223   the same as described above for memory vividness.
224   **Attention checks.** Two attention checks were included in Session 1 after participants
225   rated the emotionality and vividness of the video. The first attention check was a directed query;
226   participants were presented with a sliding scale and instructed to leave the question blank and not
227   to click on the scale. The second attention check was a multiple-choice question asking
228   participants what image appeared at the end of the trauma film. Response options included: a
229   cartoon gun, playing cards, schoolbag, or penguin. The correct answer was *playing cards*.
230   Participants who failed these attention checks were excluded from analyses.

231   **Misinformation manipulation.** Misinformation was introduced to participants assigned
232   to the misinformation condition in Session 2 through a 447-word fictional eyewitness narrative
233   describing the events depicted in the traumatic video. The narrative contained 12 true statements
234   about the video (e.g., 'The emergency dispatch lady called out for *Andre* to pick up the phone')
235   and eight misinformation statements (e.g., 'Andre was in the driver's seat of the car and he was
236   wearing a *blue t-shirt*' instead of a *camouflage shirt*). Participants in the control condition did not
237   receive any misinformation, but instead completed a filler-task for five minutes. In the filler-task,
238   participants searched for thirteen words associated with ice cream flavours that were hidden
239   among a 14 × 14 grid of letters. Participants dragged the computer mouse over the letters that
240   formed a word to solve the word search puzzle. After five minutes, participants were
241   automatically taken to the next task.
242   **Recognition test.** Memory accuracy and susceptibility to misinformation was measured
243   in Session 2 using a true/false recognition test with an associated confidence rating for each test
244   item. The test comprised 24 statements about the traumatic video (e.g., "The blonde perpetrator's
245   name was Cal", "During the shooting Andre tipped over a desk"), with eight statements directly
246   referring to misinformation details presented in the misinformation narrative (e.g., "The dark-
247   haired perpetrator's name was *Chris*" instead of *Andre*, "The first gun was retrieved from the
248   *backseat* of the car" instead of *boot*). The test also contained 4 false items (foil items) unrelated
249   to information presented in the narrative and 12 true statements about the video clip. Each
250   statement was presented individually, and participants were not able to skip items or go back to
251   previous items. Participants indicated how certain they were that each statement was either true
252   or false on a 6-point scale ranging from (1) certain this is false, to (6) certain this is true.
253   Responses of 1-3 for false items were counted as correct and responses of 4-6 for true items were
254   counted as correct (hit). Incorrect answers to misinformation items indicated endorsement of
255   misinformation (false alarm), with a higher number of these questions being incorrect indicating
256   greater susceptibility to misinformation. We measured participants' ability to discriminate
257   between misinformation items and true items using signal detection and receiver operating
258   characteristic curve analyses, described below.
259   **D-prime (d').** We used d' to measure participants' ability to accurately discriminate
260   between true statements and misinformation statements at test. d' is derived from the signal
261   detection theory (SDT) of recognition memory, which assumes recognition decisions are based
262   on evidence strength of previously encountered items and new (misinformation or foil) items at
263   test (Stanislaw & Todorov, 1999). Higher values of d' indicate a greater ability to accurately
264   discriminate between true statements and misinformation statements. If misinformation effects
265   are present in our study, and misinformation participants are poorer at discriminating between
266   true statements and misinformation statements as hypothesised, d' should be smaller for the
267   misinformation group than for the control group.
268   **Response bias (c).** We used the SDT measure c to compare group differences in response
269   bias. Response bias tells us whether there is a general tendency for participants to respond either
270   "true" or "false" on test items (Stanislaw & Todorov, 1999). In our study, negative values of c

271   represent a bias toward responding "true" and positive values of c represent a bias toward
272   responding "false". More negative or positive values of c indicate a stronger bias toward
273   responding "true" or "false", respectively. We expected there would be no difference in overall
274   response bias between the misinformation group and the control group. However, if
275   misinformation effects are present in our study, we would expect the misinformation group to
276   show a greater bias toward responding "true" on misinformation items compared to the control
277   group.

278   ***Receiver operating characteristic (ROC) curves.*** ROC curves provide a non-parametric
279   and atheoretical estimate of participants' ability to accurately discriminate between true
280   statements and misinformation statements (Stanislaw & Todorov, 1999). ROC curves are
281   produced by plotting hit rates against false alarm rates for all possible certainty ratings. Chance-
282   level performance, where the hit rate is equal to the false alarm rate, is represented as a diagonal
283   line. This occurs when participants are unable to discriminate between true items and
284   misinformation items and rely on guessing. Good discrimination accuracy is represented by a
285   curve that bows toward the left. The more the ROC curve bows toward the left, the greater the
286   discrimination accuracy. Based on our hypothesis for misinformation effects, we expect the ROC
287   curve for the control group to bow more toward the left than the ROC curve for the
288   misinformation group. We also expected the curves would show similar hit rates for the
289   misinformation and control groups, but the misinformation group would have higher false alarm
290   rates than the control group.

291   ***Area under the curve (AUC).*** To quantify discrimination accuracy based on the ROC
292   curves, we calculated AUC for the misinformation and control group ROC curves. AUC can be
293   interpreted intuitively as the proportion of times in which participants correctly discriminate true
294   statements from misinformation statements (Stanislaw & Todorov, 1999). AUC values typically
295   range from 0.5 (chance-level recognition performance) to 1 (perfect recognition performance).
296   Larger AUC indicates greater discrimination accuracy. Consistent with our misinformation effect
297   predictions, we expect the control group will have a larger AUC than the misinformation group,
298   representing poorer discrimination accuracy for the misinformation group.

