Running Head: Meta-Cognition and Resistance

Meta-Cognition and Resistance to Political Persuasion:

Evidence from a Three-Wave Panel Study

Joseph A. Vitriol

Harvard University

Howard Lavine

University of Minnesota, Twin-Cities

Eugene Borgida

University of Minnesota, Twin-Cities

Editor's Note: Jason Rentfrow served as Associate Editor for this article.

Author Note

The data and data syntax are available at:

https://osf.io/t8ck9/?view_only=dd2ea70e25ce48498623c919ab5ab6d7

Correspondence concerning this article should be addressed to Joseph A Vitriol,

Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138. Tel.:

215.771.4697. E-mail: joevitriol@gmail.com.

Abstract

We investigate the temporal course of meta-cognition and resistance processes following exposure to counter-attitudinal information in the 2012 Presidential election. Using a unique 3-wave survey panel design, we tracked eligible voters during the last months of the 2012 campaign and experimentally manipulated exposure to negative political messages targeting Barack Obama and Mitt Romney on an online platform. As predicted, we found that politically unengaged (vs. engaged) individuals were less likely to counter-argue a message attacking their favored candidate. Resistance, in turn, led to increased attitudinal certainty, polarization, and correspondence with actual voting behavior over the course of the campaign. These findings provide the first analysis of the longitudinal effects of meta-cognitive processes underlying persuasion for real-world attitude change and behavior.

KEYWORDS: Meta-Cognition; Resistance to Persuasion; Attitudinal Certainty; Political Psychology; Voting Behavior

Meta-Cognition and Resistance to Political Persuasion:

Evidence from a Three-Wave Panel Study

The strategic use of mass media messaging to motivate and persuade voters is commonplace in politics. In the 2012 and 2016 presidential and congressional elections, for example, candidates, parties, and outside groups spent more than two billion dollars alone on these efforts to mobilize and persuade (Fowler & Rideout, 2013; Sultan, 2017). Two general questions stand out as central in understanding the persuasive impact of communications like these. First, does any of it matter? Do those billions of dollars manage to change anyone's mind about who to support for political office? Second, to the extent that campaigns *do* matter, by what psychological processes do they influence and move voters from one side to the other?

With respect to the first question, political scientists argue that presidential campaigns typically play only a minimal role in persuading voters to change their policy positions and other political assessments (Brady & Johnston 2006; Huber & Arceneaux, 2007). Instead, the argument goes, most voters possess stable, cognitively efficient predispositions that are commonly activated and reinforced—but rarely changed—by information disseminated by persuasion campaigns and within social media (e.g., Gelman & King, 1993; Sides & Vavreck 2014). In an age of polarization and hyper-partisanship most voters simply toe the party line, with nine in 10 Democrats and Republicans voting for the party's presidential candidate in recent elections (Abramowitz 2010; Erikson & Wlezien, 2012). Importantly, the candidate preferences and voting intentions of politically engaged (vs.unengaged) citizens are especially strong and durable over time, suggesting that campaigns may have little impact on those who pay the most attention to politics (Abramowitz, 2010; Prior, 2013).

Perhaps, then, rather than inquiring into the processes by which campaigns produce persuasion, the more pertinent question is: what do we know about how voters *resist* counter-attitudinal messages? Although a substantial body of work has investigated the dynamics of political persuasion (e.g., Druckman, Slothuus & Peterson, 2013; Luttig & Lavine, 2015; Mutz, Sniderman & Brody 1996; Valentino & Nardis, 2013; Zaller 1992), virtually none of it has explored in real time—as a campaign actually unfolds—if (or which) voters manage to defend against counter-attitudinal information and what consequences this might have for longitudinal change in attitudes and behavior. In this research, we rely on the meta-cognitive model of attitudes (Tormala & Petty, 2002; Petty, Brinol & DeMarree, 2007) to understand the processes by which and individuals for whom counter-attitudinal information strengthens or weakens candidate preferences over the course of a presidential election, and its consequences for behavior.

Critically, the temporal course of resistance processes, while theoretically implied, has been empirically neglected previously. This is no small matter. Existing work on meta-cognition in persuasion settings has generally been limited to cross-sectional designs in laboratory contexts, and so it remains unclear if (or how) the effect of these processes on attitudes persists over time and with what consequences for actual behavior. Evaluating these dynamics longitudinally provides a particularly critical test of meta-cognitive perspectives on resistance to persuasion, especially whether these processes do, in fact, translate into meaningful long-term changes in attitudes or behavior in real-world contexts, as is not always found for other psychological phenomena (e.g., Lai et al., 2016).

Therefore, our focus is on whether the temporal dynamics of meta-cognitive resistance processes, which have been theoretically implied but not empirically examined, account for why

voters (especially those who are highly engaged in politics) are often unaffected by persuasion

efforts, and why political campaigns and social media environments tend to produce partisan

polarization. In doing so, we provide the first investigation of the implications of meta-cognitive

processes for real-world attitudes and behavior. We also demonstrate the utility of this

framework for understanding the electoral consequences of political communication across

media platforms.

The Meta-Cognitive Approach to Resistance Processes

Central to our approach is the role of meta-cognitions in the persuasion process.

According to Tormala and Petty (2002), meta-cognitions are the conscious inferences that people

make about their own thought processes, affect, or behavior. Past research indicates that

weathering a credible persuasion attempt produces attitudes that are held with more confidence

(or certainty), less vulnerable to subsequent persuasion attempts, and more likely to correspond

to behavior (Tormala & Petty, 2002, 2004a, b, c). As Tormala, Clarkson, and Petty (2006, p.

423-424) explain, "when people resist persuasion they can perceive this resistance, reflect on it,

and form specifiable attribution-like inferences about their own attitudes. These inferences, in

turn, affect attitude certainty." In particular, when people believe that they have successfully

resisted an attempt to change their attitudes, they infer that their view must be valid and correct

(Tormala & Petty 2002). This work suggests that when voters resist political messages they may

also increase the certainty of their prior attitudes, thereby strengthening both their initial

candidate preferences as well as the link between their attitudes and, over time, their subsequent

voting behavior. Polarization may thus be exacerbated by meta-cognitive resistance processes

that result from exposure to counter-attitudinal messages.

However, the impact of meta-cognitive resistance processes depends on the ability and motivation of the message recipient, as defending one's attitudes against attack requires a relatively high level of cognitive effort and skill (Petty, Brinol & Tormala, 2002). It is important to note that similar perspectives have been advanced in political science. According to Zaller's (1992, p. 44) model of mass opinion change, for example, "the likelihood of *resisting* persuasive communications that are inconsistent with one's political predispositions rises with a person's level of political attentiveness." He finds that when elites divide on an issue, unengaged individuals tend to be persuaded by whichever information stream (pro or con) is "louder," whereas engaged individuals tend to resist the counter-attitudinal message and change their opinions in the direction of the pro-attitudinal message (for similar results, also see Chong & Druckman, 2010 and Taber & Lodge, 2006).

