Title: Split Feelings: Understanding Implicit and Explicit Political Persuasion

Timothy J. Ryan
Associate Professor
Department of Political Science
361 Hamilton Hall, CB #3265
The University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
919-962-0403
tjr@email.unc.edu
ORCID: 0000-0002-8992-6760

Yanna Krupnikov
Professor
Department of Political Science
Social and Behavioral Sciences Building
Stony Brook University
Stony Brook, NY 11794
631-632-4166
yanna.krupnikov@stonybrook.edu
ORCID: 0000-0002-4708-0074

Abstract: Research in psychology has established that people have visceral positive and negative reactions to all kinds of stimuli—so-called "implicit attitudes." Implicit attitudes are empirically distinct from explicit attitudes, and they appear to have separate consequences for political behavior. However, little is known about whether they change in response to different factors than explicit attitudes. Identifying distinct antecedents for implicit and explicit attitudes would have far-reaching implications for the study of political persuasion. We hypothesized that implicit attitudes would change primarily in response to political advertisements' emotional valence, but this turned out to be wrong. In contrast, our next hypothesis that implicit (but not explicit) attitudes would improve in response to increased familiarity with an attitude object was supported across several tests. Aside from this finding, our studies illustrate how routine pre-registration helps researchers convey what they learned from each test—including when predictions are not borne out.

Key words: implicit attitudes, mere familiarity; Affective Propositional Evaluation Model; persuasion, pre-registration

Acknowledgments: We thank Elijah Cole Stanley, Caitlin Young, Marianna Abdalla, and Derek Robinson for indispensable help producing video advertisements. We thank J Ehlinger, Josh Gubler, Eric Groenendyk, Chris Johnston, Lily Mason, Rahsaan Maxwell, Keith Payne, Kathleen Searles, our peer reviewers, and seminar participants at Duke University and the University of Michigan for feedback at earlier stages of this project. We thank Erika Franklin Fowler and the staff at the Wesleyan Media Project -- Laura Baum, Markus Neumann and Pavel Oleinikov -- for sharing data. We gratefully acknowledge support from the University of North Carolina Junior Faculty Development Award.

One of the most effective television ads to air during the 2016 campaign season was a spot created for Bernie Sanders' unsuccessful primary run. The ad, titled "America," is striking for its minimalistic style. It is sixty seconds long, but aside from a legally required statement in the final three seconds ("I'm Bernie Sanders, and I approve this message"), it includes no narration whatsoever. No list of accomplishments; no policy statements; and no criticism of his opponents. Instead, there are fleeting shots of regular-looking people cheering and high-fiving during Sanders campaign events, interspersed with a mélange of Americana: farmers milking cows, boats in a harbor, and so forth—all set to an uplifting Simon & Garfunkel melody. Although the ad offers no explicit reasons whatsoever for why the viewer should support Sanders, a professional analysis of viewer reactions dubbed "America" to be "the ad that moved people the most" (Vavreck 2016).

"America" is an extreme example, but few people would dispute that campaign ads are routinely thin on substance—and the academic literature agrees (e.g. Kaid & Johnston 2001). Still, the scholarly understanding of how messaging elements *beyond* substance work remains minimal. For many years, research focused primarily on framing effects—the capacity of political messages to make some substantive considerations more salient than others (e.g. Chong & Druckman 2007a). This focus certainly captures one important facet of campaign messaging, but as Brader (2006) discusses, it puts aside an array of other psychological mechanisms via which campaign advertisements might be influential. As researchers have begun to systematically study the emotional components of political messaging, they have discovered that music, imagery, metaphors, and other non-substantive components of political advertising indeed influence voters in a host of ways (see Brader & Marcus 2013 for a review).

Our purpose herein is to increase understanding of how the emotional components of political advertising work. So far, the study of non-substantive facets of campaign messaging has relied almost entirely on self-reports—about the emotions citizens experience and their liking for

particular candidates relative to others. While critical, this work puts aside an extensive body of research in psychology finding that many aspects of human behavior are influenced by visceral positive and negative reactions that are not conducive to being measured with self-reports—so-called implicit attitudes (e.g. Banaji & Heiphetz 2010). Implicit attitudes appear to have an independent influence on political judgments (e.g. Ryan 2017), and as we discuss below, they might be especially ripe for manipulation via political advertising. As such, the current literature might be missing an entire domain of important campaign effects.