299   **Procedure**

300   Participants were recruited via the online participant pool Prolific (www.prolific.co), and
301   the experiment was administered in Qualtrics (www.qualtrics.com). The experiment was
302   advertised as a study of memory for traumatic events involving viewing a 10-minute video of a
303   fictional school shooting and answering some questions about the video. The study was made
304   available for participants to complete on a desktop computer or laptop, and for those whom
305   English is their first language.

306   **Session 1.** Participants began by viewing the trauma video. If they did not watch the full
307   video, they were asked to report the timepoint in which they stopped the video. After the video
308   ended, participants rated the emotionality and vividness of the traumatic video and completed the
309   two attention checks. For exploratory purposes, participants were then invited to comment on
310   their experience of viewing the traumatic video. Finally, participants were thanked for their time,

311   provided contact details for various mental health support services, and reminded they may be
312   asked to participate in another session in 5 days' time. Session 1 took approximately 20 minutes
313   to complete.
314   **Session 2.** Five days later, participants who completed Session 1 and passed the attention
315   checks were invited to participate in Session 2. Participants were given five days to complete
316   Session 2. They again rated the emotionality and vividness of their memory for the traumatic
317   video and indicated their perceived change in emotionality and vividness from Session 1 to
318   Session 2. Those assigned to the misinformation condition read the misinformation narrative,
319   while those in the control condition completed a filler-task for 5 minutes. Following this, all
320   participants completed a filler-task for a further 5 minutes before completing the recognition test.
321   At the end of the session, participants were thanked for their time, debriefed as to the purpose of
322   the study, and provided with contact details of mental health support services.
323
Results
325   We used the signal detection measure d-prime (d') and receiver operating characteristic
326   (ROC) curves to assess participants' ability to correctly distinguish between true statements and
327   misinformation statements one week after viewing the traumatic video. We excluded non-
328   misinformation false items from the analysis so we could compare true items directly to
329   misinformation items. A flattening constant was applied to account for floor and ceiling effects
330   for hit rates and false alarm rates. Standard calculations were used for d', response bias (c), and
331   ROC (Stanislaw & Todorov, 1999). We report 95% confidence intervals for Cohen's d effect
332   size for each comparison. Post-hoc Bayes' factors for independent samples t-tests with a default
333   Cauchy prior (0, .707) were used to determine the relative evidence in favour of the null or
334   alternative hypothesis for each analysis. Where the assumption of equal variances was violated
335   for between-group comparisons, a Mann-Whitney U test (U) was used.
336   Results comparing discriminability (d') for the misinformation and control groups
337   showed a significant difference in correct discrimination between the misinformation ($M = 0.21$,
338   $SD = 0.56$) and control group ($M = 0.49$, $SD = 0.75$), $t(97) = -2.10$, $p = .04$, Cohen's $d = 0.42$,
339   95% CI [0.02, 0.82]. These results suggest exposure to misinformation one week after viewing a
340   traumatic video interfered with participants' ability to correctly distinguish between true and
341   false statements at test. However, we found contradictory results for the post-hoc $BF_{10} = 1.46$;
342   indicating only anecdotal evidence of this effect. We calculated c and AUC to see whether we
343   could further clarify the effect of misinformation on discrimination accuracy.
344   Group comparisons for c showed participants in the misinformation group ($M = 0.79$, SD
345   $= 0.28$) were slightly more biased towards responding "true" compared to control participants (M
346   $= 0.55$, $SD = 0.35$), $U = 699$, $p < .001$, $d = 0.76$, 95% CI [0.35, 1.17]. The moderate-to-large
347   effect size, and post-hoc $BF_{10} = 91.4$ indicates very strong evidence of the effect of
348   misinformation on a bias towards responding "true". There are two possible explanations for
349   this: first, the bias towards responding "true" may reflect misinformation effects; that is,
350   misinformation group participants responded "true" to a greater proportion of misinformation

351 items thereby inflating overall response bias. Second, exposure to misinformation after one week
352 may have weakened participants' memory for the original event, leading them to agree with
353 statements and respond "true" more often than disagreeing and responding "false" across all test
354 items. We further examined response bias using ROC analyses.
355   In addition to signal detection measures, we used area under the ROC curve as a more
356 sensitive measure of discriminability. We plotted hit rates against false alarm rates for both
357 groups at each level of certainty (from "certain this is false" through to "certain this is true").
358 Figure 2 displays the ROC curves for the misinformation and control groups. The ROC curves
359 further support the signal detection analyses above, with the control group demonstrating a
360 greater bow to the left, moving further away from chance-level performance, and indicating
361 superior discriminability compared to the misinformation group. The ROC curves also show
362 similar hit rates at each certainty level for the misinformation and control groups. However, the
363 misinformation group show higher false alarm rates than the control group. This suggests the
364 bias toward responding "true" for the misinformation group can more likely be explained by a
365 greater tendency for misinformation participants to respond "true" to misinformation statements
366 than control participants. If it were the case that the misinformation group had an overall
367 tendency to respond "true" across all test items, we would expect to see a greater effect on hit
368 rates.
369   We calculated AUC for the misinformation and control group ROC curves using the
370 inverse of the trapezoidal rule for AUC. This involved dividing the area above the curves into a
371 series of trapezoids and summing the area of each trapezoid. We subtracted this sum from 1 to
372 give the AUC. An independent samples t-test comparing AUC revealed a significant difference
373 between groups in discrimination accuracy. Participants in the control group correctly
374 determined whether a statement was true or false 77% (SD = 0.12) of the time when true
375 statements and misinformation statements were presented during the recognition test. However,
376 participants in the misinformation group correctly identified true and false items 70% (SD =
377 0.11) of the time, $t(97) = -2.91$, $p = .004$, $d = 0.59$, 95% CI [0.18, 0.99]. We found a post-hoc
378 $BF_{10} = 8.34$, indicating the data is 8.34 times more likely under the alternative hypothesis (i.e.,
379 that participants in the misinformation condition are worse at discriminating between true and
380 false statements) than the null. Although d' results showed some ambiguity, our AUC
381 comparisons show a clear effect of misinformation on participants' discrimination accuracy after
382 one week.
383 **Exploratory Analyses**
384   The analyses in this section were not pre-registered and are therefore exploratory. To
385 explore whether differences in discriminability were due to endorsement of misinformation, we
386 compared hit rates and false alarm rates for both groups. If differences can be attributed to the
387 effects of misinformation, we would expect the misinformation group to have a higher false
388 alarm rate than the control group, but no difference in hit rates between groups. Figure 3 displays
389 the group comparisons of hit and false alarm rates. We found no significant difference in hit rates
390 between the misinformation and control groups, $U = 1146$, $p = .60$. However, the false alarm rate

