CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Fool me twice: How effective is debriefing in false memory studies?
Authors	Murphy, Gillian;Loftus, Elizabeth;Hofstein Grady, Rebecca;Levine, Linda J.;Greene, Ciara M.
Publication date	2020-08-07
Original Citation	Murphy, G., Loftus, E., Hofstein Grady, R., Levine, L. J. and Greene, C. M. (2020) 'Fool me twice: How effective is debriefing in false memory studies?', Memory, 28(7), pp. 938-949. doi: 10.1080/09658211.2020.1803917
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1080/09658211.2020.1803917
Rights	© 2020, Informa UK Limited, trading as Taylor & Francis Group. All rights reserved. This is an Accepted Manuscript of an item published by Taylor & Francis in Memory on 7 August 2020, available online: https://doi.org/10.1080/10.1080/09658211.2020.1803917
Download date	2026-11-01 08:36:31
Item downloaded from	https://hdl.handle.net/10468/10516



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Murphy, G., Loftus, E., Grady, R. H., Levine, L. J., & Greene, C. M. (2020). Fool me twice: how effective is debriefing in false memory studies? *Memory*, *28*(7), 938-949.

Fool Me Twice: How Effective is Debriefing in False Memory Studies?

Gillian Murphy[1]

Elizabeth Loftus[2]

Rebecca Hofstein Grady[2]

Linda J. Levine[2]

Ciara M. Greene[3].

[1]School of Applied Psychology, University College Cork, Ireland.

[2]Department of Psychological Science, University of California, Irvine, CA, USA.

[3]School of Psychology, University College Dublin, Ireland.

Correspondence may be sent to Gillian Murphy, Ph.D., School of Applied Psychology, University College Cork, Cork, Ireland. gillian.murphy@ucc.ie. Tel: 00353 21 4890 4507

Abstract

Deception is often necessary in false memory studies, especially when the study aims to explore the effect of misinformation on memory. At the end of the study, participants are debriefed, but does this eliminate the influence of misinformation? In the current study, we followed up 630 participants six months after they participated in a study in which they were exposed to fabricated political news stories. We compared the memories of these 'continuing participants' for both novel and previously seen news stories to the memories of 474 newly recruited participants. Relative to new recruits, continuing participants were less likely to report a false memory for a story that they had been previously exposed to, and they were also less likely to report a false memory for a novel fake news story. Continuing participants were more likely to report a memory for previously seen true events than novel true events. Both groups of participants reported enjoying the experience and feeling confident that they understood which stories were fabricated. Importantly, this study did not find any negative long-term effects of participating in our false memory experiment, and even exhibited some positive effects.

Keywords: False memory, misinformation, debriefing, politics

False memory studies often involve necessary deception. Participants are presented with some misinformation and their acceptance or rejection of that misinformation is recorded. Then, typically, participants are debriefed and told about the misinformation before they leave the study. For example, a participant might see a thief with a green jacket steal a wallet and later be told that the jacket was brown. Many will later 'remember' seeing a brown jacket. The participant is debriefed at the end of the study and we assume we have 'taken back' that misinformation. But what has happened to their memory? If tested later, would they now remember the true detail (green) or would they remember the misinformation (brown)? In other areas of psychological research, studies have shown that retracting information does not entirely remove its sway (Cabalo et al., 2020; de Keersmacker & Roets, 2017; Thorson, 2016). This 'continued influence effect' (Johnson & Seifert, 1994) is often evident even when warnings are provided to participants before exposure (Ecker et al., 2010). Individuals may experience source-confusion, where they remember the misinformation but not where it came from (Johnson et al., 1993) or the correction may even backfire, increasing belief in the misinformation (Lewandowsky et al., 2012).

It is clear from the literature that retraction of misinformation is difficult and the consequences can be complex. Clark et al., (2012) had participants perform actions in the lab and two days later showed them a video that was doctored to appear as if they had performed additional actions. Participants were debriefed, which reduced *belief* that they had performed these additional actions. However, debriefing did not eliminate participants' false *memories* of performing these actions. There is a large body of literature that suggests that recollection and belief are independent-co-occurring constructs, and 'non-believed memories' can remain, after a memory is refuted (Mazzoni et al., 2010). Belief can be described as 'the truth value attributed to the occurrence of an event' (Scoboria et al., 2014, p 1243). As pointed out by Scoboria and colleagues (2004),

one can believe that an event happened without remembering it (such as believing you were born). One can also recollect something without believing it really happened (such as remembering seeing Santa Claus come down the chimney as a child). Otgaar et al. (2013) planted false memories for a childhood experience and later debriefed participants, informing them that the event had never occurred. Of those who had described a false memory during the study, forty percent reported a non-believed memory after the debriefing. Outside of laboratory settings, research has shown that non-believed autobiographical memories are relatively common amongst the general public (Mazzoni et al., 2010; Otgaar et al., 2019; Scoboria & Talarico, 2013). This raises the concern that reducing belief in a fabricated event may not entirely eliminate a false memory (Otgaar et al., 2014).

There are many contexts where the deliberate and persistent alteration of memory and beliefs may be especially problematic. One such area of study is the effect of 'fake news', and many studies have attempted to plant false memories for fabricated political events (e.g. Frenda et al., 2013; Murphy et al. 2019; Nash, 2018; Saachi et al., 2007). A 2018 meta-analysis showed that misinformation related to real-world events is more challenging to correct than constructed misinformation (e.g., referring to a fictional crime), and that misinformation pertaining to politics is especially resistant to retraction (Walter & Murphy, 2018). In particular, misinformation that is in line with existing beliefs is generally processed more fluently and thus less likely to be successfully retracted (Lewandowsky et al., 2012). It is important to assess whether the debriefing procedures in these political false memory studies are sufficient. After participants are debriefed and return to their daily lives, are they more vulnerable than non-participants to forming false memories for the event that they were presented with during the study?

When assessing whether exposure to a news story during a research study might increase false memories or beliefs for that news story later, we might also consider the

wealth of research on the illusory truth effect. Hasher, Goldstein & Toppino (1977) presented participants with trivia statements on three occasions and had them rate their accuracy. They found that statements that had been previously encountered were rated as more accurate than novel statements. Numerous studies have replicated this effect (see Dechêne et al. (2010) for a meta-analysis). The dominant explanation for this 'illusory truth effect' is that repeated exposure increases processing fluency, leading to increased perceptions of accuracy (Unkelbach & Greifeneder, 2013). Recently, Pennycook et al., (2018) assessed the illusory truth effect in relation to fake news. Participants were exposed to political fake news headlines such as *'Mike Pence: Gay Conversion Therapy Saved my Marriage'* in a familiarization stage and then again in an assessment stage where they were asked to rate the accuracy of both novel and previously-seen headlines. Participants were twice as likely to perceive a story as true if they had previously encountered it, both when the second stage took place immediately and after a one-week delay. Across seven studies investigating the illusory truth effect for trivia statements and fake news headlines, De keersmaecker et al., (2019) found that this effect was not moderated by individual differences in cognitive ability, need for cognitive closure or cognitive style. Thus it appears the illusory truth effect is a robust phenomenon and repeated exposure to inaccurate information is likely to increase perceived truthfulness. As participation in a false memory study can involve repeated exposure to fabricated information (during the study and perhaps again during the debriefing process), it is possible that this illusory truth effect may increase rates of false memory at follow-up.

