

Poison Parasite Counter: Turning Duplicitous Mass Communications into Self-Negating
Memory Retrieval Cues

Robert Cialdini
Department of Psychology
Arizona State University

Jessica Lasky-Fink*
Goldman School of Public Policy
University of California, Berkeley

Linda J. Demaine
Sandra Day O'Connor College of Law
Arizona State University

Daniel W. Barrett
Department of Psychology
Western Connecticut State University

Brad J. Sagarin
Department of Psychology
Northern Illinois University

Todd Rogers
Harvard Kennedy School
Harvard University

***Corresponding author.** 2607 Hearst Ave., Berkeley, CA 94720. Email: jlaskyfink@berkeley.edu

Acknowledgments: The PPC idea is related to unpublished work by R.B. Cialdini, L.J. Demaine, K.v.L. Rhoads, D.W. Barrett, B.J. Sagarin, P. Petrova, J. Maner, and N.J. Goldstein. We thank Tanya Kent, Sam Madison, Steve Masnjak, Jennifer Murphy, Anna Valuev, and Alexa Weiss for research assistance. We thank Peter Clark, Adam Galinsky, Emma Levine, Malia Mason, Michael Norton, and Maurice Schweitzer for feedback on the research. In addition, we thank Steven Neuberg for exquisite counsel on the article's overall disposition. **Funding:** We thank the Harvard Foundations of Human Behavior, and Robert Cialdini. **Author contributions:** RBC, along with LJD, BJS, KvLR, and DWB conceived and developed the initial idea and designed an early version of the first and fourth studies. RBC, TR, and JLF developed the idea further and designed the studies reported here. JLF implemented and analyzed the studies. JLF, TR, and RBC wrote the paper with feedback from LJD, DWB, and BJS. **Competing interests:** The authors declare no conflict of interest. **Data and materials availability:** All materials and data available at osf.io/aps8h.

Statement of Relevance

The dissemination of false information is a serious societal problem that is worsened when deceptive communicators have greater access than their rivals to communication channels. To redress this imbalance and to increase the efficacy of counter-argumentation against disinformation, we propose the Poison Parasite Counter (PPC). The PPC leverages associative memory to turn a deceptive communication into a retrieval cue for a negating counter-message. This work demonstrates that the PPC more enduringly and effectively undermines the persuasive effects of a duplicitous communication than traditional counter-argumentation strategies. The research brings associative memory insights to the area of behavioral policy, and informs our understanding of political information processing, as well as of communication strategies in unequal information environments. The PPC represents an important step toward preventing legitimate critiques by disadvantaged voices from being drowned out by more advantaged communicators who exploit their position to disseminate deceptive information.

Abstract

Disinformation in politics, advertising, and mass communications has proliferated in recent years. Few counter-argumentation strategies have proven effective at undermining a deceptive message over time. This paper introduces the Poison Parasite Counter (PPC), a cognitive science-based strategy for durably countering deceptive communications. The PPC involves inserting a strong (poisonous) counter-message, just once, into a close replica of a deceptive rival's original communication. In parasitic fashion, the original communication then "hosts" the counter-message, which is recalled upon each re-exposure to the original communication. The strategy harnesses associative memory to turn the original communication into a retrieval cue for a negating counter-message. Seven experiments ($N = 3,679$) show that the PPC lastingly undermines a duplicitous rival's original communication, influencing judgments of communicator honesty and favorability, as well as real political donations.

In recent years disinformation in politics, advertising, and mass communications has proliferated (Lazer et al., 2018; Pennycook & Rand, 2019; Vosoughi et al., 2018). This paper introduces and tests a new cognitive science-based strategy that, while general in its applicability to persuasion processes, was structured specifically to increase the durability of persuasive counterarguments that aim to correct deceptive communications. The Poison Parasite Counter (PPC) leverages associative memory to increase the accessibility and salience of counterarguments by embedding a counter-message in a close replica of a rival's original communication. In doing so, the PPC effectively turns the original communication into a retrieval cue for the counter-message, which is then activated and evaluated upon each re-exposure to the original communication. Seven online experiments show that the PPC more enduringly undercuts an original communication than traditionally-presented counterarguments, and that this effect is driven by associative memory processes. This work contributes to behavioral policy research by applying novel behavioral insights to pressing political and social issues (Benartzi et al., 2017).

Commonly used strategies for combating disinformation (e.g., disclosures, forewarnings, fact checks) have generally yielded mixed results (Schwartz et al., 2016; Wahlheim et al., 2020). One notable exception entails presenting strong counterarguments after exposure to a rival's original message (Petrova & Cialdini, 2011). Even sound counterarguments, though, become less salient and accessible over time, thereby decreasing their persuasiveness and efficacy (Koriat et al., 2000). Consequently, traditional counter-argumentation is most—or only—effective when it is encountered immediately after every original message. This ensures that the countervailing message is accessible and salient each time the original message is re-encountered. In practice, however, this can be difficult to achieve. Even if a communicator is able to quickly craft a refutational message, unequal access to mass communication channels makes it difficult—if not impossible—for disadvantaged communicators to ensure equal exposure to counter-messages (Koerth-Baker, 2018; OpenSecrets, 2018). This asymmetry is particularly consequential as research has shown that increasing exposure to an argument increases viewers' perceptions of its veracity (Fazio & Sherry, 2020; Unkelbach et al., 2019).

This paper introduces the PPC as one strategy for increasing the efficacy of counterarguments against deceptive communications. The PPC involves two components—the “parasite” and the “poison”—each corresponding with distinct psychological processes. The “parasitic” component of the PPC relies on associative memory. Specifically, it leverages Tulving's (1983) encoding specificity principle, which suggests that memory recall is facilitated when conditions present when a memory is first encoded are also present at the time of retrieval (Petrova & Cialdini, 2011; Roediger et al., 2017). By this process, associating a counter-message with the imagery of a rival's original and existing communication effectively embeds the counter-message into the original communication. When employed, the perceptual similarity between the PPC-embedded counter-message and the original communication turns the latter into a retrieval cue for the countervailing message upon each re-exposure to the original communication. This strategy can mitigate the natural memory decay that decreases both the

accessibility and salience of previously encountered counterarguments over time (Wixted, 2004). By deliberately turning the original communication into a “host” for a countervailing message, the PPC also mitigates the persuasive effects of repeated exposure to the original communication.

Once counterarguments are made accessible and salient upon re-exposure to an original communication they can be used for online evaluation. The counterarguments must be sufficiently “poisonous” to effectively neutralize the original message upon evaluation. Tautologically, the PPC cannot be effective unless the counter-message is, by itself, effective. A particularly effective form of counter-argumentation is one that reveals the dishonesty of the original communicator. There appears to be an evolved human sensitivity and aversion to dishonesty such that, once deception is recognized, targets tend to resist both the communicator and information (Cosmides & Tooby, 1992; Farrelly et al., 2005; Fehr & Gächter, 2002).

We hypothesize that, together, the parasitic and poisonous components of the PPC will counter deceptive communications, even when the original communicator benefits from asymmetric exposure to target audiences. Seven randomized experiments test this in the realm of political and commercial advertising, two areas where both disinformation and asymmetric reach in communication channels are common and consequential (Gerber, 1998; Gerber, 2004; Lowery, 2014). We show that the PPC more effectively undercuts a rival’s original communication than traditionally-presented counterarguments at a single point in time (Studies 1-3); these effects persist across time and in the face of large frequency-of-exposure asymmetries (Study 4); the PPC can affect a consequential behavior (political donations) in a real-world setting (Study 5); the PPC can be effectively implemented with either static or video ads (Study 6); and the PPC effect is driven by associative memory processes (Study 7).

Sample sizes for all studies were determined prior to data collection based on pilot studies that yielded medium effects ($d = 0.70$, $d = 0.49$). Accordingly, for each study reported here (and in the SOM), we targeted a minimum of 70 participants per condition. With 80% power and a standard alpha level of 0.05, this yields a minimum detectable effect of $d = 0.48$ for each pairwise comparison. All materials and data are available on Open Science Framework (<https://osf.io/aps8h/>).

Study 1: The Poison Parasite Counter Effect

Study 1 tested whether embedding strong counter-messages into a replica of a deceptive rival’s political ad subsequently undermines the original communication relative to presenting the same counter-messages in a traditional, visually independent fashion.

Method

Participants. Participants were recruited through Amazon’s Mechanical Turk (MTurk) to complete a 10-minute online survey for which they were paid \$1.00 each. Only MTurk workers

located in the United States, who had an MTurk approval rating of at least 95%, who had not participated in previous similar studies, who consented to participate, and who passed an initial attention check were eligible to complete the survey. All studies reported here employed these standard qualifications (see SOM). A total of 297 participants (mean age = 38 years, $SD = 11.9$; 50% female) completed the study.

Procedure. All workers who consented to participate and passed the attention check were randomly assigned by the survey platform to one of two experimental conditions: the Traditional Counter condition ($N = 148$) or the Poison Parasite Counter condition ($N = 149$). Participants were not aware of their condition assignment.