Findings consistent with the proposition that the influence of counter-attitudinal information on attitudes is conditioned on political engagement also result from studies of the impact of television advertising in U.S. House and Senate races: while persuasion can occur across levels of engagement, the effects are stronger among citizens low in engagement (Ridout & Franz, 2011). Importantly, studies of counter-framing effects also indicate that message recipients with weak (vs. strong) preexisting opinions become less (vs. more) certain in their pre-existing opinion following repeated exposure to counter-attitudinal messages (Chong & Druckman, 2013; but see Kam, 2006). A more recent investigation finds that exposing partisans to counter-attitudinal information on social media platforms (i.e., Twitter) failed to induce persuasion effects, but instead led to increased polarization and entrenchment in pre-existing preferences (Bail et al., 2018). Nevertheless, the underlying psychological mechanisms by which such backfire effects to counter-attitudinal information occurs has not, to our knowledge, been

directly investigated[1]. This is a serious oversight, as these findings make clear that meta-cognitive processes may help explain why counter-attitudinal information can lead to this kind of polarization.

The Present Research

In sum, previous work indicates that the failure to resist persuasion can undermine attitude certainty, polarization, and correspondence with behavior. Moreover, research in both social and political psychology suggests that resistance is most likely to occur among those who are highly knowledgeable about and interested in politics (i.e., the politically unengaged), as they are the most able and motivated to react critically to the political arguments they receive (see Taber & Lodge 2006). In contrast, the politically unengaged are less able to defend their attitudes against attack and, consequently, may experience reductions in attitude certainty and behavior correspondence following exposure to counter-attitudinal information (Chong & Druckman, 2013).

Thus, research in political science regarding the kind of voters who are most susceptible to persuasive communication (i.e., politically unengaged) maps on quite nicely to existing perspectives in social psychological research indicating that counter-attitudinal information can have diverging effects on attitudes as a function of pre-existing levels of processing motivation and ability among message recipients. Despite the potential for these lines of research to generate complementary insight for understanding the psychological effects of persuasive political communication, each has thus far proceeded independent of the other.

---

[1] Taber & Lodge (2006) demonstrate that people high (vs. low) in political engagement are more (vs. less) likely to counter-argue political messages that challenge their beliefs, and Chong & Druckman (2013) demonstrate that counter-attitudinal messages also increases attitude certainty for these individuals. However, prior work has not tested the direct link between counter-argumentation, as a function of levels political engagement, to attitudinal certainty over time.

In the present research, we integrate and extend these perspectives to examine the impact of meta-cognitive resistance processes in the context of a 3-wave panel study during the 2012 American presidential election. As suggested earlier, this longitudinal methodological design is atypical in psychological research and, to our knowledge, has not been used in studies of resistance to persuasion, which have largely involved cross-sectional analyses of attitudes and behaviors in laboratory contexts. By leveraging multiple measurement points over a two-month period and manipulating exposure to messages in an online setting, this panel design allows us to establish causality more convincingly and to examine intra-individual change over time in a real world context as a function of the experimental messages and mediating processes.

Hypotheses

Because highly engaged voters have strongly held preexisting political opinions, it is unlikely that resistance to counter-attitudinal information would bolster these convictions further. Instead, what's more likely is that counter-attitudinal information has no net effect on the attitudinal certainty of the politically engaged, rendering them unpersuaded by such information and every bit as confident in their preferences. We nonetheless explore the possibility that counter-attitudinal information can *bolster* attitudinal certainty among the engaged. However, our primary predictions concerns how exposure to counter-attitudinal information *undermines* attitudinal certainty among voters with low levels of engagement, individuals who are least able to defend their attitudes against attacks—who political scientists describe as "persuadable" (Chong & Druckman, 2010; Taber & Lodge, 2006; Zaller, 1992).

First, we expect individuals low (vs. high) in political engagement to show a *decrease* in the certainty of their candidate preferences from T1 (pre-message) to T2 (post-message) when exposed to a counter-attitudinal but not a pro-attitudinal message (*Hypothesis 1*). Next, we

examine whether in the counter-attitudinal condition, a T1 to T2 *change* in certainty mediates the relationship between T1 political engagement and (a) whether respondents voted at T3 (immediately post-election) for the candidate they preferred at T1 (September; *Hypothesis 2)* and (b) polarization of candidate evaluations from T1 to T3 (*Hypothesis 3).* Finally, we seek to explain the effect of engagement on change in certainty in the counter-attitudinal message condition. To do this, as is common in meta-cognitive persuasion research, we code for the presence of active counter-arguing in respondents' open-ended cognitive responses, which were assessed immediately upon reading the persuasive message. We expect that engagement will facilitate counter-arguing (*Hypothesis 4*), and that the number of listed counter-arguments will mediate the impact of T1 political engagement on T1 to T2 change in certainty in the counter-attitudinal message condition (*Hypothesis 5*).

Method

Overview

We utilized an on-line 3-wave panel design (*T1 baseline*, September 9 to September 23, 2012, *n=410*; *T2 pre-election*, October 19-November 4, 2012, *n=375*; *T3 post-election*, November 7 to November 14, 2012, *n=373*; election day was November 6, 2012). We retained a 91% of our sample across all 3 measurement periods. The persuasive message was embedded in the Time 2 (T2) pre-election survey, having been pilot tested on an independent sample prior to administration to ensure the comprehensibility and validity of the experimental stimuli (see online appendix). The message was conveyed in an online context and targeted either the policies or personality of President Obama or Governor Romney.

Participants

The final sample of participants was composed of 295 undergraduate students (199 females, 95 males; mean age = 19.24, SD = 2.45) recruited from the research participant pool at a large public university. We excluded from analyses participants in the original sample recruited for the T1 survey who did not complete all 3-waves of the survey (n=37, 9% of T1 sample), non-U.S. citizens (n=41), and U.S. citizens who either failed to correctly identify the target of the persuasive communication (n=26) or who responded incorrectly to more than one of the instruction tests administered (described below) at each time point (n=10). We recruited as many participants as possible at T1 and did not allow new participants to enroll once the T2 survey was in the field.

Of the eligible participants, most identified as white (83.7%) and were relatively affluent (43% report a family income greater than 100K). The mean political ideology was moderate (M = 3.66, SD = 1.71; 1 = "very liberal" to 7 = "very conservative") and the mean partisan identification leaned slightly Democratic (M = 3.74, SD = 1.89; 1 = "strong Democrat" to 7 = "strong Republican"). More participants at Time 1 preferred Obama (n=192) than Romney (n=104; Table 2).

As participants were randomly assigned to a persuasive message targeting one candidate or the other, some were exposed to a *pro*-attitudinal message (e.g., Obama supporters at T1 exposed to a negative ad about Romney), whereas others were exposed to a *counter*-attitudinal message (e.g., Romney supporters at T1 exposed to a negative ad about Romney). Table 2 provides the distribution of participants who, at T1, indicated support for Obama or Romney, and whether they were assigned to a pro- or counter-attitudinal message.

Procedure

Participants completed an on-line survey at each measurement period. At the outset of

T2, participants were randomly assigned to a message targeting either the character or issue-

positions of Barack Obama or Mitt Romney. The experimental stimuli were based on publicly

available advertisements and communications used in the 2008 general presidential election and

the 2012 primary elections, and therefore reflect actual political rhetoric used by candidates,

parties, and advocacy groups. The content of the persuasive messages, along with the results of a

pilot test validating these materials on an independent sample (N = 263), is available in the

supplementary materials2. Messages were designed to be of comparable length, interest, and

comprehensibility, and to emphasize both the personal character and issue-positions of each

candidate (see supplementary materials). After message exposure, participants were instructed to

"Please take a moment and write down all of your thoughts while you were reading the

message." Participants were given the opportunity to record up to 10 independent thoughts and

were not given the opportunity to revise their responses. Two independent raters coded the

content of these responses for the presence of counter-argumentation (described below).