The largest contribution of the studies below is to demonstrate the critical relevance of implicit/explicit attitude *dissociation* to the enterprise of political persuasion. Each of our tests serves to illustrate that implicit and explicit attitudes can change independently, and in response to different stimuli. As such, this work provides an impetus to revisit a wide array of findings related to political messaging, asking what effects the field might have missed, given its almost exclusive focus on explicit attitudes. Secondly, we begin to characterize the kinds of campaign messages that move implicit versus explicit attitudes. We find that, while implicit and explicit attitudes are interrelated and many stimuli affect both domains, interventions that increase familiarity with a candidate work primarily by influencing *implicit* attitudes. Psychologists have observed such "mere exposure" effects going back at least to Zajonc (1968). The contribution of this manuscript is 1) to show that mere exposure effects act primarily on implicit, rather than explicit attitudes and 2) to elucidate the important role that familiarity effects and implicit attitudes play in the context of political campaigning. For instance, as we discuss in our conclusion, the patterns we uncover may point to a key psychological mechanism that makes common low-information campaign tools (such as lawn signs) effective.

This slate of studies is also distinctive in that it reflects a gradual change in our thinking. At the outset, we subscribed to a framework that is prominent in Psychology: the Associative-

Propositional Evaluation (APE) Model (Gawronski and Bodenhausen 2006; 2014). However, our initial tests did not support this model. Instead, they seemed consistent with the "mere familiarity" effects we allude to above. Thus, we put the APE Model aside, and developed new experiments focused on testing mere familiarity relationships, finding support. We documented this evolution in our thinking in five distinct pre-registration documents. Because these documents clarify the theoretical focus of each test and what we learned from it, this series of studies illustrates how commitment to routine pre-registration helps researchers understand and communicate research findings—including for studies where predictions are not borne out.[1]

Implicit and Explicit Attitudes

The distinction between implicit and explicit attitudes relates to the two-part process that psychologists believe characterizes human mental evaluations (Sherman et al. 2014).[2] First, when a person encounters an attitude object (another person, a kind of food, or almost anything else), one set of mental processes rapidly draws on associations in memory and generates positive or negative affect. These processes are termed "automatic" because they are distraction-proof: they occur irrespective of motivation and any concurrent mental processes (Bargh & Chartland 1999). Second, a separate set of mental processes forms an evaluation. These processes are termed "controlled" because they engage higher-order cognition and have the potential to revise the automatic response (Fazio & Olson 2008).

[1] Our thinking in this regard was influenced by Platt's classic (1964) essay on "strong inference," which calls for iterative tests focused on specific, narrow theoretical propositions.

[2] The dual process framework we describe here is similar to, and shares common ancestry with, the two-systems framework that is now a major paradigm in Psychology (Kahneman 2011). However, the two-systems framework is somewhat broader in scope, as it encompasses not only attitudes, but also judgments, habits, and heuristics.

Answering a survey question requires higher-order cognition.[3] For this reason, self-reports by definition capture evaluations that have been influenced by the controlled processes (Gawronski & Bodenhausen 2007, 691). To capture more visceral responses before they are influenced by higher-order mental processes, psychologists devised several measures of implicit attitudes (Petty et al. 2008 for an overview). The most popular of these is the Implicit Association Test (IAT) (Greenwald et al. 1998). The IAT is a latency-based measure in which subjects are asked to associate objects from categories of interest to a researcher (e.g. cans of Coke vs. cans of Pepsi) with positive and negative words. By examining the speed with which subjects make these classifications across several trials, it is possible to determine whether they have stronger visceral positive associations with one category (Coke) than the other (Pepsi).[4]

When measured within the same individual, implicit and explicit attitudes are routinely correlated with each other—but not to the extent to be considered different manifestations of one mental construct. The correlations vary by attitude domain, but typically range from $r = 0.10$ to $r = 0.54$ (Greenwald et al. 2009, Table 3; see also Greenwald & Nosek 2008). This divergence is significant, since it opens up the possibility that each class of attitude might predict behaviors that the other does not.

And so it seems they do. Across dozens of studies, in almost every instance in which implicit and explicit measures are used to predict the same behavior simultaneously, each has incremental predictive validity relative to the other (Greenwald et al. 2009). Put another way, this research suggests that implicit and explicit attitudes may have separable effects on behavior. Characterizing *which* behaviors are governed by implicit versus explicit attitudes remains a vibrant area of research.

[3] Even if one is engaged in satisficing, there is still some basic translation from thought to response option – though this translation may be cursory (Cannell et al. 1981).

[4] See Lane et al. (2007) for an excellent general overview of the IAT.

