Specific Retractions May Counteract Personal Bias in Reducing Misinformation

Jordan S. Dimmick '18 and Robert T. Herdegen III

Department of Psychology, Hampden-Sydney College, Hampden-Sydney, VA 23943

ABSTRACT

The misinformation effect occurs when people continue to rely on older information even when newer information has been presented. This is often seen in situations involving newspaper articles or television reports, when a previous report is corrected in a later article. Factors shown to influence the effectiveness of follow-up reports have been racial bias and the retraction type. Specific retractions (in which the original incorrect information is repeated before the correction) are more effective than nonspecific retractions (Ecker, Lewandowsky, and Tang (2010)), and personal bias causes interference with the retraction if the bias runs counter to the follow up report (Ecker, Lewandowsky, Fenton, and Martin (2014)). This study aimed to see if retraction type and bias interact such that high bias undermines the effect of even a specific retraction. This was done by using an implicit association test (IAT) to group participants as either high or low in ethnic bias. Then, participants read a newspaper-type report about a crime involving a person identified of Arab ethnicity, and a subsequent retraction. A specific retraction led to more accurate recall of retracted information, but bias did not affect recall accuracy. The results of this study suggest that specific retractions are more effective than non-specific retractions, and may even counter the effects of bias in recall of information. These results highlight the importance of framing retractions in reports in a way that resembles a specific retraction, by drawing in the older information before following up with a subsequent retraction.

INTRODUCTION

The misinformation effect, or the idea that people often cite outdated information, is a problem that presents itself in a variety of real-life situations, such as in articles that are written to retract information in a previously published report. As Ecker, Lewandowsky, Chang, and Pillai (2014) state, misinformation effects "arise when misinformation is automatically retrieved and strategic monitoring fails" (p. 323). This suggests that misinformation can occur in a variety of contexts, ranging from details of a news report to the names of items on a list. Any situation that involves the recall of stored memory is susceptible to misinformation. Frequently, however, attempts to correct the information only result in a strengthening of the misinformation. Researchers are faced with two questions: what factors increase how a subject encodes and relies on false information, and what reduces an individual's reliance on the false information?

The misinformation effect occurs when past memories are influenced by false information, even though other information is readily available. In a review of research and theory on the misinformation effect, Rapp (2016) lists memory errors related to misleading encoded information, retracted false reports, and misunderstandings about scientific concepts as all stemming from the misinformation effect. Rapp highlights the results of numerous studies in which participants who read a realistic story that contained false information ("the Alps separate Asia from Europe") were more likely to rely on that false information rather than previous knowledge. Even when participants were aware that some of the information they would come across would be inaccurate, they were more likely to rely on it shortly after encoding it. After reading from a list that occasionally contained false assertions ("George Washington was not the first elected president of the United States"), participants displayed difficulty in gauging the validity of these statements. Furthermore, participants who were warned that some of the information they would be reading would be inaccurate showed little to no reduction in the reliance on the falsehoods. In a real world setting, this could range from misconceptions about scientific discoveries to misidentification of the information presented in an article or news headline. This suggests that even if the individual is forewarned that the information they are reading might not be correct, they will still remember and potentially use the information they read.

Lewandowsky, Ecker, Seifert, Schwarz, and Cook (2012) describe the four major problems associated with the misinformation effect: the Continued Influence Effect, the Overkill Backfire Effect, the Worldview Backfire Effect, and the Familiarity Backfire Effect. The Continued Influence Effect causes people to be influenced by, and as such rely on, misinformation even after a retraction of the false information. Lewandowsky et al. (2012) write that including an alternative account when retracting false information oftentimes increases the success rate of the retraction, as the specific alternative account allows for a new set of information to replace the misinformation. The Overkill Backfire Effect causes simple falsehoods to be more attractive to readers than complicated retractions. If the retraction confuses readers, they will be more likely to dismiss the retraction, and as