Measures

All continuous variables were rescaled to run from 0-1 for easier interpretation and

comparison of effect sizes. The supplementary materials include verbatim measures used at each

measurement period, including measures not included in this analysis. All measures, methods,

and exclusions are otherwise fully reported.

2 To ensure that our experimental conditions did not differ in terms of perceived bias, argument
strength, and processing requirements of the message content, we recruited 263 U.S. citizens
from the Amazon Mturk platform prior to administering the T1 survey. For participants exposed
to a counter-attitudinal message, we did not observe significant differences in perception of
message bias, argument strength, or processing motivations between messages targeting Obama
or Romney (*ps* > .2).

Administered Only at Time 1 (T1):

Vote Preference. At T1, participants responded to the following question: "If the election were held today, which of the following candidates for the President of the United States would you vote for?"

Political Engagement. This was measured by computing a measure of political knowledge (8 items) and political interest (5 items), and then taking the mean of both. This is a common approach to assessing political engagement in the political science literature (e.g., Johnston, Lavine, & Federico, 2017). *Knowledge* was indexed by the number of correct responses to such items as: (1) "What job or political office does Joseph Biden currently hold?" (2) "What job or political office does John Roberts currently hold?" (M=.54, SD=.29; Cronbach's alpha=.66). *Political interest* was indexed by the mean of responses to such items as: (1) "In your opinion, how personally relevant is the outcome of the upcoming presidential election?" (2) "How interested are you in the presidential campaign?". Higher values represent greater interest in politics (M=.49, SD=.20, Cronbach's alpha=.81). Higher values represent comparatively greater levels of political engagement (M=.51, SD=.20, Cronbach's alpha=.73).

Demographics and Political Predispositions. In addition to gender, age, and race, participants also reported their family income, partisan identification, and political ideology (see supplementary materials).

Administered Only at Time 2 (T2):

Measures administered only at T2 served as checks on the validity of the pro/counter-attitudinal message manipulation or as dependent variables. These measures were assessed immediately after participants were exposed to the experimental stimuli.

Manipulation Checks. Participants were asked to identify which political candidate was targeted by the message—those who failed this manipulation check were excluded from analyses (*n*=26).

Counter-Argumentation. Two trained independent raters content-coded the open-ended cognitive responses that participants generated on the thought-listing task following exposure to the message. Both coders were blind to the hypotheses and experimental condition. Each cognitive response was coded for the presence (coded as a 1; absence coded as 0) of each of the following five dimensions: message content or source was (a) unfair, (b) biased, (c) inaccurate, (d) not credible, or (e) containing weak arguments. We also coded for negative thoughts for the candidate opposed to the target of the message (i.e., negative thoughts about Romney if Obama was the message target, or negative thoughts about Obama if Romney was the target). Table 3 reports the average Kappa for each dimension across all thoughts (Kappa M=.53). An indicator of counter-argumentation was computed by taking the sum of the average judgment of both coders for each dimension noted above (M=.21, SD=.21).

Repeating Measures at T1, T2, or T3

Instruction tests and attention checks. Following the recommendations of Berinsky, Margolis, and Sances (2014) and Oppenheimer, Meyvis, and Davidenk (2009), we embedded a single instruction test item in each of the 3 waves to screen inattentive participants and reduce measurement error. The instruction test asked participants to "select from the options below" the "color red" (T1), "number 2" (T2), and the "color blue" (T3). Participants who selected the incorrect response option on more than one occasion were excluded from analyses (*n*=10).

Candidate Evaluations. At T1 and T3, participants evaluated the extent to which each candidate was competent, compassionate, qualified, effective, strong, likable, stubborn, reckless,

trustworthy, dishonest, and caring on a 7-point scale (after recoding the negative trait terms, T1 Romney M=3.80, SD=1.31, Cronbach's alpha=.94; T1 Obama M=4.80, SD=1.31, Cronbach's alpha=.95: T3 Romney M=4.01, SD=1.22, Cronbach's alpha=.95; T3 Obama M=4.86, SD=1.25, Cronbach's alpha= .96). To gauge *attitude polarization,* we took the absolute value of the difference score of the evaluations of Obama and Romney at each wave, producing an indicator of attitude polarization for responses at both T1 1 and T3. Items were rescaled such that higher values represent more polarized evaluations at each wave (T1 M=.32, SD=.24; T3 M=.31, SD=.23).

Vote Certainty. At T1 and T2, participants were asked to indicate how confident they were in their choice for president on a 7-point scale ranging from "1= Not at All Confident" to "7 =Very Confident" (T1 M=.71, SD=.33; T2 M=.76, SD=.30). Higher values represent greater vote certainty.

Administered Only at Time 3 (T3)

Attitude-Behavior Correspondence. At T3, participants responded to the following question: "Which candidate for the President of the United States did you vote for?" Based on self-reported vote preferences at T1, we recoded T3 vote choice to create an indicator representing attitude-behavior correspondence (0 = voted for a different candidate at T3 than was preferred at T1; 1 = voted for the same candidate at T3 that was preferred at T1).

Results

The data files and syntax are available at:

https://osf.io/t8ck9/?view_only=dd2ea70e25ce48498623c919ab5ab6d7

OLS regression was used to estimate the interaction between exposure to pro- vs. counter-attitudinal messages and political engagement on T1 to T2 change in certainty in

candidate preferences (H1) and in counter-argumentation (H4). Robust standard errors were used

to protect against heteroscedasticity and we report p-values of one-tailed tests. All analyses

collapse across the candidate targeted by the message (Obama vs. Romney). Structural equation

modeling (SEM) was performed for all mediation analyses (i.e., H2, H3, H5) with Mplus 7.4,

using maximum likelihood parameter estimation, which accommodates binary dependent

variables. No post-hoc modifications were undertaken for any model, as each indicated relatively

good fit. Analyses that treat attitude certainty or counter-argumentation as a mediator are limited

to participants exposed to a counter-attitudinal message (H2, H3, H5).

        To estimate the change in a dependent variable over time, we included its lagged value

(e.g., the value from a previous wave) as a predictor. In this way, the coefficient on engagement

estimates the impact of a *change* in vote certainty from Time 1 to Time 2 (a period of six weeks,

on average; see Finkel, 1995; Lenz, 2013). Correlations of the variables of interest, including

means and standard deviations, can be found in Table 1.

Hypothesis 1: Does Exposure to a Counter-Attitudinal Message Decrease Certainty among

the Politically Unengaged?