There is some evidence that implicit attitudes better predict "uncontrolled" behaviors—spontaneous, non-deliberative, non-verbal actions such as physical posture (Dovidio et al. 2002; Quillian 2008), though this pattern might be limited to socially sensitive topics, such as race and discrimination (Greenwald et al. 2009, 30). In the political domain, there is evidence that implicit attitudes about candidates and parties predict voting decisions, policy support, and biases in the processing of political information—though the significance of implicit attitudes might depend on the intensity of *explicit* attitudes (Lundberg & Payne 2014; Pérez 2016; Ryan 2017).

Research on the political consequences of implicit attitudes should continue, but our objective herein is to make progress in a separate research area: the ways in which elites can use political messages to manipulate implicit attitudes. Research on the antecedents of implicit attitudes is far less extensive than research on consequences—both in the study of political behavior and in psychology more generally.[5] Yet, if we are to comprehend what role implicit attitudes play in the broader political system, a crucial part of the account will be to characterize what tools political actors can use to influence implicit attitudes. We next turn to what the research in social and cognitive psychology has to say on this matter.

Associative and Propositional Routes to Attitude Change

Implicit attitude change, as well as its relationship to explicit attitude change is an active area of research in Psychology. There are several competing theoretical models, though also some points of consensus (Cone et al. 2017, for a review). We theoretically ground our investigations by focusing on one model in particular: the Associative-Propositional Evaluation (APE) Model proposed by

[5] Though see Alberston (2011) for a study showing that exposure to a religious speech improves implicit attitudes toward George W. Bush among Christian—but not non-Christian—respondents.

Gawronski & Bodenhausen (2014). We focused on the APE for two reasons. First, it reflects one important point of consensus: that implicit attitude change is best understood in the dual-process framework, wherein implicit attitudes represent fast, automatic, uncontrolled responses, while explicit attitudes reflect the output of slower, controlled processing. Alternatives to the APE, such as the Meta-Cognitive Model (Petty et al. 2007), still work from the premise that implicit and explicit attitudes arise from two distinct (though potentially interrelated) mental systems.[6] Second, it is arguably the most prominent model in the literature, and draws on more than a decade of empirical work (Gawronski & Bodenhausen 2006; 2018). This extensive body of research helped us to understand testable implications for political messaging.

The key distinction in the APE model is between *propositional* and *associative* mental processes. Propositional processes are a form of higher-order cognition and focus on assessing whether propositions are true or false. To form an evaluation, the brain assesses the truth value and logical interrelationship of propositions relevant to the judgment at hand. For instance, a person would be engaged in propositional thinking if she reasoned, "Higher taxes on corporations are desirable. Bernie Sanders favors higher taxes on corporations. Therefore, I should support (or be more likely to support) Bernie Sanders." The end result of reasoning like this, the APE posits, is what explicit attitude measures capture.

[6] The main points of disagreement among the models concern the nature of the interrelationship and how to classify what kinds of stimuli will be processed by each (Cone et al. 2017). In at least one case, we suspect that the differences among models are exaggerated. Independent research teams arrived at similar conclusions, but described their ideas differently. For instance, Gregg et al. (2006) attribute explicit attitude change to "abstract supposition" and implicit attitude change to "concrete learning." The differences compared to propositional versus associative mental processes (APE's focus) strike us as largely semantic.

Where propositional processes focus on truth and falsity, associative processes focus on positivity and negativity. As people interact with the world, the stimuli they encounter bring to mind the feelings, positive and negative, that previously co-occurred with similar stimuli—the principle that neurons that "fire together, wire together."[7] For instance, a person who has previously had favorable interactions with dogs will, upon seeing a new dog, experience positive affect, since for this person "dog" is associated with positive experiences such as cuddling on a couch. But a person who has been bitten by a dog in the past might have a contrary response, as the stimulus would evoke memories of pain and fear. Importantly, these two people might still agree on the truth or falsity of a specific proposition, such as "this particular dog is friendly." Associative processing helps explain why they might have very different gut reactions nonetheless.

Where propositional processing is costly (in terms of cognitive resources) and is posited to occur only when a person has sufficient motivation, associative processing occurs quickly and automatically, irrespective of motivations. Following this reasoning, Gawronski and Bodenhausen argue that implicit attitude measures capture the outputs of associative processes—gut responses that have not yet been refined by higher-order cognition.