In the current study, as well as assessing the potential risks of participating in a false memory study, we assessed some potential benefits. There is a lack of research on participant perspectives on false memory studies, though anecdotally, many participants report enjoying memory tests and fake news surveys, despite the deception involved. As stated in the British Psychological Society's Code of Human Research Ethics: 'If the

reaction of participants when deception is revealed later in their participation is likely to lead to discomfort, anger or objections from the participants then the deception is inappropriate' (British Psychological Society, 2014, p24). As there is mixed evidence as to whether participants often object to deception in psychology studies (Hertwig & Ortmann, 2008), we assessed how participants in the current study felt about the false memory manipulation and whether they would participate in such a study again. Furthermore, we assessed whether previous participation in our false memory study may have been educational; specifically, we evaluated whether participants estimated false memories to be more or less common if they had previously taken part in a false memory study and been debriefed. There is debate as to how widespread memory myths are amongst the general public (Brewin et al., 2019; Otgaar et al., 2020; Simons & Chabris, 2011) and it is unclear if participation in a false memory study impacts these opinions.

The Current Study

In the current study, we examined the memories of participants who six months previously had been exposed to fake news, and debriefed after the study. The false memory study concerned campaign events from the 2018 Irish abortion referendum (see Murphy, et al., 2019, for full details). The study took place in May 2018, one week before the referendum. Participants were shown real and fabricated news stories about the campaign and asked if they remembered them and how they had felt at the time. We found that fake news stories were more likely to be "remembered" by voters whose belief the story supported (i.e., a made-up scandal about the No campaign was more likely to be falsely remembered by Yes voters, and vice versa). In the current study, we followed up with these original participants, presenting them with previously seen and novel news stories. We compared true and false memories reported by these participants at follow-up with those reported by a group of newly recruited control participants. We tested two

competing hypotheses: 1) that debriefing would *reduce* future false memory effects, as intended, and 2) that debriefing would *increase* false memory effects, in line with the illusory truth effect. We also investigated whether participants enjoyed taking part in the study and believed it had improved their understanding of the topic, as well as whether it resulted in higher estimates of the prevalence of false memories.

Method

Participants

The sample consisted of 1106 participants, two of whom were removed after indicating in their responses that they had attended a talk given by one of the authors about the first research study and used that to identify the fabricated stories. This left a total of 1104 eligible participants, 630 'continuing participants' from the prior study and 474 newly recruited participants.

Continuing participants. For participants who had been previously exposed to the fake news stories (n = 630), the average age was 37.50 (SD = 13.68). Most (443) were female, 182 were male, and two selected Prefer Not to Say. The majority of the sample (90%, n = 569) was classified as Yes voters (537 who voted Yes to repeal the ban on abortion and 32 who didn't vote but would have voted yes). There were 60 No voters (57 who voted to retain the ban on abortion and three who didn't vote but would have voted no). One participant selected 'Didn't Vote and Wasn't Leaning Either Way'. These participants were recruited for the original study via an article on an online news site (TheJournal.ie), social media, and university staff and student emails.

Newly recruited participants. For participants who had not been previously exposed to the fake news stories (n = 474), the average age was 38.45 (SD = 12.88). Of these participants, 334 were female, 135 were male, and five selected Other or Prefer Not

to Say. The majority of the sample (88%, *n* = 418) were classified as Yes voters (383 who voted Yes to repeal the ban on abortion and 35 who didn't vote but would have voted yes). There were 52 No voters (46 who voted to retain the ban on abortion and 6 who didn't vote but would have voted no). The remaining participants selected 'Didn't Vote and Wasn't Leaning Either Way' (*n* = 3) or 'Prefer Not to Say' (*n* = 1). These participants were recruited for the longitudinal study via an article on the same online news site (TheJournal.ie), social media, and university student emails.

The continuing and newly recruited participants were not significantly different from each other on a number of measures assessed at the six-month follow-up, including age, t (1076) = 1.16, p = .25, and a range of referendum-related items: personal importance of the referendum, t (966.02) = 1.55, p = .12; referendum-related media consumption, t (1100) = 0.26, p = .79; and referendum-related discussions with family and friends, t (1100) = 0.94, p = .35.

Previous Study

Our first study was conducted via an online survey in the week preceding the abortion referendum, which took place on May 26th 2018 (for full details see Murphy et al., 2019). Along with four true news stories, participants were presented with two fabricated stories: one concerning illegal campaign posters funded by foreign lobby groups and one concerning distasteful comments made by campaigners. Both stories had two versions; for example, one version in which the Yes campaign had to destroy illegal posters, and one version in which the No campaign had to destroy illegal posters. Participants saw only one version of each story in this first study (i.e., either the Yes or No version of each item). Participants were first asked if they remembered each event. They were then told that they may have been presented with some fake news and were asked to pick any stories they suspected to be fake. Finally, participants were debriefed and told 'you were shown

two entirely fabricated events - the stories below were invented by the researchers and never happened. You should know that these events are entirely false and should in no way be taken into account when deciding how to vote.' Participants were then presented with the fake story once again, with the word FALSE printed before the story text. All continuing participants had therefore seen the fake news story and been debriefed, though not all had reported a false memory for the event.

Materials

News Stories

Participants viewed four news stories; one that was novel and true, one that was novel and fabricated, one taken from the previous study that was true and one taken from the previous study that was fabricated. The stories are shown in Appendix A. The stories were counterbalanced by block so that participants saw a mix of stories about each side of the referendum. Continuing participants were randomly assigned to view either the same story they had seen in the pre-referendum study or the other unseen version. For example, out of those that saw the Yes version of the poster story originally, half of them saw the Yes poster story again in this study, while the other half saw the No version of the poster story in this study. Participants were randomized to view one of two blocks of news stories. Block One contained the repeated fake story about illegal Yes posters, a repeated true story about actors supporting the Yes campaign, a novel true story about the Pope criticizing the referendum outcome, and a novel fake story about Facebook apologizing for allowing foreign interference in the referendum. Block Two contained the repeated fake story about illegal No posters, a repeated true story about athletes supporting the No campaign, a novel true story about a pro-choice memorial in Dublin and a novel fake story about Facebook apologizing for allowing foreign interference in the referendum (the same

story that appeared in Block One). Within each block, stories were presented in a random order. After each news story, participants were asked 'Do you remember this event?' and could select an option from 'I remember seeing/hearing this', 'I don't remember seeing/hearing this but I remember it happening', 'I don't have a specific memory of this but I believe it happened', 'I remember this differently' and 'I don't remember this'.