        First, we examine the interaction between political engagement and exposure to a pro- vs.

counter-attitudinal message on T1 to T2 change in vote certainty. Experimental condition was

dummy-coded (pro-attitudinal condition = 0; counter-attitudinal condition = 1), based on T1 vote

preferences. T2 vote certainty was regressed on this dummy-coded message variable, T1

engagement, their interaction, and T1 vote certainty. The results yielded a significant interaction

(*b*=.20 (CI= -.03, .42), *p*=.04). To explicate the interaction, T2 vote certainty was regressed on

T1 vote certainty and T1 political engagement separately for respondents in the pro- and counter-

attitudinal message conditions. For participants in the counter-attitudinal (*b*=.22 [CI= .02, 43],

p=.02), but not pro-attitudinal (*b*=.04 (CI= -.11, .20), *p=.29*) message condition, the effect of

political engagement on T1 to T2 change in vote certainty was significant, providing support for

H1. As Figure 1 suggests, unengaged respondents experienced a T1 to T2 decrease in vote

certainty as a function of exposure to a counter-attitudinal message, whereas highly engaged

respondents maintained levels of certainty from pre- to post-message exposure in both message

conditions (Table 4 summarizes the results of these models)[3].

Hypothesis 2: Does T1 to T2 Change in Vote Certainty Mediate the Effect of T1 Political

Engagement on Attitude-Behavior Correspondence for Counter-Attitudinal Messages?

In this section, we examine whether a T1 to T2 *change* in vote certainty in the counter-

attitudinal message condition mediates the impact of political engagement on attitude-behavior

correspondence (i.e., correspondence between attitudes toward the favored candidate at T1 and

reported vote choice at T3). A strength of our longitudinal design is that we avoid many of the

known problems for causal inference associated with mediation analysis on observational data

(see Bullock, Green, & Ha, 2010). In particular, we use T1 to T2 *change* in attitudinal certainty

as a mediator for the relationship between T1 to T3 variables. Thus, our independent, mediating,

and dependent variable are all measured in separate sessions, minimizing shared methodological

variance and rendering it logically impossible that our mediating variable--which, as we show

above, was manipulated experimentally by exposure to a counter-attitudinal (vs. pro-attitudinal)

message—retroactively effected our independent variable. The temporal ordering of these

---

[3] A main focus of this paper is on how and why political engagement moderates the effects of
counter-attitudinal messages on certainty, and the downstream effects of change in certainty for
political participation and polarization. Our focus is not on how exposure to pro- vs. counter-
attitudinal information on attitude change. However, we find limited evidence of attitude change
such that the counter- (vs. pro-) attitudinal messages led to slightly more negative evaluations of
the candidate attacked by the message (p = .023, b = -.09) but did not lead to change in voting
preferences (p > .2).

constructs allow for stronger claims about the mediating role of change in attitudinal certainty for the effects of political engagement on downstream outcomes.

For participants in the counter-attitudinal message, Figure 2 represents the theoretical SEM model; the fit indices suggest good fit (CFI=0.98, WRMR=.54; summarized in Table 6 in the supplementary materials). The results indicate that T1 political engagement had a significant direct effect on T1 to T2 change in vote certainty (b=.23 [CI=.08, .37], $p < .01$) and a non-significant direct effect on attitude-behavior correspondence (Exp(b)= .44 [CI=.08, 2.32], p = .21). Furthermore, T1 to T2 change in vote certainty directly increased attitude-behavior correspondence (Exp(b)=12.68 [CI=4.71, 33.78], $p < .001$). Finally, T1 political engagement indirectly predicted attitude-behavior correspondence (Exp(b)=1.77 [CI=1.16, 2.92], p = .02; Sobel = 2.17, p = .03). However, for participants in the pro-attitudinal condition, none of the direct or indirect pathways obtained significance.

Hypothesis 3: Does T1 to T2 Change in Vote Certainty Mediate the Effect of T1 Political Engagement on the T1 to T3 Polarization of Candidate Evaluations?

Here, we repeat the analysis above, but for polarization in candidate evaluations as the dependent variable. For participants in the counter-attitudinal message, Figure 3 represents the theoretical SEM models; the model is fully saturated, and the fit indices suggest good fit (CFI=0.98, SRMR=.04; summarized in Table 7 in the supplementary materials). The results indicate that T1 political engagement had a significant direct effect on T1 to T2 change in vote certainty (b= .22 [CI=.05, .40], $p < .01$). Furthermore, T1 to T2 change in vote certainty directly led to more polarized candidate evaluations from T1 to T3 (b= .18 [CI=.09, .26], $p < .001$). Finally, T1 political engagement indirectly increased the polarization of candidate evaluations

from T1 to T3 (*b*=.04 [CI=.003, .08], p = .02; Sobel = 2.15, *p* = .03). However, for participants in

the pro-attitudinal condition, none of the direct or indirect pathways obtained significance.

Hypotheses 4 and 5: Does Exposure to a Counter-Attitudinal Message Increase T2

Counter-Argumentation for the Politically Engaged, and Mediate the Effect of T1 Political

Engagement on T1-T2 Change in Vote Certainty?

We now turn to our final two hypotheses: whether in the counter-attitudinal condition,

engagement heightens counter-arguing (H4), and whether counter-arguing mediates the impact

of engagement on a T1 to T2 change in vote certainty (H5). To evaluate Hypothesis 4,

experimental condition was dummy-coded as before (pro-attitudinal message condition = 0;

counter-attitudinal message condition = 1), and T2 counter-argumentation was regressed on the

dummy-coded message variable, engagement, and their interaction. The results yielded a

significant interaction (*b*=.24 [CI= .03, .44], *p = .01*). To interpret the interaction, T2 counter-

argumentation was regressed on T1 political engagement separately for respondents in the pro-

and counter-attitudinal message conditions. For participants in the counter-attitudinal (*b*=.20

(CI= .04, .36), *p*=.01*), but not pro-attitudinal (*b*=-.03 (CI= -.17, .10), *p = .31*) message condition,

the effect of political engagement on T2 counter-argumentation was significant, providing

support for H4. Figure 4 graphically represents this analysis (Table 5 summarizes the results of

these models).

Next, we examined whether T2 counter-argumentation mediated the impact of T1

political engagement on T1 to T2 change in vote certainty. Figure 5 represents the theoretical

models and the results of the SEM; the model indices suggest good fit (CFI=.99, SRMR=0.04).

For this reason, we turn to estimates of the direct and indirect effects (summarized in Table 8 in

the supplementary materials). T1 political engagement directly increased T2 counter-

argumentation in the counter-attitudinal condition (b=.20 [CI= .04, .37], $p = .01$). T2 engagement

also had a direct effect on T1 to T2 change in vote certainty (b=.19 [CI= .02, .36], $p = .02$).

Finally, engagement had a significant *indirect* effect on T1 to T2 change in vote certainty

through counter-arguing (b=.05 [CI= -.003, .10], $p = .04$). That is, T1 engagement increased T2

counter-argumentation, and T2 counter-argumentation had a *direct* effect on T1 to T2 change in

vote certainty (b=.24 [CI= .07, .40], $p < .01$; Sobel = 2.02, p=.04).