In the approximately 10-minute survey, participants saw a total of 10 political ads for five fictional candidates—one “pro” and one “response” ad for each candidate. For one of these candidates, Walter McKinley, we developed two response ads corresponding with each experimental condition. The first utilized the PPC by overlaying counter-messages highlighting the duplicitous nature of the rival’s original message on an exact visual replica of the original ad (see Fig. 1B). The Traditional Counter ad presented the same counter-messages as the PPC ad, but with a different visual aesthetic that provided no associative links to the rival’s original ad (see Fig. 1C). The other fictional ads were used as filler to distract participants from the focus of the study and interfere with memory processes. For the same purpose, the ads were interspersed with four excerpts from current, non-partisan news articles, and filler questions were asked after every ad or article excerpt. Each ad or excerpt was shown on a separate page, and page timers ensured that each participant had to wait at least five seconds to advance from one page to the next. All participants saw the original Walter McKinley ad on the second screen, and either the PPC or Traditional Counter ad on the eighth. This paradigm forms the basis for each study reported in this manuscript. See SOM for detailed study procedures.

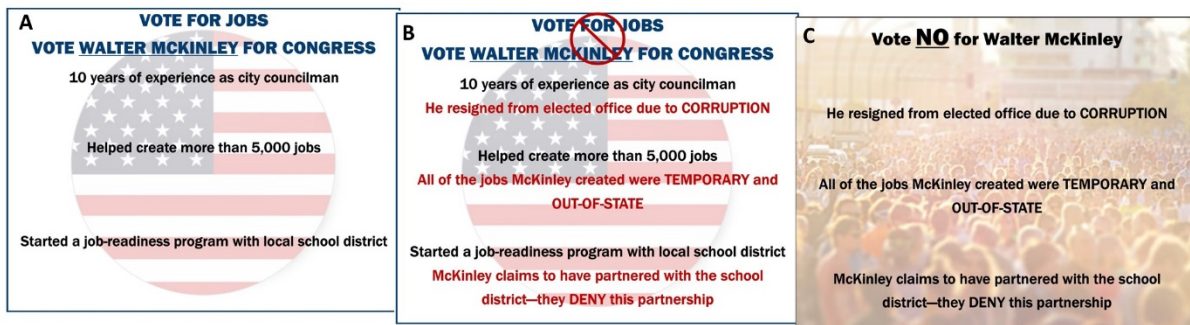


Fig. 1. Walter McKinley ads used in Studies 1 and 2. (A) Original ad; (B) PPC ad; (C) Traditional Counter ad

At the end of the survey, participants were told that they would be shown one of the candidate ads from the first section, chosen arbitrarily. All participants were again shown the original McKinley ad and were subsequently asked:

1. If Walter McKinley was running in your state, how likely is it that you would vote for him?
2. How honest do you think Walter McKinley is?

Both outcomes were ranked on a five-point scale, from extremely unlikely to extremely likely, and extremely dishonest to extremely honest, respectively.

In addition, immediately after viewing the PPC or Traditional Counter ad, participants were shown four counterclaims and asked to identify which one was *not* mentioned on the ad that they had just viewed. Testing immediate recall after viewing the PPC or Traditional Counter ad allowed us to determine whether participants were paying equal attention in both conditions. In the memory literature, this type of design is referred to as creating an A-B, A-D retroactive interference experiment.

Analysis and Results

We hypothesized that by directly challenging the claims in the original ad and creating mnemonic links between the counter ad and the original ad, the PPC would result in recurring, cue-based recall of the counter-messages whenever the original ad was re-encountered, making the counterargument resistant to normal memory degrading processes. This would, in turn, lead participants to view McKinley as more dishonest and reduce their willingness to vote for him compared to a traditionally presented counter ad.

In a standard linear model, we regressed each primary outcome on an indicator for treatment assignment, controlling for participant gender, age, party affiliation (Democrat, Republican, Independent, or Other), and an indicator for college education. To facilitate comparability across studies and outcomes, each outcome measure is standardized.

As predicted, exposure to the PPC ad subsequently reduced participants' reported likelihood of voting for McKinley relative to participants who had seen the Traditional Counter ad by 0.61 SD, $t(296) = 5.52$, $p < .001$, 95% confidence interval (CI) = [0.39, 0.83] (see Fig. 2, which reports unstandardized means for ease of interpretability). In a similar pattern, participants who saw the PPC ad rated McKinley as 0.70 SD, $t(296) = 6.40$, $p < .001$, 95% CI = [0.49, 0.92], less honest than participants in the Traditional Counter condition.

When asked immediately after viewing the PPC or Traditional Counter ad to identify which of four counterclaims was not included in the PPC or Traditional Counter ad, 81% of participants in the PPC condition answered correctly, as did 77% of participants in the Traditional Counter condition ($\chi^2(1) = 0.56$, $p = .45$). This suggests that participants in both conditions paid similar attention to the counterclaims.

Discussion

Study 1 demonstrated that the PPC more effectively undermines a deceptive rival's original communication than traditionally-presented counterarguments.

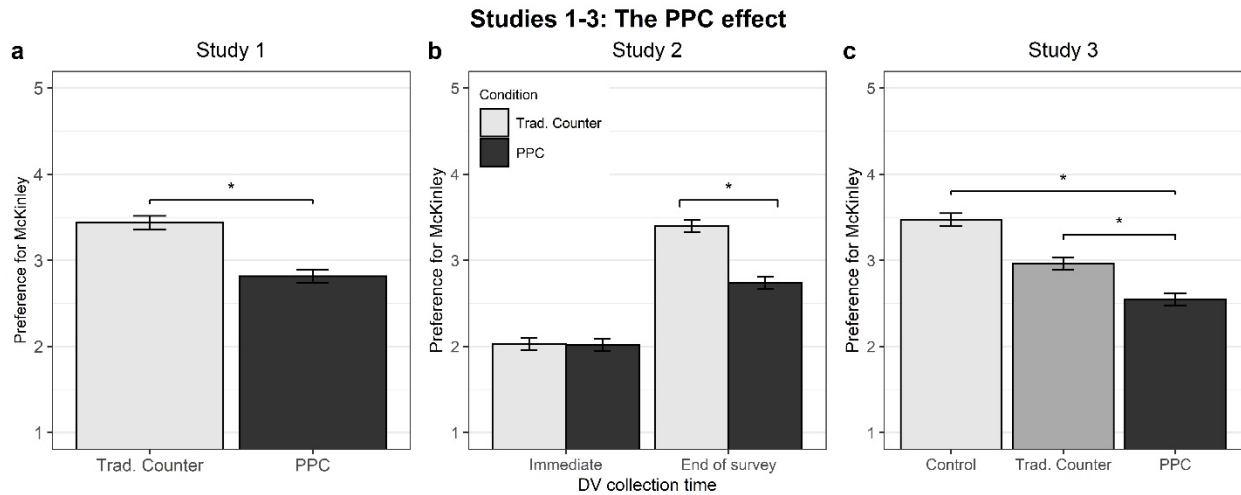


Fig. 2. Results from (a) Study 1; (b) Study 2; and (c) Study 3. Outcomes reflect regression-adjusted unstandardized means, measured on 1-5 scale where 5 represents “extremely likely to vote.” Error bars reflect ± 1 standard error * reflects differences significant at $p < .01$.

Study 2: The PPC Retrieval Cue Mechanism

The PPC and Traditional Counter ads in Study 1 had identical counter-messages and were from the same source, which ensures that any difference between conditions was not the result of source sleeper effects (Kumkale & Albarracín, 2004). Yet, it is possible that the effects were driven by differential ad quality. If the PPC ad was simply a superior ad, we would expect it to dominate the Traditional Counter ad immediately after exposure. If instead the superiority of the PPC is due to the memory retrieval cue mechanism, effects should emerge only after participants are re-exposed to the original communication. Study 2 tested this directly.

Method

Participants. Participants were 712 MTurk workers (mean age = 36 years, $SD = 10.9$; 52% female) who were recruited to complete a 10-minute online survey for which they were each paid \$1.20. Standard participant qualifications were applied (see SOM).

Procedure. The content and outcomes for this study were identical to those used in Study 1 (Fig. 1), with Walter McKinley as the focal candidate. In a pre-registered 2 x 2 factorial design, all participants were randomly assigned to either the PPC or Traditional Counter condition, as well as to either an Immediate Outcome or End-of-survey Outcome condition. One-hundred and seventy-five participants were assigned to the PPC – Immediate Outcome condition; 181 to the

Traditional Counter – Immediate Outcome condition; 178 to the PPC – End-of-survey Outcome condition; and 178 to the Traditional Counter – End-of-survey Outcome condition. Participants were not aware of their condition assignment.

The study procedures exactly followed those of Study 1, with the exception of the timing of the outcome questions. In the Immediate Outcome condition, participants answered the two dependent variable questions immediately after seeing the PPC or Traditional Counter ad. In the End-of-survey Outcome condition, participants answered the two dependent variable questions at the end of the 10-minute long survey, after being re-exposed to the original McKinley ad. In both conditions, the dependent variable questions were the same as in Study 1. See SOM for detailed procedures.

In order to keep the survey the same length for participants in all experimental conditions, those assigned to the End-of-survey Outcome condition were also asked two questions—one recall, and one filler (see SOM)—immediately after viewing the PPC or Traditional Counter ad. The recall question asked participants to identify which of four possible counterarguments was *not* mentioned on the PPC or Traditional Counter ad that they had just viewed.

Analysis and Results

All analysis procedures followed those described in Study 1, and are pre-registered on OSF. If the PPC and Traditional Counter ads are of similar quality, participants' reactions to McKinley should be the same across both conditions when assessed immediately after viewing the counter ad. As expected, when measured immediately after viewing the response ad, participants who had seen the PPC ad were as likely to vote for McKinley as participants who had seen the Traditional Counter ad, $F(1,348) = 0.00, p = .95$. There was also no difference in perceptions of McKinley's honesty immediately after viewing the PPC or Traditional Counter ad, $F(1,348) = 1.28, p = .26$. Additionally, immediate recall did not differ significantly across conditions: when participants in the End-of-survey Outcome condition were asked to identify the counterclaims they had just seen immediately after viewing the PPC or Traditional Counter ad, 71% of participants in the PPC condition and 72% of participants in the Traditional Counter condition were able to correctly do so, $\chi^2(1) = 0.02, p = .89$.