391   was significantly higher for the misinformation group than the control group, $U = 659$, $p < .001$,
392   $d = 0.91$. Post-hoc $BF_{10} = 1052$ suggests very strong evidence that between-group differences in
393   discriminability between true and false statements can be attributed to inflated false alarm rates
394   for participants exposed to misinformation.
395   **Emotionality and Vividness Ratings**
396   We used a global transition method (Anvari & Lakens, 2019) to determine the smallest
397   effect size of interest for changes in memory emotionality and vividness for the trauma video
398   from Session 1 to Session 2. We conducted these analyses to determine a cut-off point for the
399   smallest subjectively detectable change in memory vividness and emotionality that is
400   therapeutically meaningful. Emotionality ratings were missing from 3 participants (2
401   misinformation, 1 control) and vividness ratings were missing from 1 participant in the
402   misinformation group. Data from 96 participants were analysed for emotionality and 98
403   participants for vividness. Session 1 and Session 2 emotionality ratings were reverse-coded so
404   that higher ratings represented more negative emotionality.
405   Figure 4 shows the change in emotionality ratings from Session 1 to Session 2 for the two
406   groups. A mixed-model ANOVA established no significant main effect of time, $F(1,94) = 0.84$,
407   $p = .36$, $\eta^2 = 0.003$, or condition, $F(1,94) = 0.67$, $p = .42$, $\eta^2 = 0.005$. There was also no
408   significant interaction between time and condition for participants' ratings of emotionality,
409   $F(1,94) = 0.19$, $p = .67$, $\eta^2 = 0.001$. This suggests participants' memory for the school shooting
410   video was equally negative one week after viewing the video as it was immediately after viewing
411   the video.
412   To determine the minimum therapeutically meaningful change in memory emotionality
413   from Session 1 to Session 2, we grouped participants' responses into three categories: those
414   reporting "no change" ($n = 71$), "a little change" ($n = 23$), or "a lot of change" ($n = 3$). We
415   focused on participants who reported "a little more positive" or "a little more negative"
416   emotionality at session 2. These ratings were combined to form the "little change" category. We
417   conducted a Chi-Square test of independence to examine differences between conditions in the
418   number of participants subjectively reporting "no change", "a little change", or "a lot of change"
419   in memory emotionality. No significant differences were found between groups, suggesting no
420   effect of misinformation conditions on perceived change in memory emotionality, $\chi^2(2, N = 97)$
421   $= 0.81$, $p = .67$.
422   A paired-samples t-test found that, for participants who reported "a little change" in
423   emotionality from Session 1 ($M = 6.57$, SD = 3.04) to Session 2 ($M = 6.24$, SD = 1.51), there
424   was a non-significant mean decrease in emotionality of 0.33, $t(20) = 0.50$, $p = .63$, $d = 0.12$, 95%
425   CI [-0.32, 0.54]. This suggests even for those reporting a subjective sense of change, memory
426   emotionality remained stable over time. As a result, we were unable to calculate the smallest
427   therapeutically meaningful effect for emotionality; however, we were able to establish a clear
428   baseline for memory emotionality for the traumatic video over a one-week period.
429   Changes in ratings of memory vividness from Session 1 to Session 2 for the
430   misinformation and control groups are displayed in Figure 5. We found a main effect of time,

431   indicating an overall decrease in vividness for memory of the school shooting video from
432   Session 1 to Session 2, irrespective of experimental condition, $F(1,96) = 126.53$, $p < .001$, $\eta^2 =$
433   0.27. There was also a main effect of condition; participants in the misinformation group tended
434   to rate their memory for the school shooting video as more vivid than participants in the control
435   group $F(1,96) = 4.79$, $p = .03$, $\eta^2 = 0.03$. However, the time × condition interaction was non-
436   significant, $F(1,96) = 0.82$, $p = .37$, $\eta^2 = 0.002$.
437   Again, we grouped participants' responses into three categories: those reporting "no
438   change" ($n = 26$), "a little change" ($n = 58$), or "a lot of change" ($n = 15$) in memory vividness. A
439   Chi-Square test of independence examined differences between conditions in the number of
440   participants reporting "no change", "little change", or "a lot of change" in memory vividness
441   from Session 1 to Session 2. No significant differences were found between the misinformation
442   and control conditions, $\chi^2(2, N = 99) = 1.77$, $p = .41$. We focused on participants who reported
443   their memory was "a little more vivid" or "a little less vivid" at Session 2. A paired-samples t-
444   test found that, for participants who reported "little change" in memory vividness from Session 1
445   ($M = 8.10$, SD = 1.28) to Session 2 ($M = 6.26$, SD = 1.15), there was a mean decrease in
446   vividness of 1.84 over time, $t(57) = 9.87$, $p < .001$, $d = 1.30$. Memory vividness decreased
447   significantly more for the "little change" group than for the "no change" group, $t(82) = -2.19$, $p =$
448   .03, $d = 0.52$. This suggests effect sizes of at least $d = 1.30$ represent a therapeutically
449   meaningful change in memory vividness over time, with smaller effect sizes representing
450   changes that are too small to be subjectively perceived.
451