General Discussion

We investigated the temporal course of meta-cognitive resistance processes for attitude

certainty, polarization, and behavior correspondence following exposure to counter-attitudinal

information over the course of the 2012 presidential election. Drawing upon social psychological

research on the role of meta-cognition in persuasion contexts, we generated predictions about the

effects of political communications in an online context on several outcomes of interest. A key

finding from this line of work is that the failure to resist a counter-attitudinal argument may

undermine a person's confidence in her prior opinion; in contrast, success (through counter-

arguing) in defending a position against attack can bolster attitude certainty, polarization, and

correspondence with behavior (Brinol, Petty, & Tormala, 2004; Tormala, Clarkson, Petty; 2006:

Tormala & Petty, 2004a, b, c). However, prior theory and research on meta-cognitive processes

in persuasive communication has largely neglected dynamic socio-political contexts, primarily

investigating cross-sectional effects without consideration given to longitudinal dynamics that

may account for *change*, over time, in attitudes and behavior. To our knowledge, the current

research represents the first test of these hypotheses outside a laboratory setting and also the first

investigation of the implications of meta-cognitive processes for testing the robustness of effects

over time of persuasive political communication in online platforms.

We applied the meta-cognitive approach to understanding the effects of persuasive messages within the context of an ongoing presidential campaign. Theoretically, a major strength of this study is the 3-wave panel design, which allowed us to observe the hypothesized effects of resistance to persuasion on intra-individual processes of change as a function of both individual differences (i.e., political engagement) and our experimental manipulation (i.e., pro- vs. counter-attitudinal message). Consistent with our predictions, we find that exposure to a counter-attitudinal (vs. pro-attitudinal) message undermined attitudinal certainty among participants low in political engagement. Politically engaged participants maintained high levels of certainty in their attitudes following exposure to counter-attitudinal information. Furthermore, we find that politically unengaged individuals were less likely than the more engaged to counter-argue the counter-attitudinal information and, as a result, report decreased certainty in their voting preferences (from T1 to T2), polarization in their candidate attitudes (from T1 to T3), and correspondence between their baseline (T1) candidate attitudes and reported voting behavior measured on Election Day some six week later (on average). Exposure to pro-attitudinal messages, by contrast, was inconsequential for these outcomes consistent with research on both meta-cognitive resistance and motivated reasoning processes in political domains (e.g., Taber & Lodge, 2006; Tormala & Petty, 2004d).

Together, these results indicate that counter-attitudinal persuasive political messages in online contexts can undermine the electoral attitudes of the less engaged while maintaining those of the more engaged, partially through differences in counter-argumentation. This finding provides insight into *how* engaged individuals become polarized. Among the less engaged, counter-attitudinal information weakens electoral commitments, rendering them open to effective persuasive messages and changing loyalties. They are *persuadable* voters. By contrast, among

the more engaged, such information tends to lead to a strengthening of resolve, producing more

polarized evaluations and attitude-consistent voting behavior. These voters are less immune to

campaign influence. They are (for the most part) *non-persuadable* voters. The upshot of these

dynamics is that counter-attitudinal persuasive messages may paradoxically contribute to

polarization among those with the greatest motivation and ability to process political

information.

Acknowledgements

References

Abramowitz, A. I. (2010). *The disappearing center: Engaged citizens, polarization, and American democracy*. Yale University Press.

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Thousand Oaks, CA: Sage.

Albarracín, D., & Mitchell, A. L. (2004). The role of defensive confidence in preference for proattitudinal information: How believing that one is strong can sometimes be a defensive weakness. *Personality and Social Psychology Bulletin*, *30*(12), 1565-1584.

Arceneaux, K., Johnson, M., & Cryderman, J. (2013). Communication, Persuasion, and the Conditioning Value of Selective Exposure: Like Minds May Unite and Divide but They Mostly Tune Out. *Political Communication*, 30, 213-231. https://doi.org/10.1080/10584609.2012.737424

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.B.F., Lee, J., Mann, M., Merhout, F., & Volfosvky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences,* 115 (37), 9216-9221.

Benoit, W. L., & Airne, D. (2009). Non-presidential political advertising in campaign 2004. *Human Communication*, *12*(1), 91-117.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*(3), 739-753.

Bernstein, M.F. (2013). A 'Prince of polling'. *Princeton Alumni Weekly*, 20-24.

Billig, M. (2003). Political rhetoric. *Oxford handbook of political psychology*, 222-250.

Bless, H., & Forgas, J. P. (2000). *The message within: The role of subjective experience in social cognition and behavior*. Psychology Press.

Brady, H. E., & Johnston, R. (2006). *Capturing campaign effects*. Univ of Michigan Pr.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, 98(4), 550-558.

Chen, P. et al. (2014). The Minnesota multi-investigator 2012 presidential election panel study. *Analyses of Social Issues and Public Policy*, 14(1), 78-104.

Chong, D., & Druckman, J. N. (2010). Dynamic public opinion: Communication effects over time. *American Political Science Review*, *104*(04), 663-680.

Chong, D., & Druckman, J. N. (2013). Counterframing effects. *The Journal of Politics*, *75*(01), 1-16.

Converse, P. (1964). The nature of belief systems in mass publics. In David Apter (Ed.), Ideology and discontent (pp. 206–261). New York: The Free Press.

Dardis, F. E., Shen, F., & Edwards, H. H. (2008). Effects of negative political advertising on individuals' cynicism and self-efficacy: The impact of ad type and message exposures. *Mass Communication & Society*, *11*(1), 24-42.

Faber, R.J., Tims, A.R., & Schmitt, K.G. (1993). Negative political advertising and voting intent. The role of involvement and alternative information sources. *Journal of Advertising*, 22(4), 67-76.

Finkel, S. E. (1995). *Causal analysis with panel data* (No. 105). Sage.

Fowler, E. F., & Ridout, T. N. (2013, April). Negative, angry, and ubiquitous: Political advertising in 2012. In *The Forum* (Vol. 10, No. 4, pp. 51-61).

Garramone, G.M. (1984). Voter responses to negative political ads. *Journalism Quarterly,* 61,

    250-259.

Geer, J. G., & Geer, J. H. (2003). Remembering attack ads: An experimental investigation of

    radio. *Political Behavior*, *25*(1), 69-95.

Geer, J.G. (2006). In defense of negativity: Attack ads in presidential campaigns. Chicago:

    University of Chicago Press.

Gelman, A., & King, G. (1993). Why are American presidential election campaign polls so

    variable when votes are so predictable?. *British Journal of Political Science*, *23*(04), 409-

    451.

Hetherington, M. J., & Weiler, J. D. (2009). *Authoritarianism and polarization in American

    politics*. Cambridge University Press.

Hillygus, D. S., & Shields, T. (2008). The persuadable voter. *Princeton University, New Jersey*.

Huber, G. A., & Arceneaux, K. (2007). Identifying the persuasive effects of presidential

    advertising. *American Journal of Political Science*, *51*(4), 957-977.

Hulse, C. (May 21, 2014). Democrats seek issues to lure midterm votes after races buoy   G.O.P.

    NY Times, Retrieved from http://www.nytimes.com/2014/05/22/us/politics/democrats-

    seek-issue to-buoy turnout-in-midterms.html?_r=1.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology a social identity perspective on

    polarization. *Public Opinion Quarterly*, *76*(3), 405-431.

Jasperson, A.E., & Fan, D.P. (2002). An aggregate examination of the backlash effect in

    political advertising: The case of the 1996 U.S. senate race in Minnesota. *Journal of

    Advertising, 31(1)*, 1-12.