When participants were asked the outcome questions at the end of the survey—after interference-producing filler and decoy materials—the results parallel those of Study 1 (see Fig. 2). As would be expected given evidence of quick decay of persuasive effects and the memory interference effects of interpolated materials, overall favorability of McKinley was higher among participants in both conditions at the end of the survey compared to immediately after exposure to the treatment ad. Yet, as hypothesized, the PPC mitigated this decay: upon re-exposure to the original communication, participants in the PPC condition who responded to the outcome measures at the end of the survey were significantly less likely to vote for McKinley (-0.63 SD, $t(355) = -6.64, p < .001, 95\% \text{ CI} = [-0.82, -0.45]$), and rated him as significantly less honest (-0.41 SD, $t(355) = -4.23, p < .001, 95\% \text{ CI} = [-0.61, -0.22]$), than those who saw the Traditional

Counter ad. The SOM presents additional results from pre-registered analyses that use an interaction model to evaluate the differential effect of the PPC by timing of the outcome questions.

Discussion

At the time of exposure, the PPC and Traditional Counter ads spurred equal recall and produced similarly strong counter-responses to the original ad, making it implausible that the PPC ad was merely a superior ad or that the qualities of the PPC inherently facilitate a more fluid comparison with the original material.

Study 3: The PPC Versus No Counter-message Control

Study 3 added a control group and tested the effect of the PPC ad against a more externally valid Traditional Counter ad, which offered the same counterarguments as the PPC ad, but presented in a more realistic form.

Method

Participants. Participants were recruited through MTurk to complete a 10-minute online survey for which they were paid \$.60 each. Standard participant qualifications were applied (see SOM). A total of 602 participants (mean age = 40 years, $SD = 12.4$; 58% female) completed the study.

Procedure. All workers who consented to participate and passed the attention check were randomly assigned to one of three conditions: Control ($N = 197$), Traditional Counter ($N = 201$), or PPC ($N = 204$). Participants were not aware of their condition assignment.

In Studies 1 and 2, the counter-messages presented in the Traditional Counter ad were identical to the messages presented in the PPC ad (see Fig. 1). In reality, response ads often encompass cohesive narratives rather than individual and fragmented counterclaims. Accordingly, Study 3 tested the effect of the PPC against a more externally valid Traditional Counter ad, as shown in Fig. 3. While the Traditional Counter ad offered the same counterarguments as the PPC ad, it was presented in a more realistic form. In addition, Study 3 included a pure control group, in which participants still saw the original McKinley ad, but did not see a counter ad or any counterclaims against McKinley. The purpose of the control condition was to evaluate the relative effects of seeing counterclaims against McKinley delivered via the PPC or Traditional Counter ads, compared to only seeing the original ad and its positive claims about the focal candidate (McKinley). To ensure that all participants saw the same total number of ads, participants in the Control group saw an ad for an unrelated fictional candidate in place of the PPC or Traditional Counter ad.

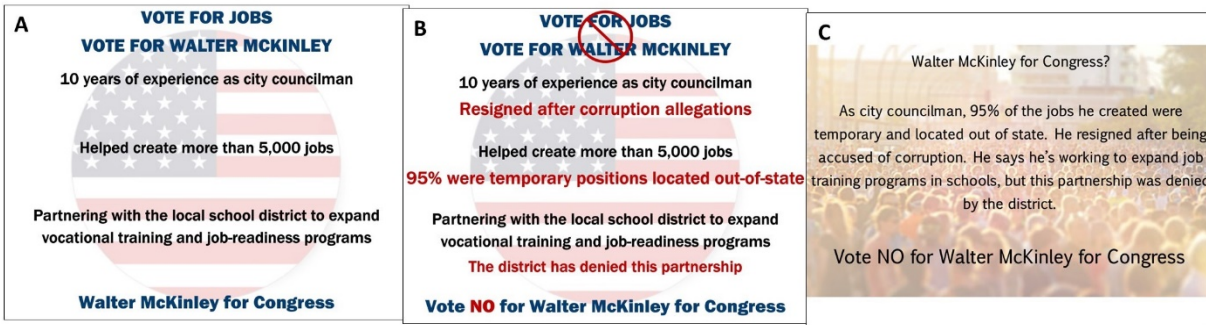


Fig. 3. Walter McKinley ads used in Studies 3, 4, and 7, as well as Supplement Studies I, II, and III. (A) Original ad; (B) PPD ad; (C) Traditional Counter ad.

Following the same procedure as Study 1, all participants saw the original McKinley ad on the second survey page, and either the PPC, Traditional Counter, or Control ad on the fourth page. These ads were interspersed with eight other ads for fictional political candidates and filler questions designed to distract participants from the focus of the study and interfere with memory processes. At the end of the survey, all participants were told they would be re-shown and asked questions about an arbitrarily chosen ad from the first section. All participants were re-shown the original McKinley ad and were subsequently asked the same dependent variable questions as in Studies 1 and 2. See online supplement for detailed procedures.

Analysis and Results

This study was pre-registered on OSF, and analysis procedures followed those described in Study 1. As shown in Fig. 2, exposure to the PPC ad significantly reduced participants' subsequent likelihood of voting for McKinley relative to the Control ad (-0.85 SD, $t(601) = -8.98$, $p < .001$, 95% CI = $[-1.03, -0.66]$), and relative to the Traditional Counter ad (-0.38 SD, $t(601) = -4.10$, $p < .001$, 95% CI = $[-0.57, -0.20]$). In a similar pattern, participants who saw the PPC ad also rated McKinley as significantly less honest than participants in both the Control condition (-0.68 SD, $t(601) = -7.04$, $p < .001$, 95% CI = $[-0.87, -0.49]$), and in the Traditional Counter condition (-0.30 SD, $t(601) = -3.11$, $p = .002$, 95% CI = $[-0.48, -0.11]$).

Relative to the Control condition, the Traditional Counter ad also reduced participants' likelihood of voting for McKinley (-0.47 SD, $t(601) = -4.95$, $p < .001$, 95% CI = $[-0.65, -0.28]$), and perceptions of his honesty (-0.38 SD, $t(601) = -3.98$, $p < .001$, 95% CI = $[-0.57, -0.19]$), although these effects are about half the size of those produced by the PPC ad. The SOM reports results from additional pre-registered analyses.

Discussion

The first three studies demonstrated that the PPC slows the decay of a counter-message's persuasive effects during a single study session, relative to presenting the same counter-message

in a traditional fashion. Supplemental Studies I and II replicated these results with variations in the frequency of exposure to both the original ad and counter ad.

Study 4: PPC Durability Over Time

A message's persuasive effect tends to decay over time, especially in the presence of memory interference (Gerber et al., 2011). Study 4 examined the durability of the PPC's persuasive effect over a 17-day period in a repeated-exposure experiment.

Method

Participants. An initial sample of 557 participants (mean age = 37 years, $SD = 11.7$; 67% female) was recruited via Amazon's MTurk to complete a series of five surveys. Standard participant qualifications were applied (see SOM). Participants were compensated \$1.10 for the each of the first and second surveys; \$0.70 each for the third and fourth surveys; and \$1.00 for the fifth survey. Additionally, all participants who completed the first four surveys received a \$0.75 bonus as an incentive to return for the fifth and final wave.

Of the 557 participants who completed the first wave of the study, 330 completed all five waves (mean age = 39 years, $SD = 12.6$; 69% female). Attrition was balanced evenly across experimental condition ($\chi^2(2) = 0.80, p = .67$). Younger participants, participants whose party affiliation was Independent, and non-college educated participants were all more likely to attrit. A joint significance test shows that we cannot reject the null hypothesis that attrition was balanced across conditions and all covariates ($\chi^2(8) = 23.05, p = .003$). All analyses thus control for party affiliation, college education, age, and participant gender.

Procedure. The study was run over a period of 17 days, with waves conducted on days 1, 3, 6, 9, and 16, and used the same materials as Study 3. Participants were told in wave 1 (day 1) that the study entailed five separate surveys. Each subsequent wave was open for 24 hours to all participants who had completed the preceding survey. Reminder emails were sent to all eligible workers when each follow-up survey opened.

In the first wave on day 1, all participants saw the original McKinley ad, interspersed with four other ads for fictional political candidates and three non-partisan news article excerpts. Filler questions designed to produce memory interference were asked after three of the ads and after each article excerpt. The original McKinley ad was shown on the third survey page for all participants. As in Studies 1-3, page timers were used to ensure that participants spent a minimum amount of time on each article and ad. At the end of the survey, participants were told that they would be shown two arbitrarily selected ads from the first section, and asked questions about each. All participants were first shown the original McKinley ad, and asked the two dependent variable questions. The second ad was randomly selected from the decoy fictional ads. Participants were asked the same dependent variable questions about the decoy candidate and ad in order to draw attention away from McKinley as the focal candidate.