such continue to hold onto the misinformation they were given earlier. To prevent the Overkill Backfire Effect from occurring it is best to keep retractions simple, to avoid confusion. The Worldview Backfire Effect shows that evidence that threatens an individual's worldview will cause an increase in their reliance on previously held beliefs. An example of this is a study mentioned by Lewandowsky et al. (2012) that found that Republicans and Democrats differed significantly in which retractions were more effective when the retractions were related to politically driven facts. For example, when articles related to the presence of Weapons of Mass Destruction (WMDS) in Iraq were read by groups of Republicans and Democrats, the study found a significant difference in the retention rate of the retraction statement, which stated that no WMDs were found in Iraq. This difference was found to be that Democrats were more likely to recall the retraction, while Republicans were less likely, as the retraction contrasted their worldview. Likewise, when reading an article that described the consequences of rising oil prices, Republicans recalled the information in the retraction statement at a significantly higher rate than Democrats. Lastly, the Familiarity Backfire Effect occurs when attempts to correct misinformation by repeating the false information in the retraction statement strengthens the usage of the misinformation in further events, due to the repetition of the falsehoods.

Ecker, Lewandowsky, and Tang (2010) further studied which methods of retracting false information were most effective. By comparing different forms of retractions and warnings (specific versus general warning, no retraction versus general retraction versus retraction with a plausible alternative), they found that "the specific warning and the provision of an alternative account reduced reliance on misinformation" (p. 1094). Furthermore, Ecker et al. (2010) describe research by Eakin, Schreiber, and Sergent-Marshall (2003) who "found immediate post misinformation warnings that explicitly identified the piece of misinformation as effective as warnings that did not specify the misleading information" (Ecker et al., 2010, p. 1095). The reason Ecker et al. (2010) give for the explicit warnings being more effective is that the specific explicit warnings, by giving concrete examples, affected both the encoding and retrieval process, allowing for participants to be less affected by the Continued Influence Effect. In a study that paired retraction statements with either subtle or explicit reminders of the misinformation being retracted, Ecker, Hogan, and Lewandowsky (2017), found that "corrections were more effective when they explicitly repeated the myth" (p. 185), suggesting the exact opposite of the Familiarity Backfire Effect. Swire, Ecker, and Lewandowsky (2017), however, warn the reader about the potential for experiencing the Familiarity Backfire Effect, stating that "if a participant is unable to correctly recall the correction of a myth because of the forgetting that primarily affects strategic memory processes, the familiarity of the myth… could lead to the myth being inaccurately accepted as true" (p. 3).

The above findings suggest that the Familiarity Backfire Effect is a factor to consider when looking at the misinformation effect, but at the same time, explicit restating of the myth when appropriate warnings that the information has the potential to be false is an effective way at dampening or even nullifying the Familiarity Backfire effect. However, one thing to consider with these findings is how they relate to real world examples. Frequently, misinformation is spread through news reports on television or in newspaper articles. In these cases, retractions are rare, and the chance of an individual being warned about the potential for misinformation before they take in the details of the report is not likely. As such, in real world cases in which warnings before the reading of the incorrect material is uncommon, the effect of specific retractions is much weaker than in most of the experimental cases.

Personal factors also contribute to the susceptibility of the misinformation effect within individuals. For example, Huff and Umanath (2017) found that age played a role, in that older (aged 65-87) adults were less likely to be misled by false information than young adults (18-22). However, strong warnings that what they were about to read would contain some falsehoods caused the overall performance of the young adults increased significantly, scoring similar results as older adults in the same conditions. Another study looked at racial discriminatory beliefs. These beliefs would theoretically interact like the Worldview Backfire effect, in that the individual would be less likely to encode and use retraction statements that retract 'facts' that support their worldview. Ecker, Lewandowsky, Fenton, and Martin (2014) examined the impact of a person's personal beliefs on the effectiveness of retractions that were directly connected to those beliefs. In their study, participant's racial prejudice was measured using their ATIA (Attitudes Toward Indigenous Australians) scores. Using these scores, participants were divided into Low prejudice and High prejudice groups. Next, all participants were given fictitious news reports that did or did not include a retraction on the race of the criminal written in the news article. They found preexisting attitudes, such as racial bias, influenced people's use of information related to those attitudes. In other words, "Preexisting attitudes codetermine people's reliance on (mis)information" (p. 303). Taken together, these findings suggest that individual factors such as age and racial bias can and do play a

role in the effectiveness of retracting false information.