Johnston, C. D., Lavine, H. G., & Federico, C. M. (2017). *Open versus closed: Personality,*

    identity, and the politics of redistribution. Cambridge University Press.

Kam, C. D. (2006). Political campaigns and open-minded thinking. *Journal of Politics, 68(4),*

    931-945.

Krupnikov, Y. (2011). When does negativity demobilize? Tracing the conditional effect of

    negative campaigning on voter turnout. *American Journal of Political Science*, 55(4),

    797-813.

Lau, R.R., Sigelman, L., & Rovner, I.B (2007). The effects of negative political campaigns: A

    meta-analytic reassessment. *The Journal of Politics*, 69(4), 1176   1209.

Lenz, G. S. (2013). *Follow the leader? How voters respond to politicians' policies and*

    performance. University of Chicago Press.

Lodge, M., & Taber, C. S. (2013). *The rationalizing voter*. Cambridge University Press.

Merritt, S. (1984). Negative political advertising. *Journal of Advertising, 13. 27-38.*

Newman, B.I. & Perloff, R.M. (2008). Political marketing: Theory, research and applications. In

    Kaid, L.L. (Ed.), *Handbook of Political Communication Research* (pp. 155-203). New

    York, NY: Routledge.

Oppenheimer, D., Meyvis T., & Davidenko N. (2009). Instructional manipulation checks:

    Detecting satisficing to increase statistical power. *Journal of Experimental Social*

    Psychology, 45, 867–872

Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of

    persuasion: The self-validation hypothesis. *Journal of Personality and Social*

    Psychology, *82*, 722-741.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects

    in simple mediation models. *Behavior research methods, instruments, &*

    computers, *36*(4), 717-731.

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*, 101-

    127.

Richardson Jr., G. W. (2001). Looking for meaning in all the wrong places: Why negative

    advertising is a suspect category. *Journal of Communication*, 51(4), 775.

Rideout, T.N., & Franz, M.M. (2011). *The persuasive power of campaign advertising*. Temple

    University Press.

Rucker, D. D., Tormala, Z. L., & Petty, R. E. (2004). Individual differences in resistance to

    persuasion: The role of beliefs and meta-beliefs. In Knowles, E.S., & Linn, J.A. (Eds.),

    Resistance and persuasion, 83. Psychology Press.

Rucker, D. D., Tormala, Z. L., Petty, R. E., & Briñol, P. (2014). Consumer conviction and

    commitment: An appraisal-based framework for attitude certainty. *Journal of Consumer*

    Psychology, *24*(1), 119-136.

Rydell, R. J., Hugenberg, K., & McConnell, A. R. (2006). Resistance can be good or bad: How

    theories of resistance and dissonance affect attitude certainty. *Personality and Social*

    Psychology Bulletin, *32*, 740-750.

Sides, J. & Vavreck, L. (2013). *The Gamble: Choice and Chance in the 2012 Election.* Princeton

    University Press.

Sultan, N.M. (2017, April 13). Election 2016: Trump's free media helped keep cost down, but

    fewer donors provided more of the cash. Retrieved from

    https://www.opensecrets.org/news/2017/04/election-2016-trump-fewer-donors-provided-

more-of-the-cash/

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political

beliefs. *American Journal of Political Science*, *50*(3), 755-769.

Tormala, Z.L., & Petty, R.E. (2002). What doesn't kill me makes me stronger: The effects of

resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology*,

83, 1298-1313.

Tormala, Z. L., & Petty, R. E. (2004a). Source credibility and attitude certainty: A metacognitive

analysis of resistance to persuasion. *Journal of Consumer Psychology*, *14*(4), 427-442.

Tormala, Z. L., & Petty, R. E. (2004b). Resistance to persuasion and attitude certainty: The

moderating role of elaboration. *Personality and Social Psychology Bulletin*, *30*, 1446-

1457

Tormala, Z. L., & Petty, R. E. (2004c). Source credibility and attitude certainty: A meta

cognitive analysis of resistance to persuasion. *Journal of Consumer Psychology*,

14, 426-441.

Tormala, Z. L., & Petty, R. E. (2004d). Resisting persuasion and attitude certainty: A meta-

cognitive analysis. *Resistance and persuasion*, 65-82.

Tormala, Z. L., Clarkson, J. J., & Petty, R. E. (2006). Resisting persuasion by the skin of one's

teeth: the hidden success of resisted persuasive messages. *Journal of personality and

social psychology*, *91*(3), 423.

Valentino, N., & Nardis, Y. (2013). Political communication: Form and consequence of the

information environment. In Huddy, L., Sears, D., & Levy, J. (Eds.), *Oxford Handbook of

Political Psychology,* 559-590. New York: Oxford University Press.

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge university press.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths

about mediation analysis. *Journal of consumer research*, *37*(2), 197-206.

Table 1

Correlations between all continuous variables used in analyses: All participants included

Variables	*M*	*SD*	1	2	3	4	5	5	6
1. T1 Political Engagement	.51	.20	--						
2. T2 Counter-Argumentation	.21	.21	.06	--					
3. T1 Vote Certainty	.71	.33	.31**	.02	--				
4. T2 Vote Certainty	.76	.30	.31**	.08	.73**	--			
5. T1 Candidate Evaluations	.32	.24	.31**	.00	.58**	.52**	--		
5. T3 Candidate Evaluations	.31	.23	.20**	-.004	.55**	.47**	.76**	--	
6. T3 Vote Choice (1=no-change)	.90	.29	.15*	.002	.47**	.47**	.25**	.32**	--

Note. †*p* < .10. *p < .05. **p < .01.

Table 2

Frequency Distribution of T1 Vote Preference, Separated by Assignment to Pro- vs. Counter-Attitudinal Condition

T1 Vote Preference	Target of Message Condition at T2		
	Obama	**Romney**	**Total**
Preference for Obama	89	100	**189**
Preference for Romney	56	48	**104**
Total	**145**	**148**	**293**

Table 3: Interrater Agreement (Kappa) for Evaluations of Cognitive Responses

Thought #	Biased	Unfair	Innaccurate	Not Credible	Weak Argument	Obama Negative	Romney Negative
1	0.57	0.37	0.29	0.1	0.22	0.67	0.68
2	0.34	0.37	0.37	0.25	0.31	0.45	0.7
3	0.27	0.39	0.42	0.14	--	0.75	0.42
4	0.66	0.49	0.24	--	0.4	0.72	0.85
5	0.33	0.44	0.66	--	--	0.83	0.39
6	--	--	--	--	--	0.86	0.44
7	--	--	--	--	--	1	0.8
8	--	--	--	--	--	1	0.57
9	--	--	--	--	--	0.86	0.67
10	--	--	--	--	--	1	--
Total	0.43	0.41	0.40	0.16	0.31	0.81	0.61

Table 4: *T1 to T2 Change in Attitude Certainty as a Function of Political Engagement and Counter-Attitudinal vs. Pro-Attitudinal Messages*