In the second wave, launched on day 3, all returning participants were randomly assigned to one of three conditions: Control, Traditional Counter, or PPC. Participants were not aware of their condition assignment. To ensure that we did not unintentionally provide associative links between the original McKinley ad and the PPC or Traditional Counter ads, participants were not initially re-exposed to the original McKinley ad in the second wave; they only saw the PPC, Traditional Counter, or Control ad, interspersed with four decoy ads for other fictional candidates and four non-partisan news article excerpts. Filler questions were asked after each excerpt, and all participants saw the PPC, Traditional Counter, or Control ad on the fourth survey page. At the end of the survey, participants were told that they would be shown two arbitrarily selected ads from either the first or second wave of the study, and asked questions about each. All participants were shown a randomly selected ad for one of the decoy candidates first, and were then shown the original McKinley ad. The standard dependent variable questions were asked after each.

Waves 3, 4, and 5 (on days 6, 9, and 16) all followed similar procedures, but participants were not re-exposed to any counterclaims about McKinley. In each wave, returning participants were again shown the original McKinley ad, interspersed with two decoy ads, three news article excerpts, and related filler questions. At the end of every wave, participants were again asked the two dependent variable questions about two “arbitrarily” selected candidates—one was always Walter McKinley, while the second subject differed in each wave. See SOM for detailed procedures.

By the end of the five-wave study, all participants had seen the original pro-McKinley ad nine times, and the PPC, Traditional Counter, or Control ad only once (in wave 2). Of the 330 participants who completed all five waves, 113 were assigned to the Control condition, 111 to the Traditional Counter condition, and 106 to the PPC condition.

Analysis and Results

This study was pre-registered on OSF, and analysis procedures followed those described in Study 1. The results from the second wave, the only wave in which participants saw either the PPC, Traditional Counter, or Control ads, parallel those of Study 3: the PPC had a large and significant effect on the likelihood of voting for McKinley relative to both the Traditional Counter ad (-0.75 SD, $t(329) = -6.27$, $p < .001$, 95% CI = $[-0.98, -0.51]$) and the Control ad (-1.18 SD, $t(329) = -9.91$, $p < .001$, 95% CI = $[-1.41, -0.94]$). Similarly, participants who saw the PPC ad rated McKinley as significantly less honest than participants who saw the Traditional Counter ad (-0.75 SD, $t(329) = -6.20$, $p < .001$, 95% CI = $[-0.99, -0.51]$) or the Control ad (-1.13 SD, $t(329) = -9.30$, $p < .001$, 95% CI = $[-1.36, -0.89]$).

After being exposed to traditionally-presented counterclaims against McKinley, participants who saw the Traditional Counter ad were also less likely to vote for McKinley (-0.43 SD, $t(329) = -3.68$, $p < .001$, 95% CI = $[-0.66, -0.20]$) and rated him as less honest (-0.37 SD, $t(329) = -3.13$, $p = .002$, 95% CI = $[-0.61, -0.14]$) than participants in the Control condition who did not see any counterclaims against McKinley.

Given evidence that the persuasive effects of advertising diminish over time, we hypothesized that the effect of the traditionally-presented counterclaims against McKinley seen in wave 2 would decay over subsequent survey waves, especially in the face of repeated exposure to the original pro-McKinley ad. At the same time, we predicted that the persuasive effects of the PPC would be more resistant to this natural decay since each re-exposure to the original ad should reactivate memory of the counterclaims embedded in the PPC ad.

As predicted, in waves 3-5, the effect of the PPC remained significantly superior to the Traditional Counter ad (see Fig. 4, which reports unstandardized means). By the fifth wave (on day 16), there was no difference in reported likelihood of voting for McKinley between participants who had seen the Traditional Counter ad in wave 2 and participants in the Control condition who had not seen any counterclaims against McKinley during the study ($F(1, 321) = 1.27, p = .26$). Yet, participants who had seen the PPC ad during the second wave (on day 3) remained significantly less likely to vote for McKinley in the fifth wave (-0.74 SD, $t(329) = -5.83, p < .001, 95\% \text{ CI} = [-0.98, -0.49]$), and continued to rate him as significantly less honest (-0.43 SD, $t(329) = -3.27, p = .001, 95\% \text{ CI} = [-0.69, -0.17]$) than those who had seen the Traditional Counter ad (on day 3). The SOM reports full results for both outcomes for each wave (Tables S3, S4), as well as results from additional pre-registered analyses.

Discussion

Despite repeated exposure to the original pro-McKinley ad, and time-induced memory interference, the persuasive effects of the PPC endured and continued to undercut the original ad throughout the two-week period of Study 4. In contrast, the initial effect of the Traditional Counter ad waned over the two-week period, consistent with work showing that political ad effectiveness decays over time (Gerber et al., 2011; Kalla & Broockman, 2018). These findings were replicated in Supplemental Study III.

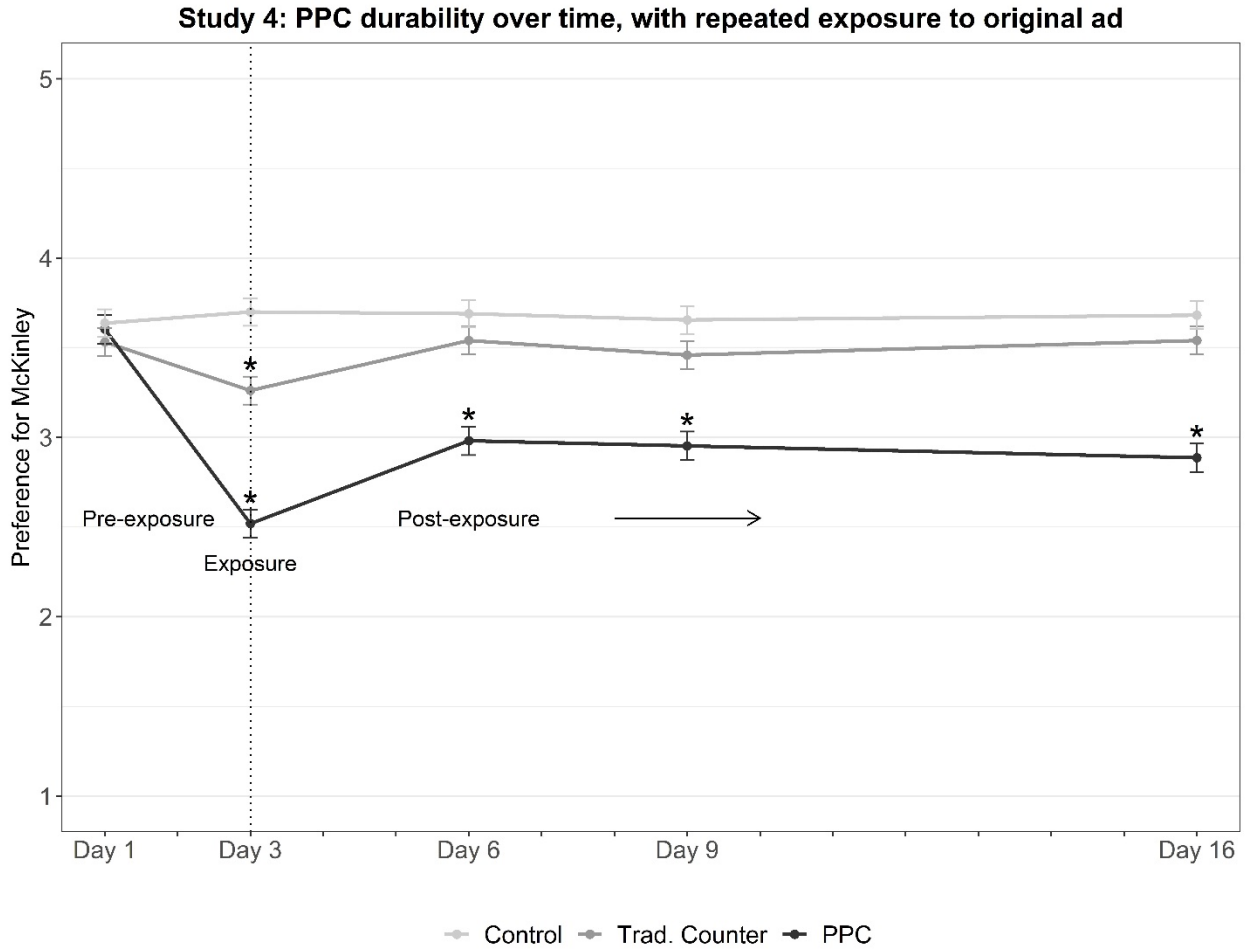


Fig. 4. Preference for McKinley each day in study 4. All participants saw the Control, Traditional Counter, or PPC ad a single time on day 3. In all other waves (on days 6, 9, and 16), participants saw the original ad interspersed with decoy ads and filler material. Outcomes reflect regression-adjusted unstandardized means, measured on 1-5 scale where 5 represents “extremely likely to vote.” Error bars reflect ± 1 standard error. * reflects differences significant at $p < .01$ relative to Control condition.

Study 5: PPC Effect on Actual Political Donations

The first four studies provided strong evidence of the efficacy and durability of the PPC in a hypothetical contest between fictional political candidates. Study 5 demonstrates that the PPC can also affect a consequential behavior—political donations—in a real campaign using real political ads.

Method

Participants. Participants were 299 Amazon MTurk workers (mean age = 36 years, $SD = 10.7$; 55% female) who received \$1.30 compensation for completing the survey. Standard participant

qualifications were applied (see SOM). Because this study used actual campaign ads from the Michigan Democratic primary election, all workers in the state of Michigan were excluded per Institutional Review Board (IRB) requirements so as not to influence any potential voters' opinions prior to the election. Additionally, since the primary outcome measure asked about voting in a Democratic primary election, all self-identified Republicans were also excluded.