Discussion

453   To our knowledge, this is the first study to test misinformation effects for a violent viral-
454   style video following a one-week retention interval using an online platform. We found only
455   anecdotal evidence that misinformation impaired discrimination accuracy when using d'.
456   However, using area under the ROC curve, we established moderate evidence that participants
457   who received misinformation were less accurate when discriminating between true statements
458   and misinformation statements than those who received no misinformation. Our results are
459   consistent with other studies finding an effect of post-event misinformation on memory accuracy
460   and misinformation endorsement (e.g., Loftus et al., 1989; Takarangi et al., 2006).
461   Our findings highlight the added value of ROC analyses for detecting subtle changes in
462   memory strength, even when misinformation may be below the acceptance threshold. Moreover,
463   the medium effect sizes we found are particularly notable given that our study was conducted
464   online where the capacity for monitoring and control is often much lower than in traditional
465   laboratory experiments.
466   This study extends our understanding of misinformation effects by using novel stimuli,
467   procedures, and context. We elicited misinformation effects outside of a laboratory using a
468   wholly online experiment and realistic target event to simulate the violent viral-style videos that
469   people may come across online. Many participants commented on the realism of our video,
470   likening it to the 2019 Christchurch mosque shootings and the 1999 Columbine High School

471   massacre. Using an online format, misinformation effects were elicited from a diverse sample

472   that may be more representative of the general population. Moreover, we showed the robustness

473   of misinformation effects under relatively extreme experimental conditions—involving reduced

474   experimental control, a high degree of participant autonomy and trust, and a one-week retention

475   interval.

476   We also used more sensitive estimates of misinformation effects by measuring

477   participants' confidence in their memory reports and using this to determine discriminability (d')

478   and points on a ROC curve. Signal detection has been used in many previous misinformation

479   studies (e.g., Chan et al., 2017; Paz-Alonso et al., 2013; van Bergen et al., 2010), however ROC

480   curves are rarely employed in such research. Some argue that a key advantage is that ROC

481   analyses are unaffected by the assumptions about recognition memory that underlie d' (Wixted

482   & Mickes, 2018). Although the direction of effects in our study was consistent across analyses,

483   ROC analyses provided stronger evidence of misinformation effects. This is because ROC

484   curves enabled us to detect instances where misinformation reduced participants' certainty in a

485   recognition judgement, even when they did not overtly accept the misinformation. Future

486   research should consider employing ROC analyses as an additional measure to allow for

487   detection of subtle changes in memory strength in misinformation effect research.

488   We also explored participants' ratings of vividness and emotionality for their memory of

489   the traumatic video, both immediately after viewing and one week later. Irrespective of exposure

490   to misinformation, participants rated their memory for the video as highly vivid and negative.

491   Our results are consistent with other trauma film paradigm studies which have found that

492   experimental analogues to traumatic events can produce intrusive and distressing memories

493   (James et al., 2016). Our findings also contribute to research on the psychological effects of

494   media exposure to traumatic events; although participants in our study were aware the traumatic

495   video was fictional, they appear to have rated their emotions and the vividness of their memories

496   as relatively intense. This provides some indication of the impact of media exposure to *real-life*

497   traumatic events, such as the live-streamed footage of the Christchurch attacks, which is likely to

498   be even greater.

499   Interestingly, participants rated their memory as being equally negative one week after

500   initially viewing the video but experienced a small decrease in memory vividness over the one-

501   week period. Our results are partially consistent with findings from similar studies, which have

502   found reductions in memory vividness over time (Houben et al., 2018; van Schie & Leer, 2019).

503   However, these studies also found memory emotionality became more positive over time, which

504   we were unable to replicate. A possible explanation for this is that our video was more

505   distressing than the staged car crash video used in previous studies, thereby we were able to

506   produce more enduring emotional reactions.

507   Memory vividness tended to be higher for misinformed participants than control

508   participants—across both sessions. According to prominent false memory and recognition

509   decision theories (e.g., fuzzy-trace theory, source monitoring framework, and ballistic

510   accumulator models), poorer memory vividness increases susceptibility to misinformation

511 (Brainerd & Reyna, 2002; Brown & Heathcote, 2008; Johnson et al., 1993). Based on this
512 assumption and our vividness results, we would expect the misinformation group to have better
513 memory performance than the control group. However, we found the opposite; despite
514 perceiving their memory as more vivid, the misinformation group had poorer discrimination
515 accuracy than the control group. Our findings suggest memory vividness did not interfere with
516 misinformation effects and demonstrate that misinformation effects can still occur for memories
517 that are experienced as being relatively vivid.
518   One limitation of this novel misinformation paradigm is that we cannot control for
519 extraneous variables as well in the online setting compared to a laboratory environment.
520 Distraction during the encoding or testing phases may have affected some participants'
521 performance on the recognition test. We attempted to minimise such effects by including
522 multiple attention checks throughout the experiment. A second limitation of this study may be
523 the objectively short interval of five minutes between exposure to misinformation and memory
524 retrieval; however, the implications of this are largely unknown. Further studies should
525 investigate whether the effect of misinformation on retrieval is affected by this interval.
526 Additionally, although participants may have incorrectly responded to recognition test items
527 based on their memory for the narrative, rather than the video, we believe this is unlikely for two
528 reasons. First, we provided participants with clear instructions to answer the questions based on
529 their memory for the video. Second, were participants responding based upon the narrative, we
530 would have expected to see a much larger increase in acceptance of the misinformation items
531 than the moderate effects we observed. Lastly, it is also possible that some participants searched
532 YouTube for the target video between sessions or took notes during initial viewing since we told
533 participants they would be asked questions about the video. However, this seems unlikely in our
534 study, given that no participants achieved recognition accuracy greater than 91.7% at test.
535 Furthermore, the fact we still found significant effects in our study highlights the robustness of
536 misinformation effects.
537