Ecker, Lewandowsky, Fenton, and Martin (2014) found that personal racial biases lowered the effectiveness of retractions of misinformation directly related to those biases, but they did not take into account different kinds of retraction statements. Specifically, their study did not include the difference that a specific or nonspecific retraction can have on the reduction of reliance on misinformation. As such, the question arises, would racial biases have an impact on the effectiveness of different types of retraction statements?

The purpose of the present study is to determine whether the effectiveness of a specific retraction is dependent upon personal bias. In light of Ecker, Hogan, and Lewandowsky's (2017) findings, it is expected that participants that receive specific retractions will better remember the corrected information than participants who receive the nonspecific retraction. This study also predicts an interaction between bias and retraction type. If one occurs, Ecker, Lewandowsky, Fenton, and Martin's (2014) study would suggest that participants high in bias, regardless of retraction type, would be less likely to recall information that counters their worldviews. Combined with Ecker, Hogan, and Lewandowsky's (2017) findings, the predicted results would suggest that participants low in bias and given the specific retraction would correctly recall more in a misinformation-based experiment than participants low in bias and given a non-specific retraction, as well as participants of high bias.

METHODS

Participants

The study included 33 traditionally-aged male college students recruited from psychology classes at a small liberal arts college. Seven participants were excluded from data analysis due to their not returning to the second part of the experiment, and one participant was excluded due to answering too few questions on the questionnaire. The participants were selected through a convenience sample. Some participants were incentivized with class credit if they participated in the study.

Materials

An Arab-Muslim - Caucasian American version of the IAT (Greenwald, Poehlman, Uhlmann, & Banjali, 2009; Roberts, Neate, & Gierasch, 2017; Howell & Ratliff, 2017) was created with the DirectRT program. This test compared the speed at which the participants paired words associated with "good" or "bad" with faces associated with "Arab-Muslim" and "Caucasian American".

Participants were given six news-type articles to read from in a quasi-random order. These articles consisted of three pairs of related articles in which one article is an initial news report and the second article is a follow-up with additional information related to information in the first article. The first of these three pairings described a car chase, but the information from the second article did not retract the statement from the first article. The second of these pairings was a story about a wildfire, with the crucial piece of evidence being how the fire started ("The fire had been deliberately lit") with the second article retracting it with the statement ("After a full investigation and review of witness reports, authorities have concluded that the fire was set off by lightning strikes.") The last article pairing was ethnically-charged in nature. The first article contained a piece of evidence related to the ethnicity of the suspect in an incident ("the suspect whose luggage contained the dangerous items is believed to have been a citizen from Iraq.") with the second article retracting the statement ("Upon completion of a full investigation, authorities have concluded that the suspect with the suspicious luggage was an American-born citizen.").

Participants were given a 22-item questionnaire that reflected on information included in the six articles. All questions were open-ended so as not to 'remind' participants of the answers if they were displayed in a multiple-choice type format. Eight of these questions were about an ethically-charged scenario, and three of these eight questions focused on the ethnic identity of the suspect in the scenario.

Design

This study used a 2 x 2 factorial design, with prejudice level (high, low), and retraction type (specific, nonspecific) treated as between subjects factors. Participants were identified as either high or low bias based on their performance on the IAT, and then randomly assigned to the retraction type.