The APE Model also provides guidance for how implicit and explicit attitudes change. Because explicit attitudes are based on propositional reasoning, they should change in response to seemingly valid information and arguments—things that cause a person to accept new premises that are relevant to a judgment. Moreover, explicit attitudes can potentially change rapidly, such as if new information sharply undermines the previous basis for a judgment. For example, suppose that a Democrat saw a political activist with a red baseball cap and white lettering that appeared to be

[7] This phrase is sometimes wrongly attributed to Hebb (1949). In fact, it appears to have been coined by Schatz (1992, 64) summarizing Hebb's ideas.

Donald Trump's "Make America Great Again!" slogan. The person would likely reason that the activist is a Trump supporter and dislike him. But suppose that upon closer inspection, the lettering turned out to be a parody, saying, "Make Obama President Again!" The construal of the activists' political loyalties would change drastically, and explicit attitudes about him likely would as well.

Where the APE Model sees explicit attitude change as grounded in the assignment of truth values, it sees implicit attitude change as grounded in associations built up via conditioning—the pairing of a stimulus and a positive or negative response. Appraisals about truth are beside the point here; what is relevant is patterns. Researchers commonly use "evaluative conditioning" to induce implicit attitude change—pairing attitude objects like "youth" and "elderly" with positive and negative words on a computer screen, for example, can affect the valence associated with these characteristics (Karpinski & Hilton 2001). There are no claims here about what old and young people are like; just the pairing with either positive or negative experiences.

The APE Model suggests that a single stimulus might have very different effects on implicit versus explicit attitudes, depending on the inferences and associations a person draws from the stimulus. For instance, in one test, participants underwent a procedure in which a neutral stimulus (a fictional creature) appeared on a computer screen two seconds before the end of an unpleasant sound (horrified screaming). This procedure caused explicit attitudes toward the creature to improve—presumably because the participants reasoned that the appearance of the creature caused an end to the sound. But it simultaneously caused implicit attitudes toward the creature to worsen—presumably because the creature appeared while a sound was causing an unpleasant feeling (Moran & Bar-Anan 2014).

In sum, the APE is a promising theoretical perspective with a substantial evidence base in pure Psychology. It posits implicit and explicit attitudes to change via independent routes, one focused on positive and negative associations, and other focused on assessments of truth and falsity.

These premises might clarify how political practitioners use different communication tools to influence separate (but both important) attitudes that guide citizen behavior. However, APE's implications for political messaging strategy have not been directly tested. We begin our investigations with a study focused on one provocative implication of the APE Model, as it might manifest in a political context.

Study 1: How Does Emotional Content Affect Implicit and Explicit Attitudes?

As we discuss above, the APE model posits that explicit attitudes change in response to information that can be assessed as true or false, while implicit attitudes change when a stimulus is paired with positive or negative experiences. Applied to political communication, a reasonable interpretation of these propositions is that the informational content in a campaign message acts primarily on explicit attitudes, while the affective content in the message acts primarily on implicit attitudes. Study 1 tests these ideas. We exposed participants to advertisements where the information provided by the ad was held constant, but the affective content varied. As we document in our pre-registration (available in the Supporting Information), our expectation was that high-quality affective content would improve implicit attitudes. Because the advertisements used provided almost nothing by way of substantive political information that could be used to gauge a candidate's political positions or likely behavior, we also predicted null effects on explicit attitudes.

Procedures. Testing the expectation above required exposing participants to ads that varied in terms of the emotions they induce, but not the information they provide. To create such ads, we follow the basic strategy used by Brader (2006) as well as Albertson and Gadarian (2015): we created political advertisements that had an identical script, but which varied in terms of other audiovisual elements (music and imagery). The audiovisual component is what we use to manipulate the affective content of the advertisement.

Creating advertisements suitable for measuring implicit attitudes toward a political candidate presented several additional challenges. One choice we faced was whether to present ads focused on a real or fictional candidate. We opted for a fictional candidate—Mike Harper—because we expected that implicit attitudes about a familiar figure would be more difficult to move (see Gregg et al. 2006) and also because focusing on a fictional candidate allowed us to more tightly control aspects of the advertisements to make for a clean test.