Conclusions

539 This study tested the viability of a novel method for investigating misinformation effects in an
540 online context. To our knowledge, this is the first study to investigate misinformation effects
541 after 1 week using a fully online sample with ostensibly realistic, viral-style, traumatic stimuli.
542 Using signal detection and area under the receiver operating characteristic curve, results showed
543 participants exposed to misinformation one week after viewing a traumatic video were
544 significantly worse at discriminating between true information and misinformation compared to
545 participants not exposed to misinformation. Our results suggest misinformation effects can be
546 established in an online experiment using candid violent viral-style video stimuli. This study also
547 provides evidence for the validity of ROC analyses in misinformation research. We hope our
548 novel materials can further contribute to understandings of memory, misinformation effects, and
549 media exposure to trauma at a faster pace and with more diverse samples than previously
550 achieved.

551

552

553

554

555

556

557

558

559 **Acknowledgements**

560 We wish to thank Malcolm Louden for technical assistance in the design of this study.

561

562

563

564

565

566

References

Anvari F, Lakens D. 2019. Using anchor-based methods to determine the smallest effect size of interest. *PsyArXiv*. DOI: https://doi.org/10.31234/osf.io/syp5a.

Besley T, Peters MA. 2019. Terrorism, trauma, tolerance: Bearing witness to white supremacist attack on Muslims in Christchurch, New Zealand. *Educational Philosophy and Theory*:1–11. DOI: 10.1080/00131857.2019.1602891.

Brainerd C, Reyna VF. 2002. Fuzzy-trace theory and false memory. *Current Directions in Psychological Science* 11:164–169.

Brown SD, Heathcote A. 2008. The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology* 57:153–178. DOI: 10.1016/j.cogpsych.2007.12.002.

Calvillo DP, Emami AS. 2019. Do lateral eye movements increase susceptibility to misinformation? A registered replication. *Psychonomic Bulletin & Review*. DOI: 10.3758/s13423-019-01641-6.

Chae Y. 2010. Application of laboratory research on eyewitness testimony. *Journal of Forensic Psychology Practice* 10:252–261. DOI: 10.1080/15228930903550608.

Coccio B. 2002. *Zero Day*. Avatar Films.

Conway MA. 2009. Episodic memories. *Neuropsychologia* 47:2305–2313. DOI: 10.1016/j.neuropsychologia.2009.02.003.

Cooke H. 2019. Christchurch terror attack: Mistruths and conspiracy theories. *Stuff*.

Forgas JP, Laham SM, Vargas PT. 2005. Mood effects on eyewitness memory: Affective influences on susceptibility to misinformation. *Journal of Experimental Social Psychology* 41:574–588. DOI: 10.1016/j.jesp.2004.11.005.

Frost P, Ingraham M, Wilson B. 2002. Why misinformation is more likely to be recognised over time: A source monitoring account. *Memory* 10:179–185.

Horry R, Colton LM, Williamson P. 2014. Confidence-accuracy resolution in the misinformation paradigm is influenced by the availability of source cues. *Acta Psychologica* 151:164–173. DOI: 10.1016/j.actpsy.2014.06.006.

Houben STL, Otgaar H, Roelofs J, Merckelbach H. 2018. Lateral eye movements increase false memory rates. *Clinical Psychological Science* 6:610–616. DOI: 10.1177/2167702618757658.

James EL, Lau-Zhu A, Clark IA, Visser RM, Hagenaars MA, Holmes EA. 2016. The trauma film paradigm as an experimental psychopathology model of psychological trauma: intrusive memories and beyond. *Clinical Psychology Review* 47:106–142. DOI: 10.1016/j.cpr.2016.04.010.

Johnson MK, Hashtroudi S, Lindsay DS. 1993. Source monitoring. *Psychological Bulletin* 114:3–28.

Loftus EF. 2005. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory* 12:361–366. DOI: 10.1101/lm.94705.

Loftus EF, Donders K, Hoffman HG, Schooler JW. 1989. Creating new memories that are quickly accessed and confidently held. *Memory & Cognition* 17:607–616. DOI: 10.3758/BF03197083.

Loftus EF, Miller DG, Burns HJ. 1978. Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory* 4:19–31. DOI: 10.1037/0278-7393.4.1.19.

McCloskey M, Zaragoza M. 1985. Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General* 114:1–16. DOI: 10.1037//0096-3445.114.1.1.