Procedure. A real print ad produced by Gretchen Whitmer's campaign, the leading 2018 Democratic gubernatorial candidate in Michigan, served as the original ad for Study 5. The Traditional Counter ad was an actual ad produced and circulated during the election by the campaign of Shri Thanedar, one of Whitmer's main Democratic opponents. Thanedar's ad provided strong and explicit counter-messages, but no associative links to the original Whitmer ad. We modified this counter ad to create two new versions using the PPC procedure: a "full" PPC ad and a "tailored" PPC ad (Fig. 5). In the "full" application, the original Whitmer ad and the Traditional Counter ad were placed side-by-side with a line down the middle and the respective headers: "Typical Gretchen Whitmer ad" and "Here's what we say in our ad." In the "tailored" application, the exact counter-messages from Thanedar's response ad were embedded in the original Whitmer ad. Both ads were created purely for research purposes; neither was used in the campaign or circulated to prospective voters. We had no a priori hypothesis as to which manifestation of the PPC would be more effective, and thus tested each relative to the real Traditional Counter ad produced by Thanedar's campaign.

The study was run prior to the primary election on August 6, 2018. The procedure for this study paralleled that of Study 1. All workers who consented to participate and passed the attention check were randomly assigned to one of four conditions: Control ($N = 76$), Traditional Counter ($N = 77$), Full PPC ($N = 75$), or Tailored PPC ($N = 71$). Participants were not aware of their condition assignment.

All participants were first told that they would see a series of ads for real political candidates currently running for office, as well as excerpts from actual news articles. All participants saw the original Whitmer ad on the third survey page and then, depending on their experimental condition, either one of the PPC ads, the Traditional Counter ad, or a Control ad for a different candidate in a different election on the twelfth survey page. As in prior studies, the purpose of the Control ad was to ensure all participants saw the same number of total ads; participants in the Control condition still saw the original Whitmer ad, but they were not exposed to any counterclaims against Whitmer.

The original Whitmer ad and the counter ads were interspersed with eight decoy ads, all of which were real campaign ads from other current political races across the United States, as well as five news article excerpts and related filler questions. At the end of the roughly 15-minute survey, all participants were told that they would be asked questions about one arbitrarily chosen ad. All participants were shown the original Whitmer ad and asked the following three dependent variable questions:

1. If you lived in Michigan, how likely would you be to vote for Gretchen Whitmer in the upcoming Democratic primary election?
2. How honest do you think Gretchen Whitmer is?
3. You have a chance to allocate real resources. We are donating \$0.10 on behalf of every worker who takes our survey. We can either donate this \$0.10 to Gretchen Whitmer's campaign or to the campaign of Shri Thanedar, her opponent. Who would you like us to donate this \$0.10 to?

In response to the third question, a total of \$21.20 was donated to Whitmer's campaign, and \$8.70 to Thanedar's campaign. All donations were made by the authors prior to the primary election.

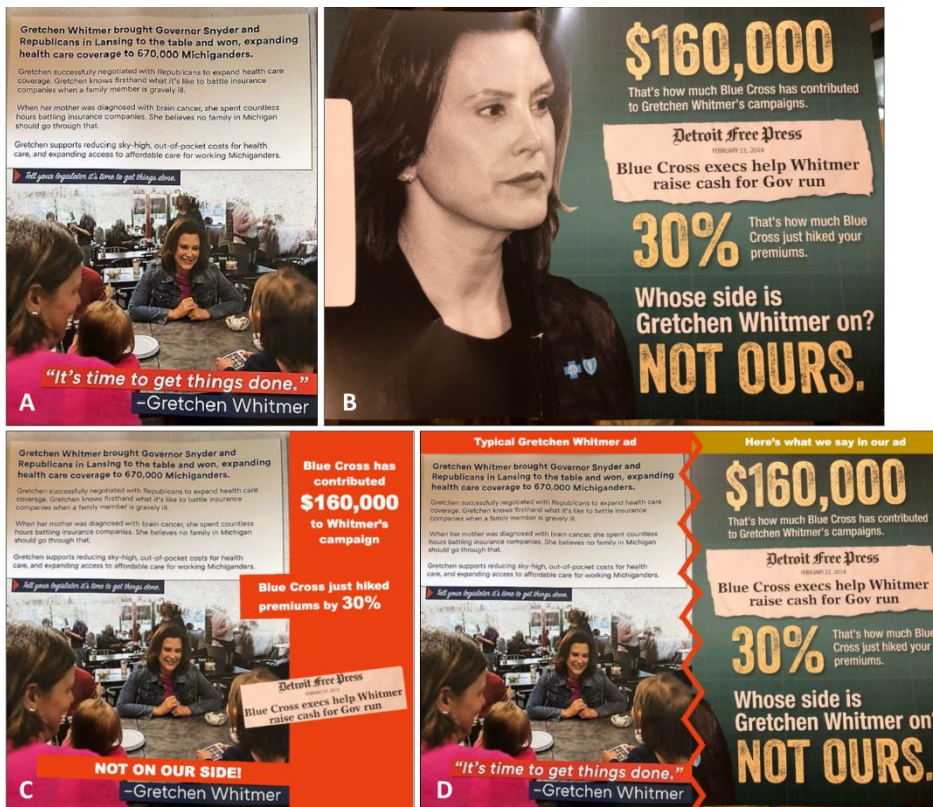


Fig. 5. Gretchen Whitmer ads. (A) Original ad; (B) Traditional Counter ad; (C) Tailored PPC ad; (D) Full PPC ad. Used in Study 5. Full size figures available in online supplement.

Analysis and Results

Consistent with results from the previous studies, participants in both PPC conditions were significantly less likely to express willingness to vote for Whitmer (see Fig. 6A, which reports unstandardized means), and rated her as significantly less honest than participants in the

Traditional Counter and Control conditions. Exposure to the Tailored PPC ad reduced participants' subsequent likelihood of voting for Whitmer by 0.40 SD, $t(298) = 2.48, p = .01$, 95% CI = [0.08, 0.71], and 0.50 SD, $t(298) = 3.15, p = .002$, 95% CI = [0.19, 0.82], relative to the Traditional Counter and Control ads, respectively. Similarly, exposure to the Full PPC ad reduced participants' likelihood of voting by 0.60 SD, $t(298) = 3.78, p < .001$, 95% CI = [0.29, 0.91], and 0.70 SD, $t(298) = 4.48, p < .001$, 95% CI = [0.39, 1.01], compared to the Traditional Counter and Control ads, respectively.

Participants in the Tailored PPC condition rated Whitmer as less honest than participants in both the Traditional Counter condition (-0.38 SD, $t(298) = -2.38, p = .02$, 95% CI = [-0.70, -0.07]) and Control condition (-0.43 SD, $t(298) = -2.66, p = .008$, 95% CI = [-0.75, -0.11]). Similarly, participants in the Full PPC condition rated Whitmer as less honest than participants in both the Traditional Counter condition (-0.62 SD, $t(298) = -3.93, p < .001$, 95% CI = [-0.93, -0.31]) and Control condition (-0.67 SD, $t(298) = -4.23, p < .001$, 95% CI = [-0.98, -0.36]).

There was no significant difference in voting likelihood or perceived honesty between participants in the Traditional Counter condition and those in the Control condition (see Fig. 6 and SOM, Table S7), demonstrating that the mere presence of counterarguments is not necessarily sufficient to produce meaningful resistance to a rival's original ad. Additionally, there was no significant difference between the two PPC conditions on either of the outcome measures, although the Full PPC ad produced larger directional results.

After the questions on voting and perceived honesty, participants were told that a \$0.10 donation would be made on their behalf to either Gretchen Whitmer or Shri Thanedar, her opponent, and were asked to direct the donation to their preferred candidate. Overall, a majority of participants directed the donation to Whitmer—in total, \$21.20 was donated to Whitmer's campaign, and \$8.70 to Thanedar's campaign. However, participants in the two PPC conditions were significantly less likely to direct their donations to Whitmer than participants in either the Traditional Counter or Control conditions. Only 58% of participants in the Full PPC condition directed the donation to Whitmer's campaign (instead of Thanedar's) compared to 75% in the Traditional Counter condition, $\chi^2(1) = 5.01, p = .03$, and 90% in the Control condition, $\chi^2(1) = 17.43, p < .001$. The Tailored PPC condition yielded very similar effects: 60% of participants in the Tailored PPC condition directed the donation to Whitmer's campaign, significantly less than in the Traditional Counter condition, $\chi^2(1) = 3.69, p = .05$, and in the Control condition, $\chi^2(1) = 15.10, p < .001$.

Discussion

Study 5 demonstrated that the PPC both reduces the viability of message, and influences consequential related behaviors like financial support for a political candidate. The relative equivalency of the Full and Tailored PPC versions illustrates that the PPC can be effectively operationalized in different ways.