Our objective in writing a script for our advertisement was to fill time with a realistic narration, but to provide as little actual substantive information relevant to a vote choice as possible, thus making experimental participants who saw an ad and control group participants who saw no ad as comparable as possible in terms of the information they received about Mike Harper. Reviewing real advertisements provided plenty of examples of pap and pablum to fill this role. We wrote a script—included as SI Section 2—in which Mike Harper offers some bland biographical details—he worked at a hot dog stand in the town where he grew up and believes in hard work—but takes no explicit policy positions. He bemoans dysfunction in Washington, but does not blame one political side over the other, and does not mention his own partisanship. We found it impossible to scrub script elements that might convey, via stereotypes, *some* inkling of Harper's political leanings, but we were careful to balance elements that might be construed as liberalizing with ones that might be construed as conservatizing. For instance, Harper talks about being a teacher (possibly a liberal stereotype), but he also discusses growing up in a small town and being the son of a war veteran (possibly conservative elements). As we discuss in more detail below, the results of an external manipulation check confirm that Harper's political affinities were ambiguous.

To film the ad, we hired a professional videographer and a local actor to play the role of Mike Harper, as well as local university students to serve as "extras." We spent two days filming scenes around the local community—outside a local school (ostensibly where Harper works), in a

classroom and faculty lounge (also ostensibly at Harper's high school), and in an author's living room, kitchen, and car (all ostensibly Harper's). We required *two* acceptable takes of each scene. One take, which was destined for the *high-quality ad condition*, was shot with professional-grade staging, lighting, and delivery. In the other take, destined for the *low-quality ad condition*, the actor delivered the identical script, but he was positioned against a neutral background in the same environs (usually a nondescript wall), the lighting was more neutral, and his delivery was more dry. Although we use the term "quality" for ease of discussion, we underscore that this refers to the visual components within the video, rather than the *video itself* (i.e. both were equally clear when viewed). Because, as we note above, implicit attitudes are thought to change through the accumulation of positive associations, we filmed ads that were between four and five minutes long.[8]

[8] Thus, our ads are considerably longer than traditional 30-second television spots. We considered this feature necessary to deliver a sufficient dose of emotional content. Real advertisers accomplish as much by repeating ads multiple times over days or weeks, but doing so was not practical within our design. We must leave questions related to dosage timing—a short ad repeated many times versus a long ad delivered once—for future studies. As concerns naturalism, however, we note that the new emphasis on online advertising (Fowler et al. 2021) affords campaigns much greater leeway as concerns ad length (Borah et al. 2018; Ridout et al. 2015). We examined the distribution of advertisement length in a newly compiled dataset of political advertisements run by candidates for federal office, including President, on Facebook and YouTube in 2020. We found that approximately 26% of Facebook advertisements and 35% of YouTube advertisements are more than 30 seconds long. To this point, our rough model for the video, in terms of feel and style, was a biographical video produced by Senator Ben Sasse, titled "The Outsider," which is nearly six minutes long. Further, in a study of YouTube campaign advertising, Ridout et al. (2015) find that "no linear relationship between ad length and viewership; viewers appear not to be turned off by long videos" (245). We thank Laura Baum and Erika Franklin Fowler for providing us with data on online advertisement length in the 2020 campaign.

As we discuss below, the tool we use to measure implicit attitudes toward Mike Harper—an Implicit Association Test (IAT)—relies on words and symbols that can be associated with the candidate. For this reason, we were careful to incorporate four suitable IAT stimuli into the ads—and to make them equally prominent in the low-quality and high-quality ad. The four ad elements designed for incorporation into an IAT were 1) Mike Harper's face, 2) Mike Harper's name, 3) Mike Harper's campaign slogan ("Start small, go far"), and 4) Mike Harper's campaign logo, which prominently features a compass. (In the ad script, Harper says he uses his father's compass as a campaign logo as a reminder to "stay on course.") Figure 1 presents still images taken from parallel sections of the low-quality and high-quality ad.

We conducted several instrumentation checks on external samples to verify important properties of the ads we filmed. First, because we hoped to isolate the effect of affective content from partisan and ideological considerations, we ensured that Harper's political leanings were ambiguous. Second (and more critical), we ensured that Harper's political leanings were indistinguishable, comparing the low-quality ad to the high-quality ad. (Otherwise, treatment effects would be confounded with something about music and imagery that conveyed political leanings.) Third, we ensured that low- and high-quality ads were comparable in other important respects: equally memorable, equally realistic, and the features we intended to use in the IAT (the slogan and logo) were remembered equally well. Fourth, we confirmed via a post-hoc check that viewers actually experienced differing degrees of positive affect while watching the ads, in accordance we with expectations. The ads passed these checks, and we report the relevant results in the SI (Section 3).

[Figure 1 here. Title: Screen Captures from Campaign Videos. No caption.]

[Figure 2 here. Title: Instructions for Implicit Association Test. No caption.]