609 Nahleen, S., Strange, D., & Takarangi, M. K. T. (2020). Does emotional or repeated misinformation
610   increase memory distortion for a trauma analogue event? *Psychological Research*.
611   https://doi.org/10.1007/s00426-020-01409-x.
612 Ost J, Vrij A, Costall A, Bull R. 2002. Crashing memories and reality monitoring: Distinguishing
613   between perceptions, imaginations and 'false memories.' *Applied Cognitive Psychology* 16:125–
614   134. DOI: DOI: 10.1002/acp.779.
615 Paz-Alonso PM, Goodman GS. 2008. Trauma and memory: Effects of post-event misinformation,
616   retrieval order, and retention interval. *Memory* 16:58–75. DOI:
617   https://doi.org/10.1080/09658210701363146.
618 Peterson J, Densley J. 2017. Cyber violence: What do we know and where do we go from here?
619   *Aggression and Violent Behavior* 34:193–200. DOI: 10.1016/j.avb.2017.01.012.
620 Pfefferbaum B, Nitiéma P, Newman E. 2019. Is viewing mass trauma television coverage associated with
621   trauma reactions in adults and youth? A meta-analytic review. *Journal of Traumatic Stress*
622   32:175–185. DOI: 10.1002/jts.22391.
623 Porter, S., Taylor, K., & ten Brinke, L. (2008). Memory for media: Investigation of false memories for
624   negatively and positively charged public events. *Memory, 16*(6), 658–666.
625   https://doi.org/10.1080/09658210802154626
626 Rapp DN, Salovich NA. 2018. Can't we just disregard fake news? The consequences of exposure to
627   inaccurate information. *Policy Insights from the Behavioral and Brain Sciences* 5:232–239. DOI:
628   10.1177/2372732218785193.
629 Redmond S, Jones NM, Holman EA, Silver RC. 2019. Who watches an ISIS beheading—and why.
630   *American Psychologist* Advance online publication. DOI: http://dx.doi.org/10.1037/amp0000438.
631 Rich PR, Zaragoza MS. 2016. The continued influence of implied and explicitly stated misinformation in
632   news reports. *Journal of experimental psychology. Learning, memory, and cognition* 42:62–74.
633   DOI: 10.1037/xlm0000155.
634 van Schie K, Leer A. 2019. Lateral eye movements do not increase false-memory rates: A failed direct-
635   replication study. *Clinical Psychological Science* 7:1159–1167. DOI:
636   10.1177/2167702619859335.
637 Sekeres MJ, Bonasia K, St-Laurent M, Pishdadian S, Winocur G, Grady C, Moscovitch M. 2016.
638   Recovering and preventing loss of detailed memory: differential rates of forgetting for detail
639   types in episodic memory. *Learning & Memory* 23:72–82. DOI: 10.1101/lm.039057.115.
640 Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behavior Research*
641   *Methods, Instruments, & Computers* 31:137–149. DOI: 10.3758/BF03207704.
642 Strange, D., & Takarangi, M. K. T. (2012). False memories for missing aspects of traumatic events. *Acta*
643   *Psychologica, 141*(3), 322–326. https://doi.org/10.1016/j.actpsy.2012.08.005
644 Takarangi MKT, Parker S, Garry M. 2006. Modernising the misinformation effect: the development of a
645   new stimulus set. *Applied Cognitive Psychology* 20:583–590. DOI: 10.1002/acp.1209.
646 Thompson RR, Jones NM, Holman EA, Silver RC. 2019. Media exposure to mass violence events can
647   fuel a cycle of distress. *Science Advances* 5:eaav3502. DOI: 10.1126/sciadv.aav3502.
648 Wixted JT, Mickes L. 2018. Theoretical vs. empirical discriminability: the application of ROC methods
649   to eyewitness identification. *Cognitive Research: Principles and Implications* 3:9. DOI:
650   10.1186/s41235-018-0093-8.
651

Figure 1

Schematic representation of the experimental procedure for Session 1 and Session 2.
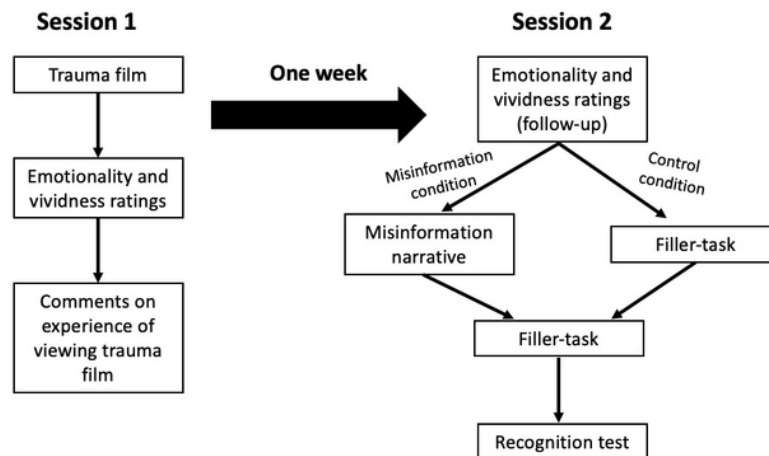
Figure 2

A comparison of receiver operating characteristic (ROC) curves plotting hit rates against false alarm rates between misinformation and control conditions.