Participants were also asked a question about when each event happened, if they did recall the event. Each question was phrased to specifically refer to a part of the story, e.g. 'According to your memory, when were the posters taken down?'; 'According to your memory, when was this apology issued?' etc. Participants could respond by entering text into a response box. This question was included to ensure that any reported memories were genuine memories of the event occurring and not participants reporting that they remembered the story from our previous study.

False Memory Estimate

Participants were asked, 'If some of the stories presented in this study were fabricated (events that never happened), what percentage of participants do you think would falsely remember them as true?' and could answer using a 0%-100% slider.

Participant Feedback

At the end of the study, participants answered questions about their experience of participating, rating their agreement with the following statements on a scale from 1 (Strongly Disagree) to 7 (Strongly Agree): 'I am confident in my understanding of which stories are true and which are fabricated', 'If I were to see these fabricated stories again, I am confident that I would identify them as fake', 'I feel I understand more about the phenomenon of false memories', 'I enjoyed participating in this part of the study', 'I find it interesting to test my susceptibility to fake news'.

Procedure

Participants who completed the first study were asked to sign up for a year-long study of memories of the referendum that would require them to fill out four online surveys (in June 2018, August 2018, November 2018 & May 2019). The 'continuing participants' in the current analyses all signed up at this point, immediately after the first study. One week after the referendum, we distributed the first wave of the longitudinal study to both these original participants, and the newly recruited participants. Each wave of the longitudinal study contained a range of questions related to the referendum, assessing current feelings, flashbulb memories and event memories. The data discussed in this paper comes from the six-month follow-up (wave 3 of 4) that was distributed at the end of November 2018. The newly recruited participants in the present study had therefore answered previous questions about their memory of the referendum, but had not been presented with any fake news stories.

In the six-month follow-up study, news stories were presented to participants at the end of the survey. The introduction to this section read 'Next we will show you four news headlines from the 8th Amendment Referendum. We would like you to indicate if you remember the event and to think about how specific your memory for that event is, using the options provided'. Participants were presented with four news stories each, reporting whether they remembered the event and when they remembered it happening.

After viewing all four stories, participants were told: 'Some participants who undertook this survey were shown fake news stories (stories concerning events that did not happen, entirely fabricated by the researchers). If you think you may have been shown any fake stories, please select any story you believe to be fake below' and could select as many of the four stories as they wished. Participants were also asked to estimate, if any of the stories were fake, how many people would falsely remember them as true.

The next page presented the two fabricated stories and explained that they were entirely fake. Participants were asked, 'If you selected either of these stories as fake, what

was it that helped you to identify the story as fabricated?' and could answer using an open-response text box. The purpose of this question was to find out whether participants relied on their memory of the previous study to identify the fake stories. Participants were also asked to once again estimate what proportion of respondents would have falsely remembered the events as having happened. Finally, participants completed the feedback questions concerning their experience of participating.

Results

All data are available at https://osf.io/9j6ka/. Participants who reported either a specific memory of hearing about an event or a more general memory that it had happened were classified as remembering that event. Those who said they did not remember an event or remembered it differently were classed as not remembering that event. In our analyses comparing memories, those who reported a belief that an event had happened, but not a memory, were excluded (unless otherwise stated). This mirrors the classification used in our original study (Murphy et al., 2019) and is motivated by a need to clearly discriminate between memories and mere beliefs (Wade et al., 2018). Across the entire sample, 22% of participants reported a false memory; 17% reported a false memory for one fabricated story and 5% reported a false memory for both of the fabricated stories shown. A further 14% reported a belief that at least one fabricated event had happened.

Fabricated Stories

Our primary interest was whether continuing participants would have a higher rate of false memories for the previously seen fabricated stories, relative both to their first exposure and to newly recruited participants. A McNemar's test showed that the rate of false memory for the poster story for continuing participants in the follow-up study (14%) was significantly lower than their rate in the first study in May (24%), $X^2(1, N = 509)$ =

18.85, $p < .001$. Continuing participants were also less likely to report a memory for the poster story (14%) than the newly recruited participants (21%), X^2 (1, N = 993) = 7.21, p = .007, V = 0.09. This did not change if we used a broader classification of memory, including both memories and beliefs; continuing participants were less likely to report a memory or belief for the poster story (22%) than newly recruited participants (30%), X^2 (1, N = 1104) = 7.30, p = .007, V = 0.08 (see Figure 1). Those who were presented with the same version of the poster story on both occasions (concerning either the Yes or No campaign) were no more likely to report a false memory for the poster story at follow-up (13%) than those who saw a different version at follow-up (15%), X^2 (1, N = 571) = 0.44, p = .51.

At follow-up, all participants were also presented with a new fabricated story concerning foreign interference in the referendum via Facebook. Continuing participants were significantly less likely to report a false memory for this story (9%) than newly recruited participants (16%), X^2 (1, N = 1010) = 12.46, $p < .001$, V = .11. This was also true if we used a broader classification of memory, including both memories and beliefs (as shown in Figure 2). Continuing participants were less likely to report a memory or belief for the new Facebook story (16%) than newly recruited participants (25%), X^2 (1, N = 1104) = 14.44, p = <.001, V = 0.11

We used a hierarchical binary logistic regression to assess what factors predicted whether continuing participants reported a false memory for an event that they had already been told was fabricated. Here the responses were collapsed across Yes/No versions of the poster story and only remembered/did not remember were included as categories; those who reported a mere belief were excluded. The predictors included were ideological congruency (whether a participant saw the version of the story that was in line with their beliefs; e.g. a Yes supporter seeing a scandal concerning the No campaign), whether a participant saw the same poster story twice, or saw a different version at follow-up, and whether the participant reported a false memory for the poster story in the

first study. The main effects were entered in the first block and all two-way interaction terms were entered in the second block.