	T2 Vote Certainty		
	b	SE	95% CI
T1 Vote Certainty	0.66***	(0.05)	0.57, 0.75
Message Type Dummy (1=Counter-att.)	-0.12†	(0.07)	-0.25, 0.01
T1 Pol. Engagement	0.04	(0.08)	-0.12, 0.19
Message Type x Pol. Engagement	.20*	(0.11)	-0.03, 0.42
Intercept	0.29**	(0.05)	0.19, 0.40
F (degrees of freedom)	83.17 (4, 288)***		
Adjusted R^2	0.55		

	T2 Vote Certainty		
Pro-Attitudinal Message Condition	*b*	SE	95% CI
T1 Vote Certainty	0.64***	(0.07)	0.51, 0.77
T1 Pol. Engagement	0.04	(0.08)	-0.11, 0.20
Intercept	0.30***	(0.06)	0.18, 0.42
F (degrees of freedom)	54.14 (2, 153)**		
Adjusted R^2	0.49		

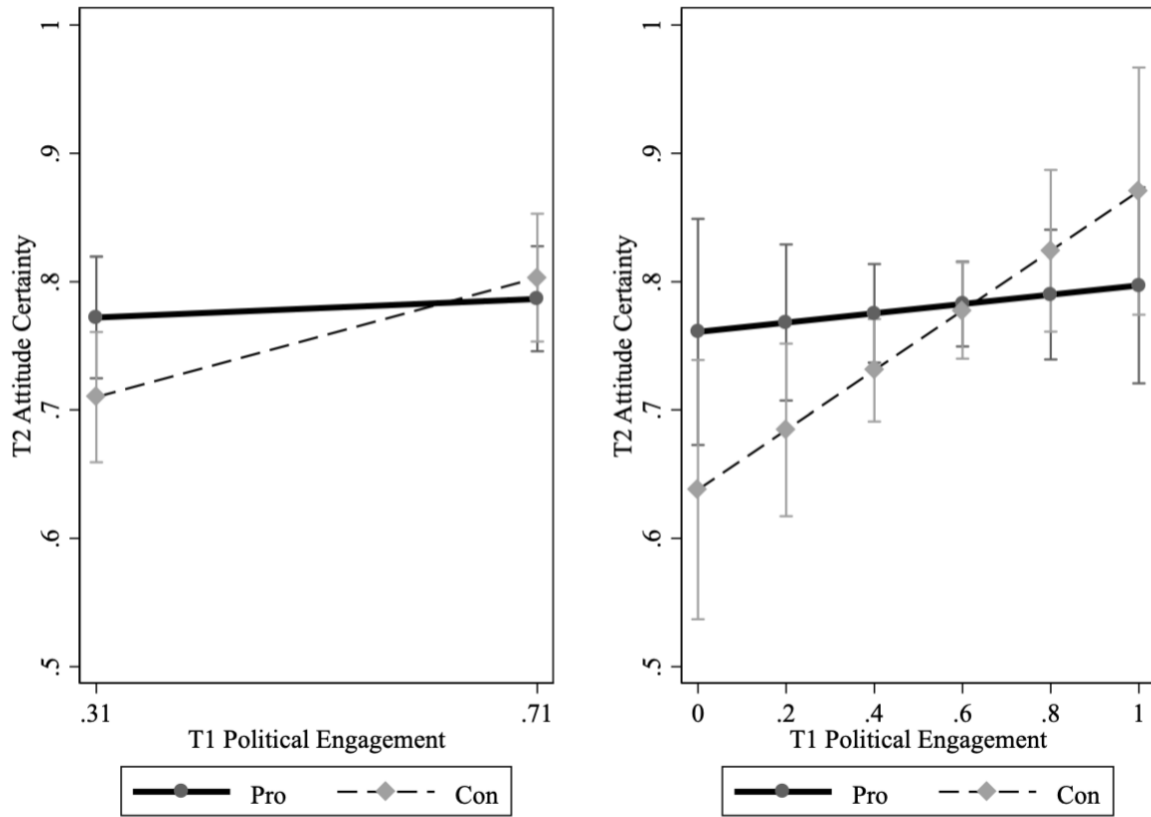|	T2 Vote Certainty		
Counter-Attitudinal Message Condition	*b*	SE	95% CI
T1 Vote Certainty	0.67***	(0.07)	0.54, 0.81
T1 Pol. Engagement	0.22*	(0.10)	0.02, 0.43
Intercept	0.16*	(0.05)	0.06, 0.27
F (degrees of freedom)	110.57 (2, 134)		
Adjusted R^2	0.59		

Note. Entries are ordinary least square regression coefficients, with robust standard errors and confidence intervals. (†$p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$)

Table 5: T2 Counter-Argumentation *as a Function of Political Engagement and Counter-Attitudinal vs. Pro-Attitudinal Messages*

	T2 Counter-argumentation		
	b	SE	95% CI
Message Type Dummy (1=Counter-att.)	-0.01	(0.06)	-0.12, .11
T1 Pol. Engagement	-0.03	(0.07)	-0.17, 0.10
Message Type x Pol. Engagement	0.24*	(0.10)	0.03, 0.44
Intercept	.17**	(0.04)	0.10, 0.25
F (degrees of freedom)	9.03 (3, 288)***		
Adjusted *R*2	0.001		

	T2 Counter-argumentation		
Pro-Attitudinal Message Condition	*b*	SE	95% CI
T1 Pol. Engagement	-0.03	(0.07)	-0.17, 0.10
Intercept	0.17***	(0.04)	0.10, 0.25
F (degrees of freedom)	0.61 (1, 154)*		
Adjusted *R*2	0.000		

	T2 Counter-argumentation		
Counter-Attitudinal Message Condition	*b*	SE	95% CI
T1 Pol. Engagement	0.20*	(0.08)	0.04, 0.36
Intercept	0.17***	(0.04)	0.09, 0.25
F (degrees of freedom)	6.39 (1, 134)		
Adjusted *R*2	0.001		

Note. Entries are ordinary least square regression coefficients, with robust standard errors and confidence intervals. (†$p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$)

Figure 1. *Effect of T1 Political Engagement x Pro/Con Messages on T1 to T2 Change in Attitude Certainty*

Error Bars Represent 95% CI

Low & High Engagement = 1 SD Below/Above the Mean of Political Engagement

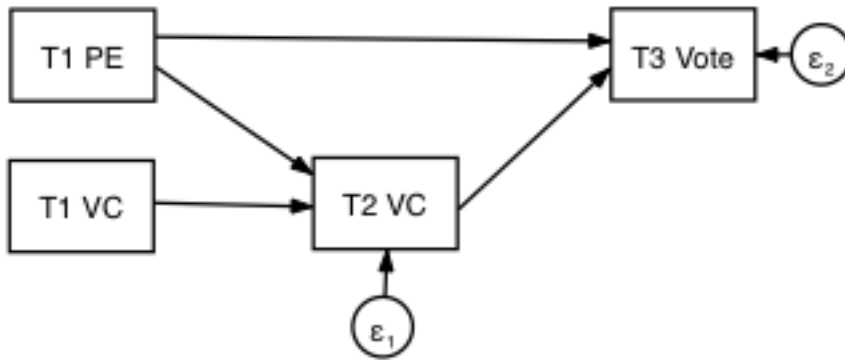*Note.* The dependent variable is T2 vote certainty while controlling for T1 vote certainty.