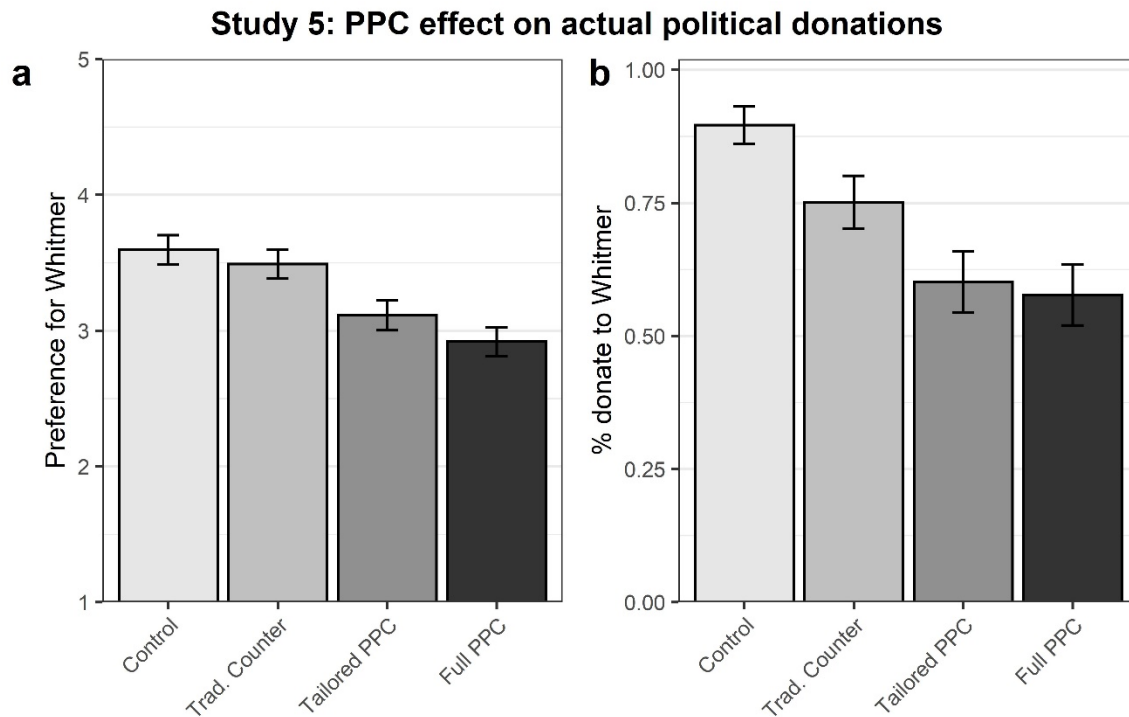


Fig. 6. (a) Preference for Whitmer, and (b) donations to Whitmer in study 5. Outcome for (a) reflects regression-adjusted unstandardized means, measured on 1-5 scale where 5 represents “extremely likely to vote.” Error bars reflect ± 1 standard error.

Study 6: Implementing the PPC using Video Ads

Study 6 examined the PPC in a nonpolitical, commercial context using video advertising, while also clarifying the underlying mechanism. In Studies 1-5, the PPC ad repeated the original ad’s false claim and directly refuted it, while the Traditional Counter ad offered the same refuting information, but without repeating the original false claims. Study 6 compares a PPC ad to a counter ad that employs an identical claim-refutation format.

Method

Participants. On the Friday before the 2020 Super Bowl, an initial sample of 2,429 participants (mean age = 41 years, $SD = 12.8$; 50% female) was recruited via Amazon’s MTurk for a three-wave study that was run over nine days. Standard participant qualifications were applied (see SOM).

Of the 2,429 participants who completed the first wave of the study (day 1), 2,152 (89%) completed the second wave (day 3), balanced evenly across conditions ($\chi^2(2) = 1.71, p = .43$). Then, anyone who completed the second wave and reported having watched the Super Bowl was invited to complete a third survey wave exactly one week later (day 9). Of the 2,152 participants

who completed the second wave of the study, 1,463 reported watching the Super Bowl between wave 1 and wave 2, and 1,172 (80%) of these participants completed the third wave. As such, our final analytic sample is comprised of 1,172 participants who watched the 2020 Super Bowl and completed all three surveys over the nine-day period (mean age = 41 years; $SD = 12.6$; 44% female).

Overall attrition across all three surveys was balanced evenly across conditions ($\chi^2(2) = 0.44, p = 0.80$), as was self-reported Super Bowl viewing ($\chi^2(2) = 1.25, p = 0.54$) among participants who completed wave 2. However, across demographics, women were less likely to report having watched the game, while those with higher incomes were more likely to have watched the game—both gender and income are controlled for in all analyses. Participants were compensated \$1.60 for the first survey; \$0.70 each for the second survey; and \$1.00 for the third survey.

Procedure. During the 2020 Super Bowl, TurboTax—one of the largest online tax preparation companies—ran a 45-second ad highlighting the simplicity and benefits of their software. They released the ad five days before the game, which afforded a unique opportunity to test the PPC knowing that participants who watched the game would be subsequently re-exposed to the original communication. Prior to the game, we developed three response ads. The PPC ad overlaid a counter-message on the original TurboTax ad that was to run during the Super Bowl. This message stated: “TurboTax says they work to make filing taxes easy for us. Yet, they’ve spent \$10 million lobbying lawmakers to prevent free automatic filing. This makes filing harder and more expensive for us, so they can make money.” This counterargument is true (see, e.g., OpenSecrets.org, 2020; Huseman, 2017; Elliott & Kiel, 2019), and offers evidence of TurboTax’s duplicity. In the PPC ad, this text scrolled twice across the screen during the 45-second video, and then ended with a static screen that displayed this message for an additional three seconds. In the Poison Only Counter ad, the same scrolling text was overlaid on a different TurboTax commercial, which was of an equivalent length. Just as in the PPC condition, the counterclaims in the Poison Only Counter ad scrolled twice across the screen during the 45-second video, and ended with a static screen that displayed the message for an additional three seconds. The underlying ad for the Poison Only Counter condition made similar claims as the original ad about TurboTax wanting filing taxes to be easier for people. And in the Pure Counterargument condition, the same scrolling text was presented with a solid black screen as the background.*

Participants were told in wave 1 (day 1) that the study would entail three waves, and were asked to signal their intention to complete all parts before proceeding to the survey. The first wave took place during the 24 hours preceding the Super Bowl. In a design that mimicked that of the previous studies, participants were randomly assigned in wave 1 to see the PPC ad, the Poison Only Counter ad, or the Pure Counterargument ad. Of the 1,172 participants who watched the Super Bowl and completed all three waves of the study, 389 were assigned to the Pure

* Ads available at <https://osf.io/aps8h/>

Counterargument condition, 398 to the Poison Only Counter condition, and 385 to the PPC condition.

Wave 1 was a 15-minute survey during which the assigned counter ad was interspersed with the original TurboTax ad—the one that was to run during the Super Bowl—and six other real television ads for a variety of services and products that served as filler. In addition, to ensure that the TurboTax response ad did not stand out, we also created and showed counter ads for three of the filler ads. Each of these decoy counter ads involved scrolling counterarguments against a solid background or against visually distinct ads, but none employed the PPC procedure. Participants were asked one or two filler questions after every ad. To help ensure that participants watched all ads in their entirety, ads were programmed to play automatically and could not be controlled by viewers, and page timers were used to ensure that participants could not proceed to the next survey page until an ad finished playing.

At the end of the survey, participants were told that they would be re-shown two arbitrarily chosen ads to answer questions about. All participants were first re-exposed to the original TurboTax ad and asked the following four questions to gauge how favorably they viewed TurboTax:

1. How positively or negatively do you view TurboTax?
2. If you were to use an online tax filing service, how likely would you be to use TurboTax?
3. There are many competing tax preparation companies. Imagine they all offer tax filing for the same price. Would you choose to file your taxes through TurboTax or one of its comparable competitors?
4. If a friend asked you for a recommendation on online tax filing services, which company would you be most likely to recommend?

On the Monday after the Super Bowl—beginning about 12 hours after the game—the initial sample of participants was re-recruited to complete the second wave survey. The original TurboTax ad was aired during the second quarter of the Super Bowl. This constituted the second exposure to the original ad for all participants who watched the game. We hypothesized that, as in prior controlled experiments with static ads, the PPC would mitigate decay of the counterclaims all participants saw in wave 1 (day 1), thereby neutralizing the persuasive effects of being re-exposed to the original TurboTax ad during the Super Bowl. As such, wave 2 asked participants the same outcome measures as in the first wave, but without explicit re-exposure to the original ad since the game served as the second re-exposure for those who watched. To draw attention away from TurboTax as the focal product, participants were also asked three questions about a randomly chosen product that they had seen an ad for during wave 1. As a manipulation check, at the end of wave 2—after the outcome measures were collected—all participants who reported having watched the Super Bowl were asked if they recalled seeing the original TurboTax ad. Sixty-seven percent recalled seeing the ad (see SOM). At the same time,

participants were also asked to recall the counterclaims against TurboTax that they had seen in wave 1.

Seven days later, all participants who completed wave 2 and reported watching the Super Bowl (i.e., participants who had presumably been exposed to the original ad a second time) were invited to complete wave 3. In wave 3, participants were exposed for a third time to the original TurboTax ad, and again asked about their attitudes toward TurboTax. See SOM for detailed procedures.

Analysis and Results

The analytic universe is limited to those participants who reported watching the Super Bowl ($N = 1,172$) and thus had the opportunity to be re-exposed organically to the original TurboTax ad during the game. If the PPC is effective, it should undercut the persuasive effects of the original TurboTax ad upon each subsequent re-exposure—during the Super Bowl, and then again during the third wave. The analysis presented here focuses on the effect of the PPC relative to both the Poison Only Counter ad and the Pure Counterargument ad. In the SOM, we also explore how removing the distraction of a background video (as in the Pure Counterargument ad) may affect recall and influence the efficacy of the ad, but that is not the main emphasis of this study.

The primary outcome of interest is a standardized aggregate “favorability index” comprised of the first two outcome measures. We evaluate the effect of treatment assignment on TurboTax favorability using a linear model controlling for participant race, age, party affiliation, gender, and income, as well as indicators for college education, prior TurboTax use, and whether self-reported weekly hours of television watched was above the median for the analytic sample.