The first model was significant, $\chi 2(3, N = 508) = 14.85$, $p = .002$, R^2 (Cox & Snell) = .03, R^2 (Nagelkerke) = .05, and correctly classified 87% of cases. As shown in Table 1, participants were twice as likely to report a false memory for the story when it was in line with their beliefs. They were also twice as likely to report a false memory at follow-up if they had also reported a false memory in the first study. Whether participants saw the same or different version of the story between the first study and follow-up was not a significant predictor. The addition of the interaction terms in the second block improved the model fit (p = .003; model; $\chi 2(6, N = 508) = 28.59$, $p < .001$, R^2 (Cox & Snell) = .06, R^2 (Nagelkerke) = .10, and correctly classified 87% of cases. While having previously reported a false memory and ideological congruency both remained as significant predictors of false memories, there was also a significant interaction between ideological congruency and whether participants saw the same or different version at follow up. Amongst those who saw the same version of the poster story on both occasions, there was no difference in rates of false memory for congruent (13%) and incongruent versions (13%). However, amongst those who saw a different version of the poster story at follow-up, participants were more likely to report a false memory for a congruent story (23%) than an incongruent story (8%).

As in our previous work, the pattern of results was broadly similar when a more liberal threshold for false memories was used and those who reported a mere belief that the event had occurred were included as 'rememberers'. All responses, including those who reported a mere belief, can be seen in Figure 1.

Table 1

Results from a logistic regression analysis of continuing participants' reports of a false memory of the poster story at follow-up, with the following predictors: ideological congruency of the story, whether participants saw the same or a different version of the story, previously reported false memory for the story, and their interactions (n = 508).

Predictors	*B*	*SE b*	Wald	*df*	*p*	*Exp(b)*	95% C.I.
Model One							
Ideological Congruency	0.72	0.27	7.03	1	**.008**	**2.06**	[1.21, 3.51]
Same Version of Story	-0.21	0.27	0.60	1	.439	0.81	[0.48, 1.37]
Previous False Memory	0.78	0.28	7.76	1	**.005**	**2.18**	[1.26, 3.78]
Constant	-2.39	0.27	78.94	1	<.001	0.09	-
Model 2							
Ideological Congruency	2.06	0.56	13.65	1	**<.001**	**7.82**	[2.63, 23.28]
Same Version of Story	0.91	0.58	2.48	1	.115	2.48	[0.80, 7.70]
Previous False Memory	1.43	0.60	5.62	1	**.018**	**4.16**	[1.28, 13.55]
Ideological Congruency*Same Version of Story	-1.84	0.61	8.96	1	**.003**	**0.16**	[0.48, 0.53]
Ideological Congruency*Previous False Memory	-1.07	0.61	3.08	1	.079	0.34	[0.10, 1.13]
Same Version of Story*Previous False Memory	0.27	0.61	0.20	1	.654	1.32	[0.40, 4.30]
Constant	-3.35	0.53	39.57	1	<.001	0.04	-

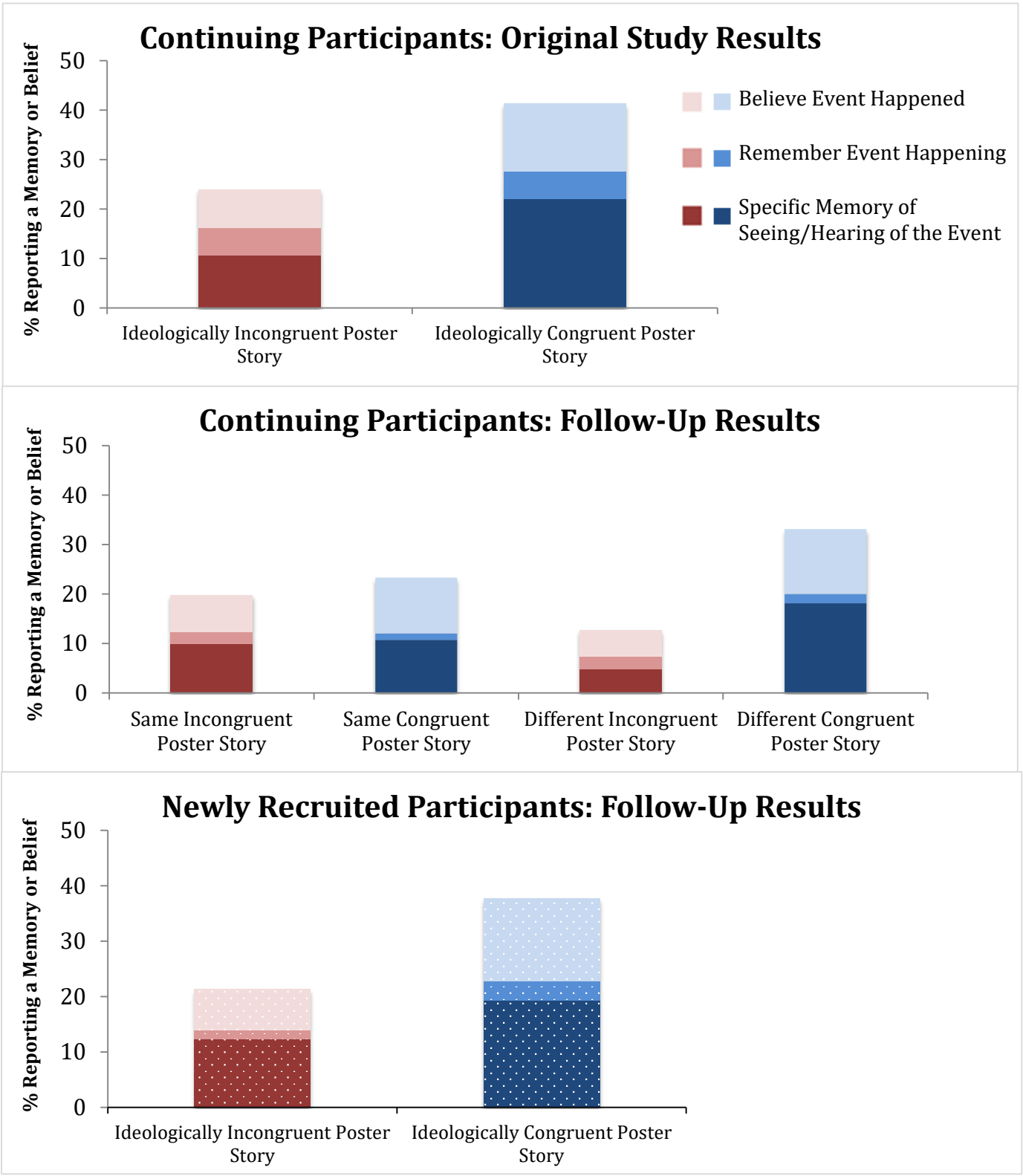


Figure 1: False memories and beliefs for the fabricated poster story amongst continuing participants (solid colour) and newly-recruited participants (dotted pattern). A) Continuing participants (n = 630) in the original study in May saw a fabricated news story concerning illegal campaign posters that was either congruent (blue) or incongruent (red) with their beliefs. B) These same continuing participants (n = 630) completed the follow-up study six months later, in November, where they saw either the same version of the poster story, or the other version. C) Newly recruited participants (n = 476) saw one story in November only. Note that this figure distinguishes between reported memories and mere beliefs that the event had occurred; the regression analysis excludes mere beliefs.