Procedure

Participants were recruited in psychology classes and reported individually to a computer lab in the psychology building. Using a mild deception, the participants were informed that the test was studying the effects of memory and comparing performance on memory-based tasks with the performance on other mental, non-memory, based tasks. This was done to avoid drawing the participant's attention to the ethnic-aspects of the key articles. After signing the consent form, participants completed the Arab-Muslim - Caucasian American IAT. After all participants completed the IAT, their scores divided them into either high or low bias. The participant's scores were calculated by averaging the reaction times of each item in each section in which the

participant had to press the same key for both a racial ethnicity (Caucasian-American / Arab-Muslim), and a descriptor (good / bad). Once two averages were received (one where Arab-Muslim and 'good' were on the same key and one where Arab-Muslim and 'bad' were on the same key), the difference was taken between those scores. The larger the difference in the two scores, the greater the implicit bias of the individual.

After completing the IAT, participants were allowed to leave, but were informed about a second, "unrelated" study that would be happening during the next week, and that participation in that study would be greatly appreciated. For students receiving class credit, they were informed that this second study was a way to earn more. Participants returning for the second part of the study reported individually to a psychology classroom and were given the six articles in a quasi-random order. The order was controlled in that the paired articles always had the opening article precede the retraction, though the pairings were not guaranteed to be presented one after the other. Additionally, the ethnically-charged article's retraction statement was always the fifth article that the participant read. Each participant had a minute and fifteen seconds to read each article, and after reading all six, participants worked on a sudoku puzzle for ten minutes as an interpolated task.

After the ten minutes passed, participants were given the test to measure recall accuracy for information in the articles. Tests were scored in that a question could only be correct or incorrect; no partial scores were given. An answer key was made ahead of time to avoid researcher bias.

RESULTS

The dependent variable in this experiment was the number of questions related to the ethnic aspect of the articles the participants correctly guessed. To properly account for this, three sets of analyses were run.

The results of the experiment were analyzed with a series of 2 x 2 factorial ANOVAs, with retraction type (specific, non-specific) and bias (high, low) treated as between subjects factors. A test was conducted to look at the group differences for total number of questions answered correctly throughout the entire questionnaire. There were neither significant main effects for retraction type and bias, nor a significant retraction type x bias interaction, meaning that no group of participants performed better due to personal differences, such as one group possessing a naturally better memory capacity.

A second analysis was run for the total number of 'key' (ethnically-charged) questions each participant answered correctly. The analysis did not reveal significant main effects for implicit bias nor for retraction type, not a significant bias x retraction type interaction. However, an examination of the results suggests that individuals of a higher implicit bias were answering more of the key questions correctly, on average, than those of low bias. This finding, though not significant, is inconsistent with earlier findings reported by Ecker, Lewandowsky, Fenton, and Martin (2014), who saw a significant effect in the opposite direction.

To correct for the possibility of differences between groups in general memory for information in the critical story, an adjusted accuracy score was computed. This score was derived by taking the total number of ethnically-based questions correct (out of three), and dividing that number by the total number of questions answered correctly for the airport scenario (out of eight). Fig. 1 displays the mean adjusted accuracy scores as a function of retraction type, for the high and low bias participants. This figure shows that the specific retraction group, regardless of bias, possessed a higher mean adjusted accuracy score than the non-specific retraction group. It also appears that adjusted accuracy scores were higher for the high bias than the low bias participants. A factorial ANOVA on these data revealed a main effect for retraction type, $F(1, 21) = 6.425$, $p < .05$, but not for bias type nor the retraction type x bias interaction. The participants who received the specific retraction in the third article pairing had a greater proportion (.41) of the ethnically-based questions as their correct questions for the airport report as opposed to those who received the non-specific retraction (.27). Moreover, the apparent difference between high and low bias participants in this measure was not statistically reliable.