Figure 2. Results of SEM for Hypothesis 2. CFI=.99; WRMR=.054; chi-square=2.35; degrees of freedom=1. Sobel=2.17, p=.01. VC= vote certainty. T3 vote= T3 vote choice. PE= political engagement. e= error. Coefficient estimates for direct and indirect effects are reported in the supplemental materials.
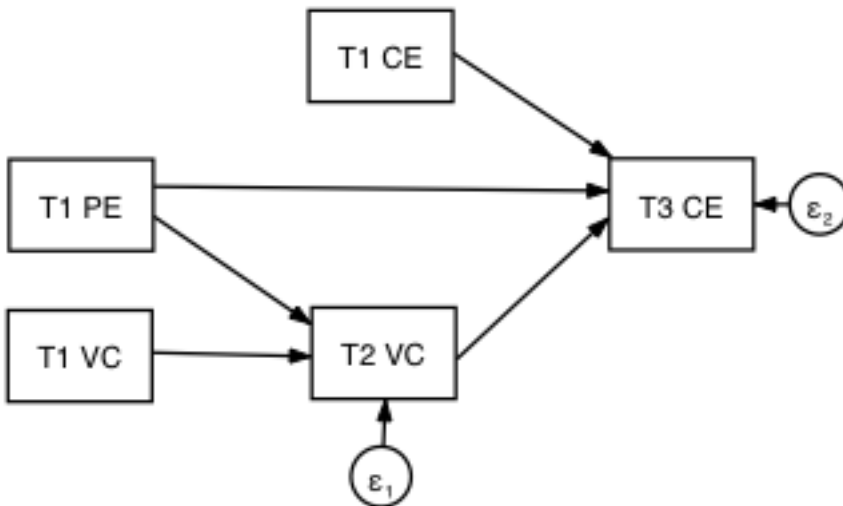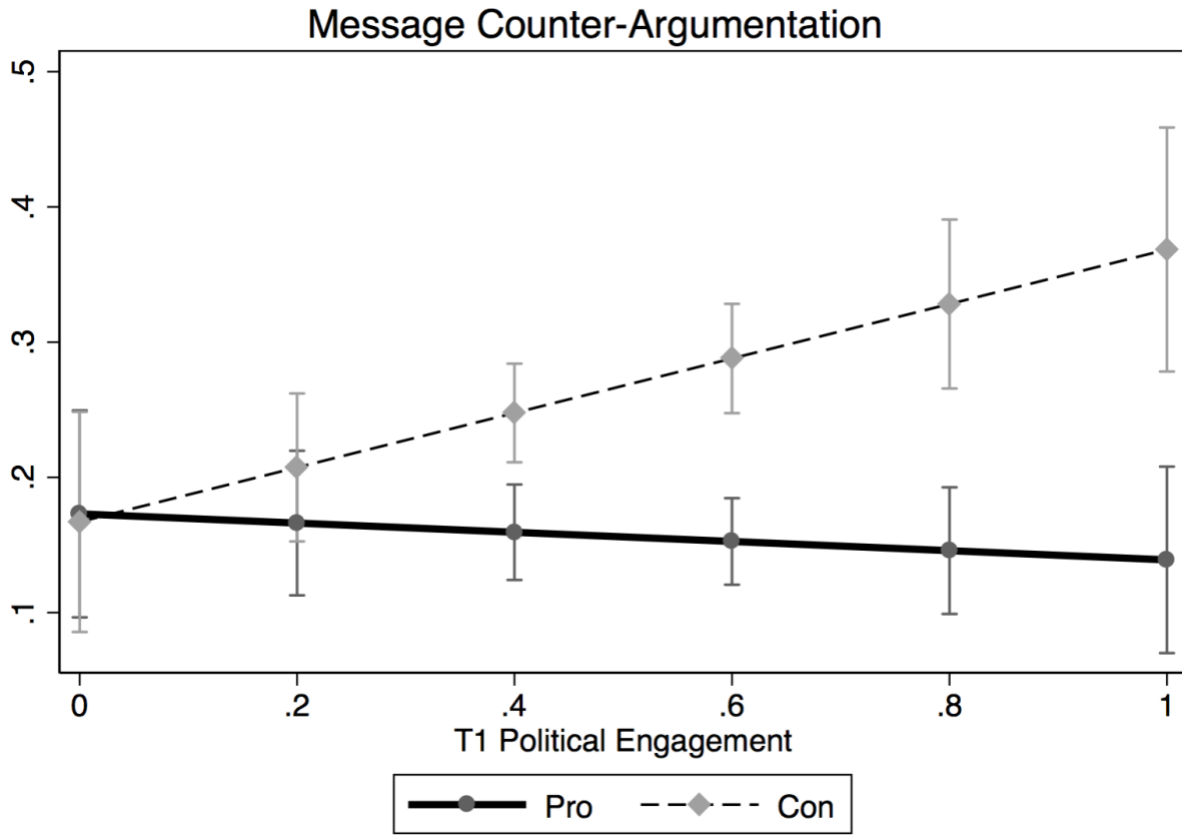

Figure 3. Results of SEM for Hypothesis 3. CFI=.98; SRMR=.04; chi-square=8.18; degrees of freedom=2; Sobel=2.15, p=.03. CE=candidate evaluations. VC= vote certainty. PE= political engagement. e= error. Coefficient estimates for direct and indirect effects are reported in the supplemental materials.

Figure 4. *Effect of T1 Political Engagement x Pro/Con Messages on T2 Counter-Argumentation*
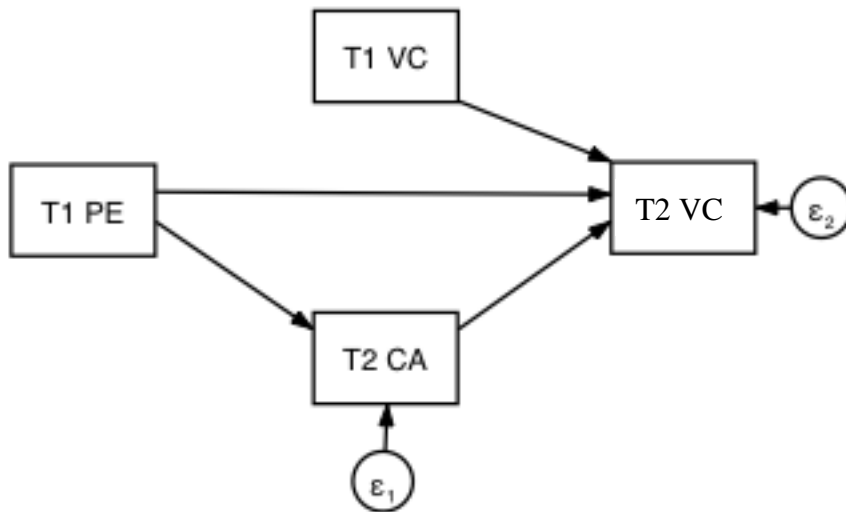
Figure 5. Results of SEM for Hypothesis 5. CFI=.99; SRMR=.04; chi-square=1.292; degrees of freedom=1; Sobel=2.15, p=.03. CA=counter-argumentation. VC= vote certainty. PE= political engagement. e= error. Coefficient estimates for direct and indirect effects are reported in the supplemental materials.