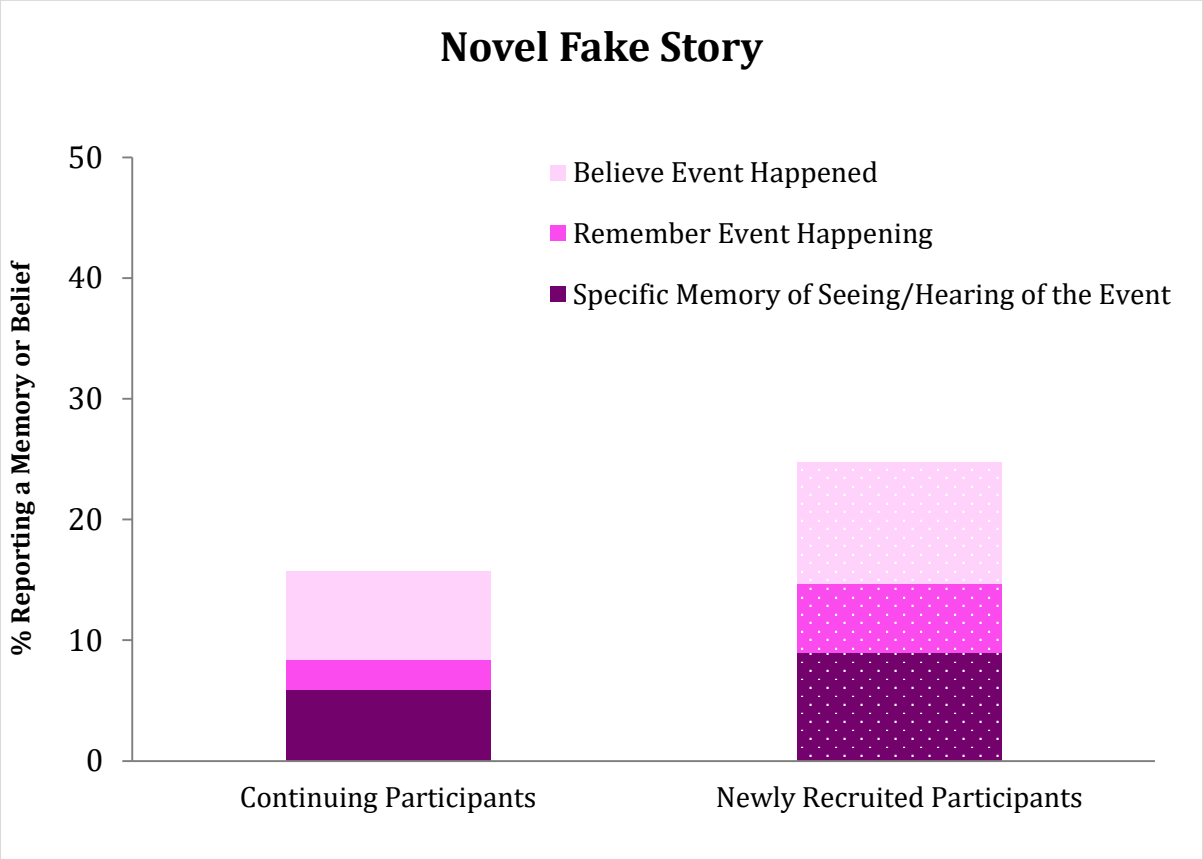


Figure 2: False memories and beliefs for a novel fabricated story concerning Facebook amongst continuing participants (solid colour) and newly-recruited participants (dotted pattern). Continuing participants (n = 630) and newly recruited participants (n = 476) saw this fake story during the follow-up survey.

True Stories

We also compared memories for true stories, to assess whether previous participation impacted recall of real events. In our first study, all participants saw a number of true stories, including either a story about athletes calling for a No vote or actors calling for a Yes vote. A McNemar's test found significantly lower rates of remembering this pair of true stories among continuing participants in the follow-up study (48%) than in the first study in May (72%), X^2 (1, N = 575) = 83.44, p < .001. As shown in Figure 3, continuing participants were shown either the same story they saw in the first study or the story about the other campaign. Continuing participants who saw the same story twice were more likely to report a memory for the event at follow-up (70%) than those who saw a different story (44%), X^2 (1, N = 529) = 35.87, p < .001, V = .26. This was true for both stories; Actors for Yes X^2 (1, N = 252) = 15.53, p < .001, V = .25 and Athletes for No X^2 (1, N = 277) = 18.26, p < .001, V = .26. This suggests that participants had not entirely forgotten their participation in the first study and there was some remaining memory of the stories presented at that time.

At follow-up, all participants were presented with one of two new true stories concerning either a memorial that was established in Dublin the day after the referendum or comments made by Pope Francis concerning the referendum. Continuing participants were less likely to report a memory for the Pope's comments (15%) than newly recruited participants (25%), X^2 (1, N = 430) = 6.85, p = .009, V = .13, but equally likely to report a memory for the memorial story (79%) as newly recruited participants (78%), X^2 (1, N = 477) = 0.01, p = .907, V = .01. It is possible that the continuing participants were more suspicious of the Pope Francis story, due to their previous exposure to our fake news paradigm. This did not seem to affect rates of memory for the memorial story, perhaps because it was more widely publicized and better remembered

(as evidenced by the higher recall rate for the memorial story across both the

continuing and newly recruited participants).