DISCUSSION

This study expected to find a main effect for retraction type, in that participants who received a specific retraction would correctly answer significantly more questions than those who received the non-specific retraction. Additionally, the study expected an interaction between bias level and retraction type, such that high bias would not be affected by the type of retraction, but low bias would perform better on the recall test when given the specific retraction. The results found supported the first hypothesis, as participants who received the specific retraction had a significantly greater proportion of their correct airport answers being related to the racial identity of the suspect. However, the lack of an interaction effect means the second hypothesis failed to reject the null hypothesis.

The findings in this study support the findings of Ecker et al. (2010) that specific retractions are more effective in making the participants stop relying on the original article and remember the follow-up

report. However, the current findings differ from those of Ecker et al. (2014), as bias did not play a significant role in the correct recall of the article's contents. Despite the lack of a significant difference, however, there is a notable trend for bias in that individuals with higher bias had a higher adjusted accuracy score than individuals of lower bias. This difference, although not significant, runs counter to the findings of Ecker et al. (2014). One possibility for this trend is that individuals of higher racial bias actually became more sensitive to the racial aspects of the articles, and as such were more correctly able to identify those aspects when later quizzed about the article.

One limitation in the study is that the lack of participants forced the study design to change slightly. Originally, the participants who were the top third and bottom third of the IAT scores were going to make up the participants of high and low bias. Those in the middle third would be excluded, and allow for the remaining two groups to serve as the extremes of the variable. These extremes would better allow for the differences within the groups to be observed. However, due to a lower than anticipated final participant number, participants were simply assigned to the high or low bias groups based on a median split. Future studies should aim to include a large enough sample size to allow for greater distinction between high and low bias.

This study attempted to bridge the gap between two sets of findings by seeing if the effects of bias inhibited the effectiveness of retraction type. Although the anticipated interaction was not found, the lack of an interaction can also give potential insight into these effects. Repeating the study to check for consistent results is necessary, but if these findings are repeated, it could suggest the importance of retraction type, in that they allow an individual, regardless of their bias, to better dismiss the prior information in favor of the updated article.

Overall, the best methods to combat misinformation are not so straightforward. This study aimed to find a significant interaction between two known methods of reducing misinformation: accounting for personal bias, and retraction type. Despite this study's manipulation now revealing an interaction, the lack of a significant relationship raises its own host of possibilities. It could be that the type of retraction has such an influence over recall that the effects of retraction overpower the effects of bias. If this were to hold constant in other studies, it would suggest the importance of formatting a retraction to be a specific retraction when writing a retraction statement.

REFERENCES

Eakin, Deborah & A Schreiber, Thomas and Sergent-Marshall, Susan. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology. Learning, memory, and cognition.* 29. 813-25.

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

Ecker, U. K. H., Lewandowsky, S., Change, E., and Pillai, R. (2014) The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology*, 20, 4, 323-335.

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

Ecker, U. K. H., Hogan, Joshua L., and Lewandowsky, S. (2017) Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition,* 6. 185-192.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17-41.

Howell, J. L., and Ratliff, K. A. (2017) Not your average bigot: The better-than-average effect and defensive responding to Implicit Association Test feedback. *British Journal of Social Psychology*, 56. 125-145.

Huff, M. J. and Umanath, S. (2017, August 17). Evaluating suggestibility to additive and contrary misinformation following explicit error detection in younger and older adults. *Journal of Experimental Psychology: Applied.* Advance online publication. http://dx.doi.org/10.1037/xap0000138

Lewandowsky, S., Ecker, U. K. H., Seifert, C., Schwarz, N., and Cook, J. (2012) Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest,* 13, 3, 106-131.

Rapp, David N. (2016) The consequences of reading inaccurate information. *Current Directions in Psychological Science.* 25, 4, 281-285.

Roberts, R. M., Neate, G. M., and Gierasch, A. (2017) Implicit attitudes towards people with visible difference: findings from an Implicit Association Test. *Psychology, Health & Medicine,* 22, 3. 352-358.

Swire, B., Ecker, U. K. H., and Lewandowsky, S. (2017, May 15). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory and Cognition.* 43(12), 1948-1961