For the main study, we incorporated the advertisements described above into a survey instrument. Participants—undergraduates who participated for course credit[9]—began the study by viewing either the low-quality ad, the high-quality ad, or, if randomly assigned to the control condition, no ad at all. Next, participants completed an IAT focused on Mike Harper. The IAT is a common procedure in which implicit attitudes are measured via participants' response latencies in using keyboard taps to pair target stimuli (items associated with an attitude object of interest, discussed above) with positive and negative words. Across seven blocks, respondents had to pair items related to Mike Harper with either positive words (e.g. joy, love, smile) or negative words (agony, terrible, horrible), with response latencies measured in milliseconds. The order of the pairing (Harper associated first with positive words, or negative) was counterbalanced by a random assignment. The IAT procedure standardly generates a D-score, which is the average response latency (excluding practice rounds) when Harper stimuli are paired with positive words, minus the latency when paired with negative words, divided by the pooled standard deviation across all trials.[10] The IAT standardly measures implicit preference for one target relative to another target, as in Coke vs. Pepsi, Democrat vs. Republican, or flowers vs. insects. Because we did not provide participants with any information about Harper's political competition, we measured implicit attitudes relative to a neutral target—images of gray shapes. Figure 2 shows the IAT instructions presented to

[9] We agree with Druckman & Kam's (2011) assessment that students represent a useful data source—especially for studies, like ours, where the objective is to understand psychological mechanisms, rather than to estimate averages in some larger population (such as the United States). This said, Studies 2 and 3b below use non student samples, and Section 6 of the SI tests for possible heterogeneous effects.

[10] As is standard, we exclude from analysis 3.7% of respondents (N=17) who appear to be "button-mashing"—they had more than 10% IAT trials that were implausibly fast (faster than 300ms).

participants.[11] Measured this way, D-scores about Mike Harper range from -0.985 to 1.280, are roughly normally distributed, have a mean of 0.263, and a standard deviation of 0.452.

Immediately after the IAT, participants reported their explicit attitudes toward Mike Harper by answering the question, "How much did you like Mike Harper, the candidate depicted in the ad?"[12] There were seven response options ranging from "Dislike him a great deal" to "Like him a great deal." We scale responses such that the most negative feelings toward Harper take a value of 0, the most positive feelings take a value of 1, and other responses take intermediary values. After reporting feelings toward Harper, participants reported perceptions of his political leanings (ideology and partisanship)[13] and they answered questions measuring respondent demographics.[14]

Participants had one week in which they were allowed to complete the survey instrument, after which this part of the study closed. Two weeks after the initial survey instrument, all

[11] Because respondents randomly assigned to the control group had no exposure to Mike Harper before the IAT, their instructions including the following sentence: "In this IAT, one of the categories is 'Images related to Mike Harper,' a candidate for public office. Images related to Mike Harper are displayed below."

[12] For the control group, this question was, "Without knowing much about him, how much did you like Mike Harper, the candidate you saw?"

[13] Similar to the external manipulation check, the random assignment had no effect on Harper's perceived ideology. Scaled from 0 = extremely liberal to 1 = extremely conservative, Harper's ideology was indistinguishable in the control condition (M=0.545; SE = 0.017) compared to both the Low-quality condition (M=0.553, SE = 0.017) and the High-quality condition (M = 0.535, SE = 0.017, ns). Similarly, the random assignment had no influence on his perceived party affiliation. Respondents guessed whether they supposed Harper was a Democrat, a Republican, or reported that they didn't know what party he belonged to. By a chi-squared test, responses to this question were comparable across conditions ($\chi^2(4) = 7.35$, p=0.12).

[14] Our demographics were measured post-treatment; however, we do not condition any of our treatment effects on demographic factors, rather we use the demographic components simply to track the composition of our sample.

participants were invited to complete a follow-up questionnaire—also for course credit. This follow-up questionnaire was much shorter: it simply administered the IAT and explicit liking measures a second time, testing the durability of any effects. Both waves were required for course credit, and attrition from Wave 1 to Wave 2 was low: 94% of participants who completed Wave 1 also completed Wave 2.

We implemented this full paradigm in two separate semesters, in the Fall of 2017 (N=312) and the Spring of 2018 (N=158). Because the procedure was the same, and because splitting the data by semester uncovers the same patterns, we pool the administrations in the analyses below, resulting in an N of 470.