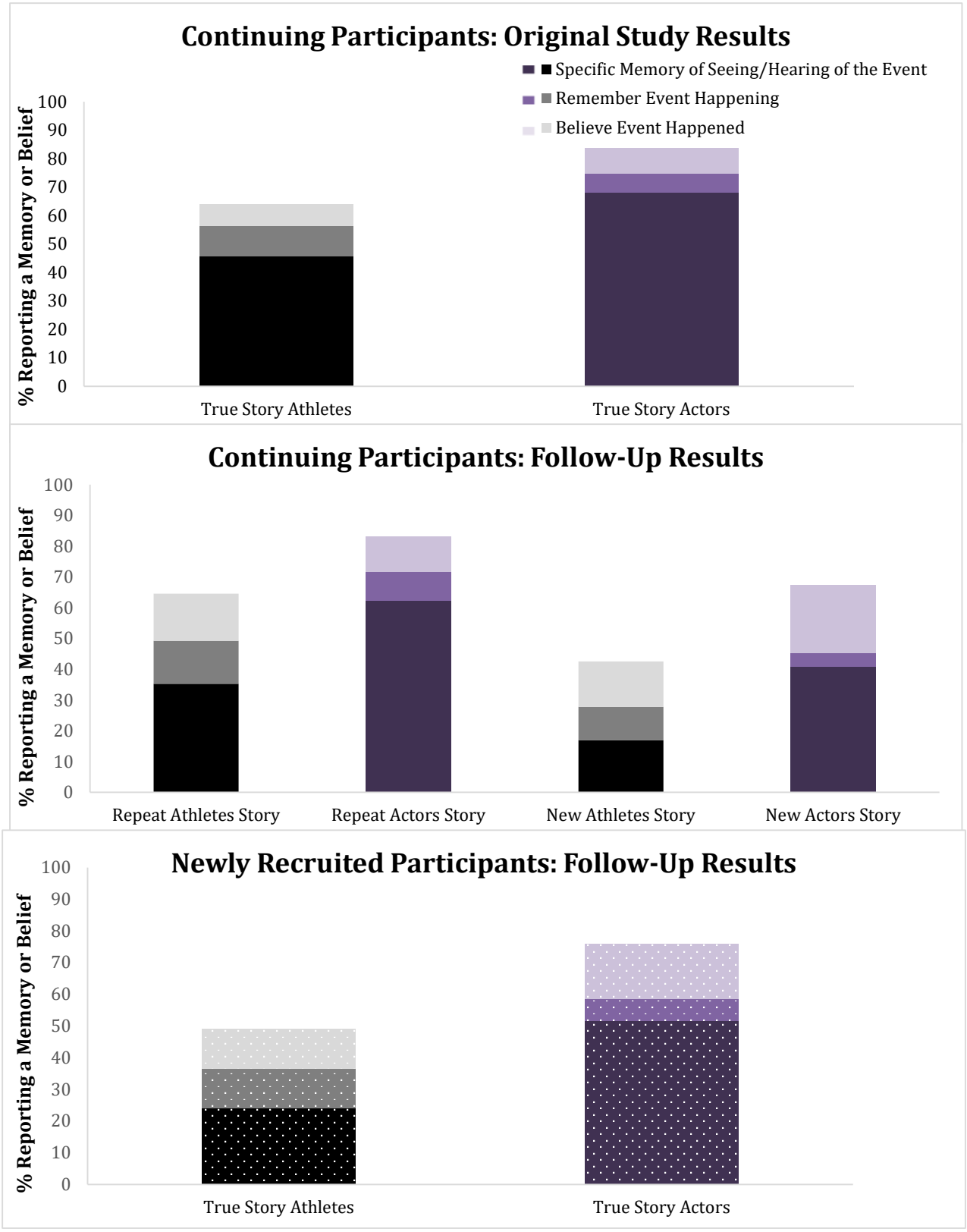


Figure 3: Reported memories and beliefs for true events amongst continuing participants (solid colour) and newly-recruited participants (dotted pattern). Continuing participants were shown one of two true stories in the first study (concerning athletes (in black) or actors (in purple). In the follow-up study, they either saw the same story again (e.g. the same athletes story on both occasions) or the other story (the athletes story when they previously saw the actors story). Newly recruited participants saw one of the two stories at follow-up for the first time.

Post-Warning Fake News Detection

Participants were told they may have been presented with some fake news stories and were asked to select any stories they believed to be fabricated. Continuing participants selected a similar number of stories (M = 1.83, SD = 0.77) as newly recruited participants (M = 1.74, SD = 0.78), t (1102) = 1.90, p = .058, d = 0.12.

When prompted, continuing participants were equally likely to identify the poster story as fake at follow-up (66%) as newly recruited participants (62%), X^2 (1, N = 1104) = 1.80, p = .179. Continuing participants were more likely to identify the poster story as fabricated if it was incongruent with their beliefs – that is, if it reflected badly on the campaign they identified with - (76%), than if the story was congruent with their beliefs (57%), X^2 (1, N = 629) = 23.09, p < .001, V = .19. This effect was also evident for newly recruited participants. They too were more likely to identify the poster story as fabricated if it was incongruent with their beliefs (70%), relative to congruent (55%), X^2 (1, N = 470) = 12.61, p < .001, V = .16. Continuing participants were equally likely to identify the novel fake Facebook story as fake (78%) as newly recruited participants (75%), X^2 (1, N = 1104) = 1.44, p = .230.

When asked why they selected the stories as fake, 16% of continuing participants did not respond to this question. Of those who gave an answer, just 5% mentioned recognizing the story from the previous survey. Responses coded as mentions of the previous survey included; '*I remembered the posters one from the first study as I think I chose that one as being true, and was struck by the fact I could be able to falsely remember things', 'I remember the last time I did this survey that you identified certain stories as fake', 'I think the poster example was flagged as being fake in your previous survey so I remember that but I think initially I thought it was true'.*

Others mentioned having heard the story before but not as part of the study, possibly suggesting source confusion; *'I remember hearing about this fake news story doing the rounds', 'I wasn't entirely sure but I had a feeling that this was a rumor started during the campaign but that it had proved false*'. Others mentioned an increased suspicion of fake news due to taking part in the study; *'I didn't remember them plus this study has made me more aware of fake news', 'I do remember there being controversies about pro-life posters, and some had to be taken down, but I didn't recollect a link to American donors. I had to think about this though and work it through. It felt familiar and possibly true, probably because it resembled a story that I had heard. I had to work it out in my head (probably because of the nature of the survey) but if I saw it in passing I doubt I'd have given it as much thought and possibly accepted it at face value.'*

Participant Knowledge

Before being debriefed, participants were asked to estimate what percentage of participants they would expect to report a false memory if they were exposed to fabricated news events. Continuing participants gave an estimate (M = 52.96%, SD = 17.42) that was almost identical to newly recruited participants (M = 52.92%, SD = 17.86), t (1079) = 0.04, p = .972. After being debriefed and told which stories were fabricated and asked to give an estimate again, both continuing participants (M = 56.38%, SD = 18.75) and newly recruited participants (M = 56.35%, SD = 19.69) increased their rating by a similar amount, t (1075) = 0.01, p = .992.

Merging the two samples, those who reported a false memory for either the poster or Facebook story gave similar pre-debrief estimates (M = 55.09%, SD = 18.22) to participants who did not report a false memory (M = 52.31%, SD = 17.42), t (973) = 1.85, p = .065. Post-debrief, there was a larger difference between those

who reported a false memory (M = 62.18%, SD = 17.36) and those who did not (M = 54.79%, SD = 19.43), t (260.85) = 4.92, $p < .001$, d = 0.40, suggesting that those who were told that they had reported a false memory revised their estimates upwards. We also assessed whether those who had reported a false memory in our first study six months previously (and been debriefed) would give higher estimates during the follow-up. We found no difference in the pre-debrief ratings of continuing participants who had previously reported a false memory (M = 54.26%, SD = 16.38) and those who did not (M = 51.87%, SD = 17.68), t (545) = 1.38, p = .170.