Exposure to the PPC ad in the first wave (before the Super Bowl) effectively reduced TurboTax favorability during each subsequent wave. In the first wave (day 1), TurboTax favorability among participants who had seen the PPC ad was significantly lower than favorability among participants who had seen the Poison Only Counter ad ($M_{PPC} = -0.21$ SD; $\beta = -0.32$ SD, $t(1171) = -4.91$, $p < .001$, 95% CI = [-0.45, -0.19]; see Figure S1). Favorability was also significantly lower in the PPC condition compared to the Pure Counterargument condition ($M_{PPC} = -0.21$ SD; $\beta = -0.30$ SD, $t(1171) = -4.54$, $p < .001$, 95% CI = [-0.43, -0.17]).

In the second wave (day 3)—after being re-exposed to the original TurboTax ad during the Super Bowl—participants in the PPC condition still rated TurboTax as significantly less favorable than participants in both the Poison Only Counter condition ($M_{PPC} = -0.18$ SD; $\beta = -0.32$ SD, $t(1171) = -4.82$, $p < .001$, 95% CI = [-0.45, -0.19]) and the Pure Counterargument condition ($M_{PPC} = -0.18$ SD; $\beta = -0.21$ SD, $t(1171) = -3.22$, $p = .001$, 95% CI = [-0.34, -0.08]).

By the third wave (day 9), participants who had seen the PPC ad in wave 1 still rated TurboTax significantly less favorably than participants who had seen the Poison Only Counter ($M_{PPC} = -0.18$ SD; $\beta = -0.30$ SD, $t(1171) = -4.54$, $p < .001$, 95% CI = [-0.43, -0.17]) or Pure Counterargument ad ($M_{PPC} = -0.18$ SD; $\beta = -0.24$ SD, $t(1171) = -3.62$, $p < .001$, 95% CI = [-0.37,

-0.11]). Similar patterns can be seen across all four outcome measures independently (see SOM, Tables S8 and S9).

If the PPC procedure induces cue-based recall as hypothesized, we would expect recall of the counterclaims after subsequent re-exposure to the original ad during the Super Bowl to be highest among participants in the PPC condition. In wave 2, all participants who reported having watched the Super Bowl were asked to identify which of five possible counterclaims against TurboTax they had seen in the first survey wave. Of participants in the PPC condition, 74% correctly identified the counterclaim they had seen against TurboTax in wave 1, versus only 53% in the Poison Only Counter condition and 76% in the Pure Counterargument condition. False alarm rates in the PPC and Pure Counterargument conditions were 8% and 7%, respectively, and nearly 14% in the Poison Only Counter condition (see SOM, Table S10). This suggests that participants in the PPC condition did recall the counterclaims they had seen in wave 1 more clearly than participants in the Poison Only Counter condition. The SOM discusses possible implications of the comparatively high recall among participants in the Pure Counterargument condition.

Discussion

By demonstrating that the PPC can be effective via video ads, Study 6 extends the usefulness of the PPC. Additionally, the PPC and Poison Only Counter ads in Study 6 employed identical claim-refutation structures. The only difference between the two conditions is the associated underlying ad, suggesting that the effects of the PPC are not being driven by a more effective claim-refutation structure.

Study 7: Cue-Based Recall Mechanism

The PPC is effective across domains and modalities. Study 7 tested our hypothesis that the PPC leverages associative memory to overcome natural memory decay.

Method

Participants. Participants were 267 Amazon MTurk workers (mean age = 35 years, $SD = 11.5$; 51% female) who received \$1.80 for completing the survey. Standard participant qualifications were applied (see SOM).

Procedure. Study 7 used the same materials as Studies 3 and 4 (Fig. 3), with Walter McKinley again serving as the focal candidate. All workers who consented to participate and passed the attention check were randomly assigned to one of three conditions: Control ($N = 88$), Traditional Counter ($N = 88$), or PPC ($N = 91$). Participants were not aware of their condition assignment.

Following the same procedures as Studies 1 and 2, all participants saw the original McKinley ad on the third survey page, and either the PPC, Traditional Counter, or Control ad on the thirteenth survey page. These ads were interspersed with eight other ads for fictional political candidates, six excerpts from non-partisan news articles, and filler questions designed to distract

participants from the focus of the study and interfere with memory processes. At the end of the survey, all participants were told that they would be asked questions about one arbitrarily chosen ad from the first section. All participants were then shown the original McKinley ad and were subsequently asked the same dependent variable questions as in Studies 1-4, plus two additional recall questions as follows:

1. When you see the claims in McKinley's ad, how clearly do you recall the specifics of any arguments you may have viewed against those claims?
2. Which of the following anti-McKinley claims do you recall seeing, if any? *Participants were asked to choose from a list of eight options. Three of the choices were the exact counter-messages included in both the Traditional Counter ad and the PPC ad.*

See SOM for detailed procedures.

Analysis and Results

Following the same analysis procedures as in the previous studies, participants assigned to the PPC condition were again significantly less likely to vote for McKinley than participants in the Control condition (-0.70 SD, $t(266) = -5.02$, $p < .001$, 95% CI = $[-0.98, -0.43]$) and in the Traditional Counter condition (-0.43 SD, $t(266) = -3.06$, $p = .002$, 95% CI = $[-0.71, -0.16]$). Similar effects of the PPC can be seen on perceived honesty (see SOM, Table S11).

On the first recall measure, 27% of participants in the PPC condition reported they recalled seeing specific arguments against McKinley's claims "extremely clearly," versus 15% of participants in the Traditional Counter condition ($\chi^2(1) = 3.98$, $p = .05$) and 5% of participants in the Control condition ($\chi^2(1) = 13.27$, $p < .001$). Corroborating these reports, when asked to identify which anti-McKinley messages they saw—from a list of nine possible choices—participants in the PPC condition correctly identified, on average, 53% of the three counter-messages ($M_{PPC} = 1.6$, SD = 1.0) versus 42% among participants in the Traditional Counter condition ($M_{TC} = 1.3$, SD = 0.97), $t(266) = 2.65$, $p = .008$, 95% CI = $[0.09, 0.60]$, and 5% among participants in the Control condition ($M_C = 0.16$, SD = 0.40), $t(266) = 11.31$, $p < .001$, 95% CI = $[1.19, 1.70]$. Relatedly, 24% of participants in the PPC condition correctly identified *all* three counter-messages, compared to 15% of participants in the Traditional Counter condition ($\chi^2(1) = 2.69$, $p = .10$) and 0% of participants in the Control condition. The false alarm rate was 6% in both the PPC and Traditional Counter conditions ($F(1, 258) = 0.01$, $p = .92$), and 16.3% in the Control condition (compared to PPC condition, $F(1, 258) = 55.46$, $p < .001$). This suggests that participants in the PPC condition remembered the counterclaims they had seen against McKinley more clearly than participants in the Traditional Counter condition.

Discussion

The PPC spurs recall of its counter-messages upon repeated exposure to a rival's original ad. This explains at least some of the effect of the PPC on perceptions of the rival candidate's dishonesty and participants' subsequent voting preferences.

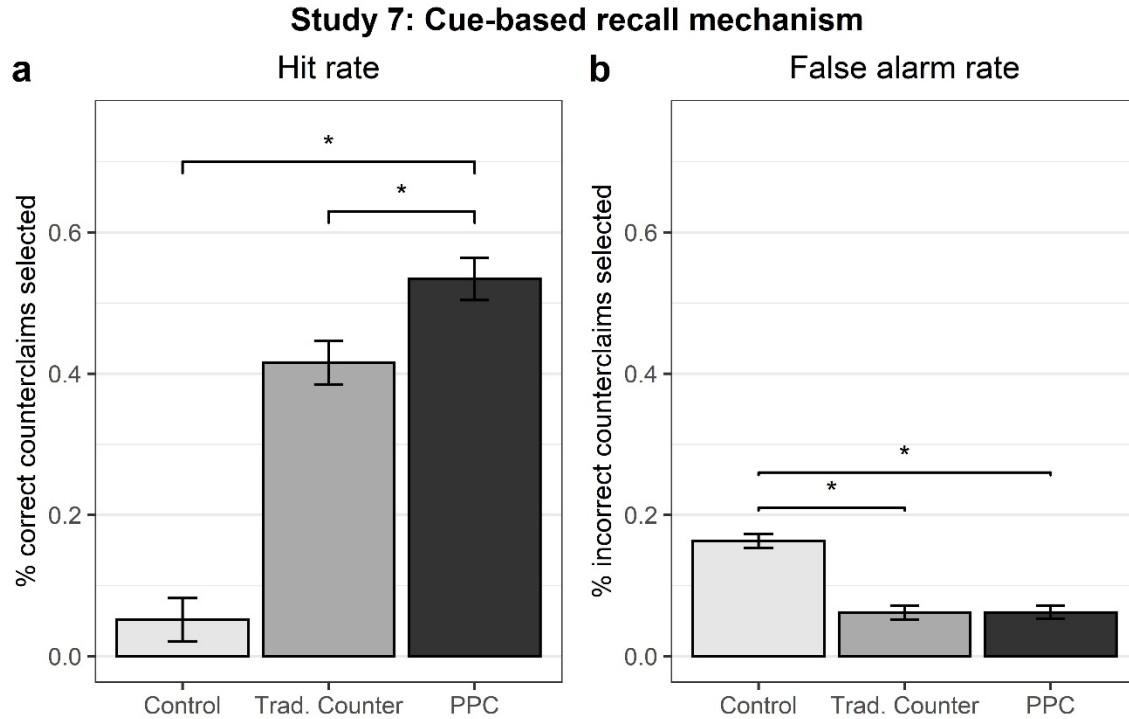


Fig. 7. (a) Hit rate in test of recall of counterclaims, and (b) false alarm rate in test of recall of counterclaims in study 7. Outcomes reflect regression-adjusted means. Error bars reflect ± 1 standard error. * reflects differences significant at $p < .01$.