Results. Our primary interest is in how the random assignment influenced implicit and explicit liking of Mike Harper. Figure 3 presents the main results, which are broken down by implicit and explicit measures, as well as survey wave. In Wave 1, the High-quality ad improved explicit attitudes relative to the control ($p<.001$), but the Low-quality ad was indistinguishable from the control ($p=0.78$). Both ads improved implicit attitudes relative to the control ($p<.01$ for both contrasts). But the Low-quality ad was not distinguishable from the High-quality ad ($p=.92$). By Wave 2, the effects for explicit attitudes had become more muted, though the High-quality advertisement was still distinguishable from control ($p<.02$). In Wave 2, no implicit attitude contrasts were statistically significant.

Discussion. This pattern of results does not conform to the pre-registered expectations we derived from the APE Model. We predicted no strong effects on explicit attitudes, since the ad provided little substantive information (and identical substantive information in comparing the Low- to the High-quality ad). But the High-quality ad substantially improved explicit attitudes. Moreover, although our prediction that the Low-quality ad would improve implicit attitudes relative to the

control was supported, we also predicted that the High-quality ad would improve implicit attitudes relative to the Low-quality ad—a result that did not occur.

Although the APE model was not supported, this test has an important result that underlines the promise in studying implicit attitudes: the treatment conditions moved implicit and explicit attitudes *separately*. We see this in contrasting the Low-quality ad to the Control, where implicit attitudes are improved while explicit attitudes remain unchanged. We see the result again in contrasting the High-quality ad to the Low-quality ad, since explicit attitudes are improved while implicit attitudes remain essentially unchanged. Thus, we find two clear examples of implicit/explicit attitude dissociation, and evidence that the two types of psychological orientations change in response to different influences. A third sign of implicit/explicit attitude dissociation is the greater durability of treatment effects on explicit attitudes that can be seen via the right-hand column. This result is a possible hint that political appeals targeting implicit attitudes require repetition to stick—a point to which we return in our concluding section.

An Alternative to the APE Model: Mere Exposure Effects

After Study 1, we conducted a second test of the APE Model. The advertisement conditions above were crossed with a text-based manipulation providing facts about Mike Harper—an effort to separately manipulate propositional and associational processing about him. We drew participants from Amazon Mechanical Turk (MTurk), providing some ability to extrapolate beyond the university students.[15] Once again, pre-registered expectations derived from the APE Model were not supported. We report this study (Study 2) fully in the SI.

[15] MTurk is a crowdsourcing service where human participants complete tasks for small payments. See Coppock (2019) and Coppock et al. (2018) on the generalizability of studies conducted on MTurk.

Looking across these two studies, we noticed a pattern: exposure to *any* advertisement about Mike Harper improved implicit attitudes toward him, relative to a pure control condition (as seen in the bottom-left panel of Figure 3). The same is not true of explicit attitudes. For instance, mere exposure to the low-quality advertisement did not move explicit attitudes in Study 1.

The pattern for implicit attitudes, we realized, is aptly described by a framework with a long history in social / cognitive psychology. In a seminal paper, Robert Zajonc (1968) used the term "mere exposure" effects to refer to the tendency for attitudes toward a stimulus to be enhanced whenever the stimulus becomes "accessible to the individual's perception" (1). Zajonc had in mind patterns such as the tendency to rate fictional words, jagged polygons, and unknown Japanese ideographs more favorably, the more times a person has seen them. In the ensuing years, more than 400 studies would document mere exposure effects in a wide range of contexts (Bornstein & Lemley 2017), and the principle would become a cornerstone approach in product marketing (e.g. Fang et al. 2007).

[Figure 3 here. Caption: Explicit attitudes are scaled from 0 = disliking Mike Harper to 1 = liking Harper. Implicit attitudes range from -0.985 to 1.280 in Wave 1 and -0.886 to 1.288 in Wave 2. For implicit attitudes, negative scores reflect implicit dislike of Harper, and positive scores reflect implicit liking. Whiskers represent 95% confidence intervals.]

Mere exposure effects are more than just a recurrent empirical pattern. There are reasons to expect them to emerge more reliably for implicit than explicit attitudes. Years after noting the phenomenon, Zajonc explicitly connected it to the "two systems" framework that was beginning to take hold in psychology. He argued that raw positive or negative affect emerges from a set of mental processes that is both uncontrollable and independent of those associated with effortful thinking. Such attitudes can be enhanced, he thought—even in the absence of information—because certain

neurophysiological systems are designed to identify stimuli that, in the past, have not proven to be threatening (Zajonc 1980). The two-systems framework also would acquire the status of conventional wisdom in Psychology (e.g. Kahneman 2013). For instance, it undergirds Petty & Cacioppo's Elaboration Likelihood Model of persuasion, which describes a peripheral route to persuasion in which attitudes change "without necessitating scrutiny of the true merits of the information presented" (1986, 125).