At the end of the study, participants were asked about their experience, with all questions answered on a 1-7 scale, where 1 indicated strong disagreement and 7 indicated strong agreement. Participants reported understanding which stories were true and which were fabricated (M = 5.09, SD = 1.44), being confident that they would be able to identify the stories as fake if they saw them again (M = 5.37, SD = 1.41), that they enjoyed participating in this part of the study (M = 5.68, SD = 1.09), that they had learned more about false memories (M = 5.08, SD = 1.18) and that they found it interesting to test their susceptibility to fake news (M = 5.98, SD = 1.03). There were no significant differences between continuing and newly recruited participants on any of these measures (all $p > .05$).

Discussion

The current study followed up participants six months after they completed our experiment where we exposed them to fabricated political news stories. We compared their rates of false memories at follow-up both to their rates in the original study and to newly recruited participants. Across the entire sample, 36% of participants reported a memory or belief for at least one fabricated event. As in our first study (Murphy et al., 2019), ideological congruency was a strong predictor of

participants reporting a false memory for the fabricated poster story and of identifying the story as fake after being warned of the purpose of the study. This again demonstrates that false memories are likely to form in line with existing beliefs.

Though research has shown that misinformation can be difficult to correct (Lewandowsky et al., 2012), continuing participants were in fact slightly less likely to fall for the fake news story concerning illegal posters than newly recruited participants. This was true whether we assessed reported memories only, or included mere beliefs in a less conservative classification of memory (Wade et al., 2018). Amongst continuing participants, there was no increase in false memories for those who saw the same version of the poster story (e.g. concerning the Yes campaign on both occasions) compared to those who saw a different version (e.g. the No campaign story in the first study but the Yes campaign version at follow-up). This is in contrast to the illusory truth effect literature that might predict increased belief after repetition (Dechêne et al., 2010). The qualitative responses suggest that a small minority of participants (5%) remembered seeing the events in our previous study, or at least explicitly stated that that they did. Our primary concern was that participants from our study would be more vulnerable to falling for the same fabricated story in future, but we have found no evidence of that.

In addition, there was a significant interaction between seeing the same version of the poster story and ideological congruency. Participants who saw the same version of the poster story on both occasions demonstrated no effect of congruency. In contrast, those who saw a different version of the poster story at follow-up were significantly more likely to report a false memory for a story that was in line with their beliefs, a pattern similar to that observed amongst newly

recruited participants at follow-up. This is in line with what we observed in our original study (Murphy et al., 2019) and has been demonstrated in other contexts (Frenda et al., 2013). It appears that participating in the study (and being debriefed) eliminates the congruency effect if exposed to the story again, but only if the exact same story is presented on both occasions. These findings speak to the narrow effect of previous participation, which did not generalise even to the exact same story presented about the opposite campaign. There is mixed evidence concerning the effect of attitude congruency on misinformation retraction (Ecker et al., 2014), and further study is warranted.

Interestingly, continuing participants were significantly less likely to report a false memory for the novel fake story presented at follow-up (concerning foreign interference in the referendum via Facebook). This may be due to the fact that they were aware we were investigating fake news and thus were more suspicious and better at detecting the fabricated event. Indeed, once we alerted participants to the fact that they may have seen fake news stories, newly recruited participants were as likely as continuing participants to select the story as fabricated. This further supports the idea that the naiveté of the new recruits drove the difference in initial reported memories for the story. Research has shown that warnings can sometimes reduce false memory rates (Ecker et al., 2010; Gerrie & Garry, 2011). It is difficult to draw firm conclusions based on a single item, but this warrants further investigation, as it may evidence a benefit of participating in a false memory study; greater suspicion of fake news stories. Future research might also investigate whether this reduced tendency to report false memories would generalize to news presented from other sources, i.e. whether after participating in a false memory

study, participants might be more suspicious of fake news presented on a news site or in another study not associated with the original researchers.

We observed a clearer effect of previous participation for the true news stories. We directly compared memories for two news stories concerning either athletes or actors endorsing either side of the campaign. We found that when participants were presented with the same story at follow-up that they had seen in our first study (e.g. athletes on both occasions), they were more likely to report a memory for that event than if they saw two different news stories. This suggests that even though there was quite a long time period between participation and follow-up, there was some residual effect of participation and some memory of the study was intact. Of the two novel (non-repeated) true stories presented in this study, continuing participants were less likely to report a memory for one of the events, relative to newly-recruited participants. It may be that previous participation increased suspicion or encouraged stricter source-monitoring, as with the novel fake story, but this warrants further investigation in future research.

Finally, we assessed whether there were any benefits from participating in such a study. We found participants generally enjoyed taking part in the study and found it interesting. This is an important finding, as previous studies have found mixed reactions to deception in psychology research (Hertwig & Ortmann, 2008). There was no evidence that participants became more aware of the phenomenon of false memories or estimated them to be more common. This was difficult to assess due to a ceiling effect – both new and continuing participants gave high estimates of false memory rates (> 50%) and those rates increased if they found out that they themselves had reported a false memory. This average estimate was significantly higher than the rate of false memory we observed in the current study, as well as

that observed in most false memory studies (Loftus, 2005). Given the ongoing debate concerning public perceptions of memory reliability and whether non-psychologists perceive memory to work 'like a video camera' (Brewin et al., 2019; Otgaar et al, 2020; Simons & Chabris, 2011), this high estimate is noteworthy. Participants, both newly recruited and those who had previously participated in a false memory study, may actually overestimate the likelihood of false memories forming in response to fake news articles.

It is important to note the long delay between the first study and follow-up. Our follow-up period of six months might have resulted in participants simply forgetting about the study and thus any effects that may have been evident may have been lost. The qualitative evidence supports this, as very few participants reported remembering the story from the first study. However, we would note that there was a clear effect of previous participation in the case of the true events, despite the long delay.

A sizeable minority of continuing participants reported a false memory for the fabricated event (14%), despite being debriefed after the original study. Those who had reported a false memory in the first study were also more likely to report a false memory in the current study. Future research might examine the characteristics of these false memories reported by continuing participants. It is possible that the debrief was entirely forgotten or not believed to begin with, as some studies have noted that participants can reject the debriefing and insist the fabricated event did happen (Otgaar et al., 2009). Alternatively, participants may have been reporting non-believed memories at follow-up (Mazzoni et al., 2010; Otgaar et al., 2014). That is, some who reported a false memory of the fabricated event after debriefing may no longer have believed that the event actually occurred.

If so, this would suggest that debriefing was partially successful for these participants. Thus, future research might assess confidence and belief in these memories at follow-up to further examine any effects of debriefing.