General Discussion

The Poison Parasite Counter is a cognitive science-based strategy for durably countering deceptive rival communications. By using associative memory to parasitically link a counterargument to a deceptive rival's original communication, the PPC turns the original communication into a memory retrieval cue for a negating counterargument. Associating the counterargument with the original communication increases the accessibility and salience of the counterargument with each re-exposure to the original communication. Once the counterargument is made accessible and salient, it can influence (and subvert) online evaluation of the original communication. Seven studies (and three additional studies in SOM) show that, in some contexts, the PPC can enduringly undercut the persuasive effects of a rival's original communication, and more potently than traditional counter-argumentation strategies.

This research makes four contributions. First, it theoretically develops and empirically validates a strategy to counter deceptive communications from better resourced rival communicators. This is particularly important because other corrective strategies for countering disinformation show mixed effects (see, e.g., Dias et al., 2020; Ecker et al., 2017; Pennycook et al., 2020). This work can inform political psychologists who wrestle with how political information is processed and remembered, especially when communicators have asymmetric

access to communication channels (McGraw, 2000; Taber & Young, 2013). Similarly, it is relevant to marketing and communications scholars who study the dynamics and strategies of participants in unequal information environments (Mohr & Nevin, 1990).

Second, this work brings associative memory insights to the area of behavioral policy (Benartzi et al., 2017; Thaler & Sunstein, 2003). The PPC directly addresses one central challenge for behavioral policy: how to develop communications that bridge the time between when they are initially administered and when they are intended to have influence (Rogers & Frey, 2015; Rogers & Milkman, 2016).

Third, this work contributes to our understanding of multi-stage communication strategies. Recent work has explored the psychology of multi-stage social (Huang et al., 2017; Cooney et al., 2020) and negotiation communications (Bitterly & Schweitzer, 2020; Rogers et al., 2017). This research suggests that associative memory can play a powerful strategic role in these contexts.

Fourth, the PPC contributes to work on delayed persuasion, which examines sleeper (Kumkale & Albarracin, 2004) and inoculation effects (van der Linden & Roozenbeek, 2020). These effects are driven by source forgetting and defense fortification processes, whereas the PPC is driven by associative memory processes.

We highlight two key directions for future work. First, the studies reported here examined three paid advertisement stimuli responding to an original message that was duplicitous. The PPC was implemented in a similar fashion and yielded similar effects for all three ad subjects (McKinley, Whitmer, and TurboTax). This suggests that the manifestation of the PPC tested here—overlaying counterclaims on a rival’s original communication—may be an effective method of durably undercutting the persuasive effects of a paid political or commercial communication when the counterclaims offer evidence of the original communicator’s duplicity and dishonesty. Additional research should extend this work to other modalities, formats, communicators, sectors, message types, and social interactions—including speeches, negotiations, and dyadic conversations.

Second, we examined the PPC only in the context of communicators who were deceptive. Further research should examine the effectiveness of the PPC when the poisonous counter-messages simply counter rather than directly refute the original message.

References

- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8). doi:10.1177/0956797617702501
- Bitterly, T. B., & Schweitzer, M. E. (2020). The economic and interpersonal consequences of deflecting direct questions. *Journal of Personality and Social Psychology*, 118(5), 945-990. doi:0.1037/pspi0000200
- Cooney, G., Mastroianni, A. M., Abi-Esber, N., & Brooks, A. W. (2020). The many minds problem: disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, 31, 22-27. doi:10.1016/j.copsyc.2019.06.032
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). Oxford University Press.
<https://www.cep.ucsb.edu/papers/Cogadapt.pdf>
- Dias, N., Pennycook, G., & Rand, D. G. (2020, January). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *The Harvard Kennedy School Misinformation Review*, 1(1).
<https://misinforeview.hks.harvard.edu/article/emphasizing-publishers-does-not-reduce-misinformation/>
- Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, 6(2), 185-192.
- Elliott, J., & Kiel, P. (2019, October 17). Inside TurboTax's 20-year fight to stop Americans from filing their taxes for free. *ProPublica*. <https://www.propublica.org/article/inside-turbotax-20-year-fight-to-stop-americans-from-filing-their-taxes-for-free>
- Farrelly, M. C., Davis, K. C., Haviland, M. L., Messeri, P., & Healton, C. G. (2005). Evidence of a dose-response relationship between "truth" antismoking ads and youth smoking prevalence. *American Journal of Public Health*, 95(3), 425-431.
doi:10.2105/AJPH.2004.049692
- Fazio, L. K., & Sherry, C. L. (2020). The effect of repetition on truth judgments across development. *Psychological Science*, 31(9), 1150-1160. doi:10.1177/0956797620939534

- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137-140. doi:10.1038/415137a
- Gerber, A. (1998). Estimating the effect of campaign spending on senate election outcomes using instrumental variables. *American Political Science Review*, *92*(2), 401-411. doi:10.2307/2585672
- Gerber, A. (2004). Does campaign spending work?: Field experiments provide evidence and suggest new theory. *American Behavioral Scientist*, *47*(5), doi:10.1177/0002764203260415
- Gerber, A. S., Gimpel, J. G., Green, D. P., & Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *American Political Science Review*, *105*(1), 135-150. doi:10.1017/S000305541000047X
- Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of Personality and Social Psychology*, *113*(3), 430. doi:10.1037/pspi0000097
- Huseman, J. (2017, March 20). Filing taxes could be free and simple. But H&R Block and Intuit are still lobbying against it. *ProPublica*. <https://www.propublica.org/article/filing-taxes-could-be-free-simple-hr-block-intuit-lobbying-against-it>
- Kalla, J., & Broockman, D. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review*, *112*(1), 148-166. doi:10.1017/S0003055417000363
- Koerth-Baker, M. (2018, Sep. 10). How money affects elections. *FiveThirtyEight*. <https://fivethirtyeight.com/features/money-and-elections-a-complicated-love-story/>
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*, 481-537. doi:10.1146/annurev.psych.51.1.481
- Kumkale, G. T., & Albarracín, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin*, *130*(1), 143-172. doi:10.1037/0033-2909.130.1.143

- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096. doi:10.1126/science.aao2998
- Lowery, W. (2014, April 4). 91% of the time the better-financed candidate wins. Don't act surprised. *The Washington Post*. Retrieved from: <https://www.washingtonpost.com/news/the-fix/wp/2014/04/04/think-money-doesnt-matter-in-elections-this-chart-says-youre-wrong>.
- McGraw, K. M. (2000). Contributions of the cognitive approach to political psychology. *Political Psychology*, 21(4), 805-832. doi:10.1111/0162-895X.00217
- Mohr, J., & Nevin, J. R. (1990). Communication strategies in marketing channels: A theoretical perspective. *Journal of Marketing*, 54(4), 36-51. doi:10.2307/1251758
- OpenSecrets.org. (2018). Did money win? <https://www.opensecrets.org/elections-overview/did-money-win>.
- OpenSecrets.org. (2020). Client Profile: Intuit Inc. <https://www.opensecrets.org/federal-lobbying/clients/summary?cycle=2020&id=D000026667>
- Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *PNAS*, 116(7), 2521-2526. doi:10.1073/pnas.1806781116
- Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 66(11). doi:10.1287/mnsc.2019.3478
- Petrova, P. K., & Cialdini, R. B. (2011). New approaches toward resistance to persuasion. In G. Hastings, C. Bryant., & K. Angus (Eds.). *The Sage handbook of social marketing* (pp. 107-122). Sage Publications.
- Roediger, H. L., Tekin, W., & Uner, O. (2017). Encoding-retrieval interactions. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference*, 2nd edition (pp. 5-26). Elsevier. doi:10.1016/B978-0-12-809324-5.21036-5
- Rogers, T., & Frey, E. (2015). Changing behavior beyond the here and now. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 723–748). Wiley-Blackwell.

- Rogers, T., & Milkman, K. (2016). Reminders through association. *Psychological Science*, 27(7), 973-986. doi:10.1177/0956797616643071
- Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology*, 112(3), 456. doi:10.1037/pspi0000081
- Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, 2(1), 85-95. doi:10.1353/bsp.2016.0009
- Taber, C. S., & Young, E. (2013). Political information processing. In L. Huddy, D. O. Sears, & J. S. Levy (Eds.), *The Oxford Handbook of political psychology* (2 ed.). Oxford University Press. doi:10.1093/oxfordhb/9780199760107.013.0017
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, 93(2), 175-179. doi:10.1257/000282803321947001
- Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.
- Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, 28(3), 247-253. doi:10.1177/0963721419827854
- van der Linden, S., & Roozenbeek, J. (2020). Psychological inoculation against fake news. In R. Greifeneder, M. Jaffé, E. J. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation*. Psychology Press. <http://dx.doi.org/10.4324/9780429295379-11>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1156. doi:10.1126/science.aao9559
- Wahlheim, C. N., Alexander, T. R., Peske, C. D. (2020). Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term. *Psychological Science*, 31(10), 1325-1339. doi:10.1177/0956797620952797
- Wixted, J.T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235-269. doi:10.1146/annurev.psych.55.090902.141555