The grey dotted line denotes chance performance, where the hit rate is equivalent to the false alarm rate. Each data point along the curve represents a different certainty level for recognition responses, ranging from "certain this is false" to "certain this is true".
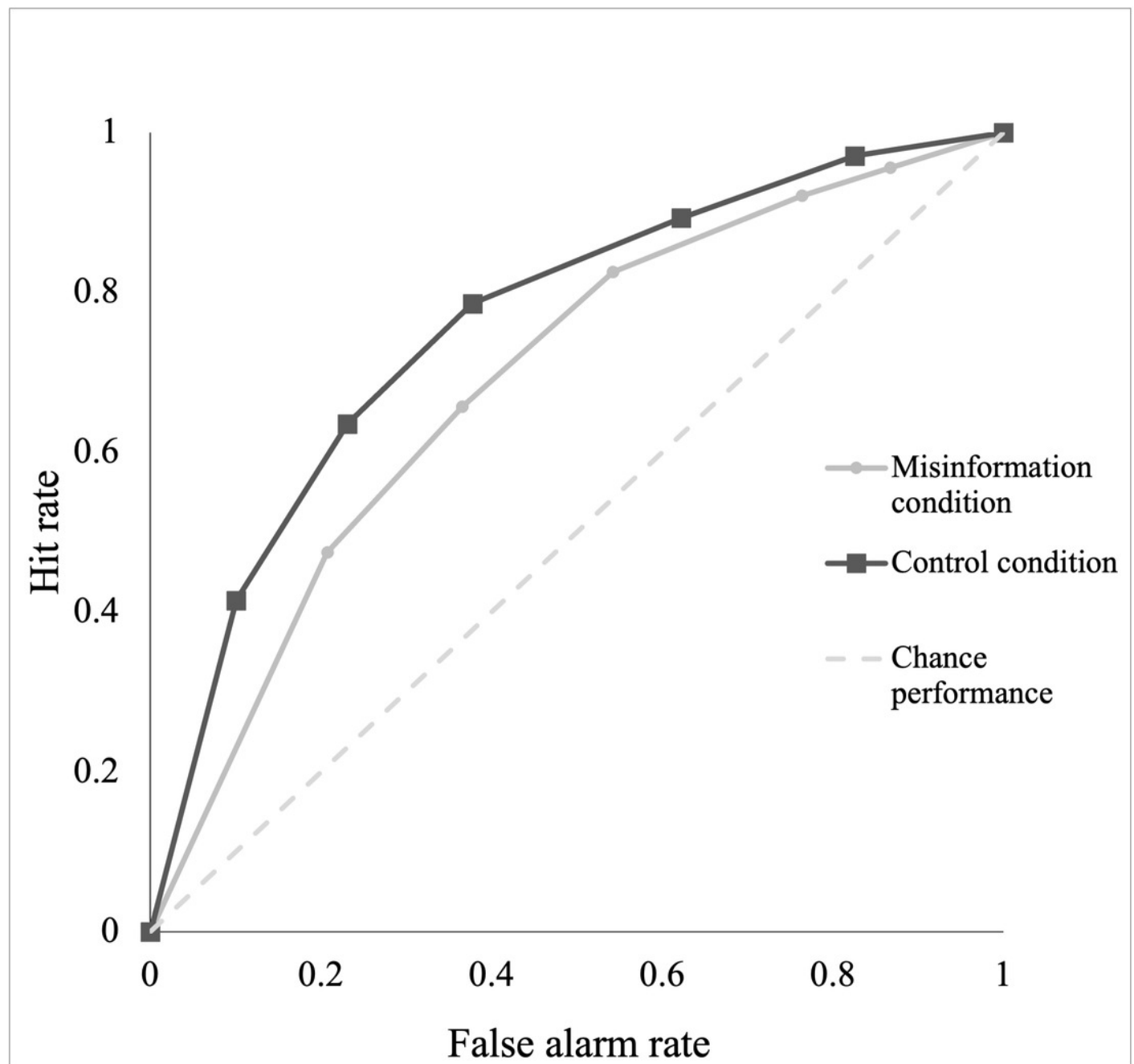
Figure 3

Recognition test hit and false alarm rates for the misinformation group and control group.

Error bars demonstrate standard error of the mean (SEM) for hit and false alarm rates.
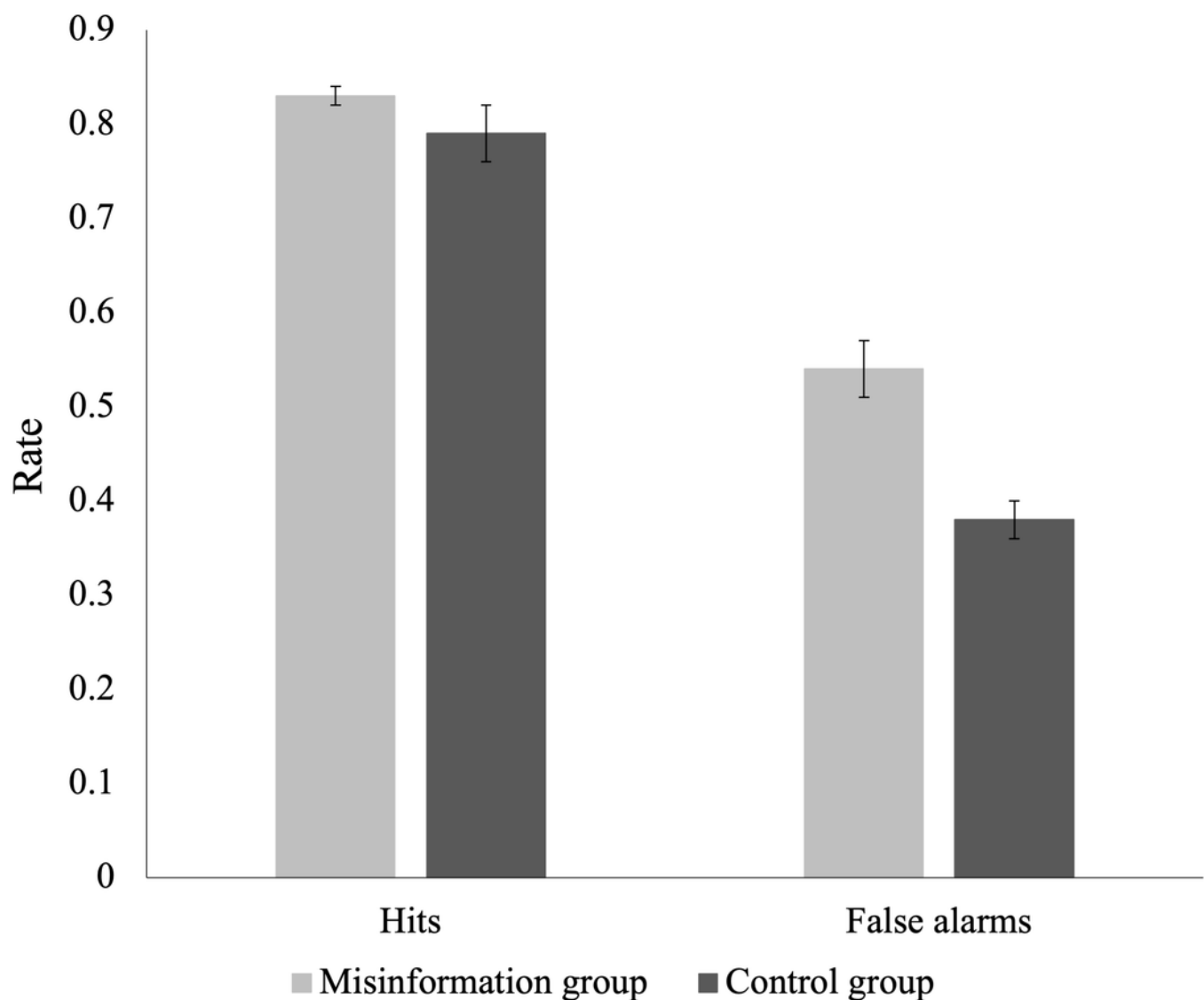
Figure 4