In conclusion, our study suggests that participating in our fake news study did not negatively impact participants' memories. Relative to new recruits, continuing participants were less likely to report a false memory for a story that they had been previously exposed to and less likely to report a false memory for a novel fake news story. Participants reported enjoying the experiments and feeling confident that they understood which stories were fake after they had been debriefed. Our study does not suggest that exposure to fabricated political events during these types of misinformation studies will negatively affect vulnerability to fake news in the medium-to-long term future.

References

Brewin, C. R., Li, H., Ntarantana, V., Unsworth, C., & McNeilis, J. (2019). Is the public understanding of memory prone to widespread "myths"?. *Journal of Experimental Psychology: General*, *148*(12), 2245-2257

British Psychological Society. (2014). *Code of human research ethics*. British Psychological Society.

Cabalo, D. G., Ianì, F., Bilge, A. R., & Mazzoni, G. (2020). Memory distortions: When suggestions cannot be easily ignored. *Applied Cognitive Psychology*, *34*(1), 106-118.

Clark, A., Nash, R. A., Fincham, G., & Mazzoni, G. (2012). Creating non-believed memories for recent autobiographical events. *PLoS One*, *7*(3), e32998.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238-257.

De Keersmaecker, J. & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, *65*, 107-110.

Ecker, U. K., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & cognition*, *42*(2), 292-304.

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087-1100.

Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, *49*(2), 280-286.

Gerrie, M. P., & Garry, M. (2011). Warnings reduce false memories for missing aspects of events. *Experimental Psychology*, *58*(3), 207.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of verbal learning and verbal behavior*, *16*(1), 107

Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, *18*(1), 59-92.112.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source

monitoring. *Psychological bulletin*, *114*(1), 3-28.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1420–1436

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, *13*(3), 106-131.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & memory*, *12*(4), 361-366.

Mazzoni, G., Scoboria, A., & Harvey, L. (2010). Nonbelieved memories. *Psychological Science*, *21*(9), 1334-1340.

Murphy, G., Loftus, E. F., Grady, R. H., Levine, L. J., & Greene, C. M. (2019). False Memories for Fake News During Ireland's Abortion Referendum. *Psychological science*, *30*(10), 1449-1459.

Nash, Robert A. "Changing beliefs about past public events with believable and unbelievable doctored photographs." *Memory* 26.4 (2018): 439-450.

Otgaar, H., Bücken, C., Bogaard, G., Wade, K. A., Hopwood, A. R., Scoboria, A., & Howe, M. L. (2019). Nonbelieved Memories in The False Memory Archive. *Journal of Applied Research in Memory and Cognition*, *8*(4), 429-438.

Otgaar, H., Candel, I., Merckelbach, H., & Wade, K. A. (2009). Abducted by a UFO: Prevalence information affects young children's false memories for an implausible event. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(1), 115-125.

Otgaar, H., Scoboria, A., & Mazzoni, G. (2014). On the existence and implications of nonbelieved memories. *Current Directions in Psychological Science*, *23*(5), 349-354.

Otgaar, H., Scoboria, A., & Smeets, T. (2013). Experimentally evoking nonbelieved memories for childhood events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 717.

Otgaar, H., Wang, J., Howe, M. L., Lilienfeld, S., Loftus, E., Lynn, S., ... & Patihis, L.

(2020). Belief in Unconscious Repressed Memory Is Widespread: A Comment on Brewin, Li, Ntarantana, Unsworth, and McNeilis (2019). *Journal of Experimental Psychology: General*.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology. General*, *147*(12), 1865.

Sacchi, D. L., Agnoli, F., & Loftus, E. F. (2007). Changing history: Doctored photographs affect memory for past public events. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *21*(8), 1005-1022.

Scoboria, A., Jackson, D. L., Talarico, J., Hanczakowski, M., Wysman, L., & Mazzoni, G. (2014). The role of belief in occurrence within autobiographical memory. *Journal of Experimental Psychology: General*, *143*(3), 1242.

Scoboria, A., Mazzoni, G., Kirsch, I., & Relyea, M. (2004). Plausibility and belief in autobiographical memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(7), 791-807.

Scoboria, A., & Talarico, J. M. (2013). Indirect cueing elicits distinct types of autobiographical event representations. *Consciousness and Cognition*, *22*(4), 1495-1509.

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the US population. *PloS one*, *6*(8).

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*(3), 460-480.

Unkelbach, C., & Greifeneder, R. (2013). A general model of fl uency effects in judgment and decision making. In *The experience of thinking* (pp. 21-42). Psychology Press.

Wade, K. A., Garry, M., & Pezdek, K. (2018). Deconstructing rich false memories of committing crime: Commentary on Shaw and Porter (2015). *Psychological Science*, *29*(3), 471-476.

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423-441.

Disclosure of Interest

The authors report no conflict of interest.

Data Availability Statement

The data is available at the Open Science Framework, here: https://osf.io/9j6ka/.

Appendix A: News Stories Presented in Follow-Up Study

True & Novel



During his official visit to Ireland, Pope Francis questioned whether a "materialistic 'throwaway culture'" has made people "increasingly indifferent to the poor and to the most defenseless members of our human family, including the unborn, deprived of the very right to life".

Used with permission of author, 'Catholic Church England & Wales'
https://www.flickr.com/photos/catholicism/8723854050



After voting in the eighth amendment referendum, many left flowers and messages at a mural of Savita Halappanavar in Dublin.

Used with permission of author, 'lusciousblopster'
https://www.flickr.com/photos/lusciousblopster/42482245331

True & Previously Seen



A number of Irish actors including Saoirse Ronan, Cillian Murphy & Liam Cunningham featured in a Together For Yes Youtube video titled 'In Ireland Today'. The celebrities called for a Yes vote in the referendum.

(Screenshot taken from Youtube)



A number of GAA stars called for a No vote in the 8th amendment referendum. The group, including Aoife Cassidy (Derry, All Ireland Camogie winning Captain) and Micky Harte (Tyrone Manager) released a number of Youtube videos detailing why they are voting No.

(Screenshot taken from Youtube)

Fabricated & Novel



Facebook offered a formal apology to Irish voters after evidence emerged that Irish voter profiles had been sold to Russian campaigners seeking to influence the referendum outcome in Ireland.

Photo used with permission of author, Anthony Quintato.
https://commons.wikimedia.org/wiki/File:Mark_Zuckerberg_F8_2018_Keynote_(41793468182).jpg

Fabricated & Previously Seen

(participants saw either the Yes or No version of this story)



The Together for Yes campaign was forced to destroy 25,000 campaign posters after evidence emerged that the posters were bought using funding received from American pro-choice lobbyists.



The Save the 8th campaign was forced to destroy 25,000 campaign posters after evidence emerged that the posters were bought using funding received from American pro-life lobbyists.