Change in emotionality ratings for the traumatic video from Session 1 to Session 2 for the misinformation and control groups.

Error bars demonstrate standard error of the mean (SEM) for ratings of memory emotionality.
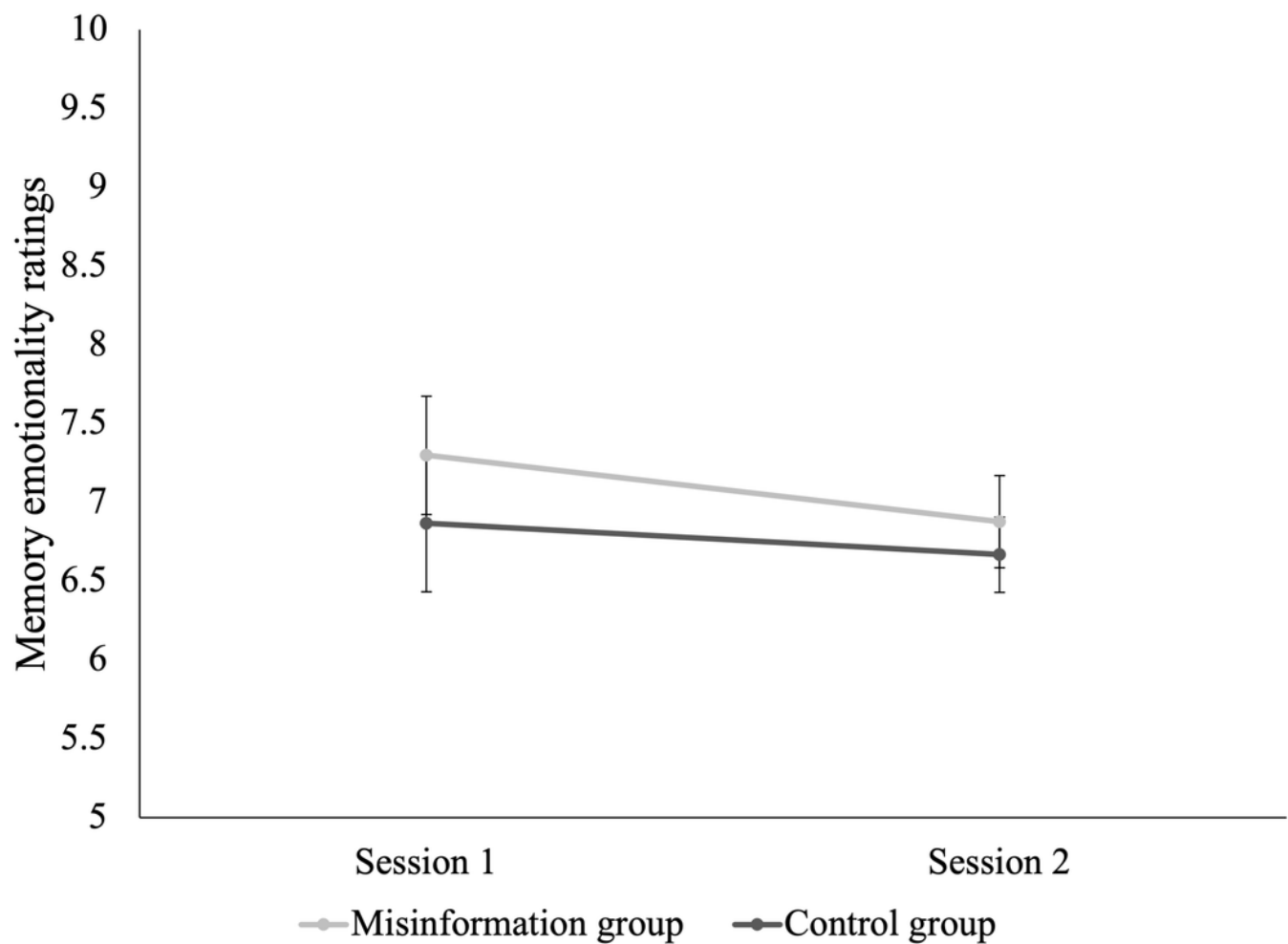
Figure 5

Change in vividness ratings for the traumatic video from Session 1 to Session 2 for the misinformation and control groups.

Error bars demonstrate standard error of the mean (SEM) for ratings of memory vividness.