These insights gave us license to posit that the implicit attitude patterns we have presented so far are not quirks, but instances of a simple familiarity induction having an effect right where Zajonc might have predicted: in the raw positive and negative associations measured by the IAT. We designed Study 3 as a more focused test of this idea.

Studies 3a and 3b: Does Mere Exposure Improve Implicit Attitudes?

Although the results in Studies 1 and 2 are consistent with an effect of mere exposure on implicit attitudes, they are not an especially demanding test of this idea. Their chief limitation is that, although the advertisements we created were thin on substantive information, it was all but impossible to create realistic advertisements that provided no information whatsoever. So, what looks like a mere exposure effect could instead be attributable (for instance) to participants learning that Mike Harper is a teacher. Thus, Studies 3a and 3b introduce a pure manipulation of candidate familiarity, with the objective of comparing how familiarity affects implicit and explicit attitudes.

Procedure. Study 3a assesses how attitudes toward Mike Harper differ, depending on whether participants have been induced to become familiar with him. Respondents were randomly assigned to one of several induction conditions, described next. Then, the survey measured their implicit and explicit attitudes toward Mike Harper. Because we measure attitudes about Mike Harper irrespective of the random assignment, we refer to Harper as the Target candidate.

Study 3a had three conditions: a Target induction, a Non-target induction, and a Control condition. In the Control condition, the protocol measured attitudes toward Mike Harper immediately after the study's consent screen. In the Target induction condition, participants underwent a procedure to induce familiarity with Mike Harper before attitude measurement. In the Non-target induction condition, participants were induced to become familiar with "Paul Coleman," a fictional candidate similar to Harper in terms of race, gender, and age. The purpose of the Non-target induction is to provide a more conservative comparison point than the pure control. For instance, this condition allows us to assess whether it is familiarity with *Harper in particular* (as opposed to, e.g., "politicians" or "white men") that affects attitudes toward him.

Study 3b was identical to Study 3a, but added a fourth condition: we induced some participants to become familiar with a woman non-target candidate ("Paula Coleman"). The purpose of this condition was to better understand the extent to which familiarity effects "spill over" onto similar attitude objects. For instance, familiarity with Paul Coleman might improve attitudes toward Mike Harper, since both are white men of a similar age. Adding a comparison point that is more dissimilar to Mike Harper than Paul Coleman helps to assess whether this is indeed a spillover effect, as opposed to chance variation.[16] We chose to create a more dissimilar candidate by modifying candidate gender because, as a matter of social cognition, another person's gender is encoded automatically (Stangor et al. 1992). If implicit attitudes toward Mike Harper benefit from induced familiarity with Paul Coleman because the two individuals are similar, this pattern should be attenuated by changing Coleman to be a woman.

[16] As can be seen in the bottom-left panel of Figure 5, the initial run of this study generated suggestive evidence of such a spillover effect, which is part of what motivated us to develop an initial treatment arm. We thank our peer reviewers for encouraging us to explore spillover effects more fully.

The familiarity induction sought to increase familiarity with the candidate without providing any substantive information about him or her. We asked participants to rate still photographs of the induction candidate, to guess the candidate's age, and to report whether a variety of adjectives (intelligent, hard-working, wealthy, competent) describe the candidate well. (Responses to these questions, of course, could only be hunches based on the still photographs.) Familiarity was further induced through a "typing test" that asked participants to type the candidate's name five times. We also presented the candidate's slogan ("Start Small, Go Far" for Harper or "Let's Get Things Done" for Coleman) and asked whether the slogan is difficult to say out loud. Of course, the real purpose of this question was to induce the participants to say the slogan out loud. The images used in the induction are presented in Figure 4.

[Figure 4 here. Title: Images from the Familiarity Induction. No caption.]

After the induction, the study measured implicit and explicit attitudes in the same way as the previous two studies. Again, *all* conditions measured attitudes toward our target candidate, Mike Harper, thereby positioning us to understand how attitudes toward Harper depend on whether participants became familiar with him, someone else, or nobody at all.

Our participants in Study 3a were undergraduates, though this time we were able to conduct the experiment at two institutions in different regions. Given that the students underwent identical procedures, we pool the two sets of students together, giving us N=457.[17] In Study 3b, we used a

[17] In Study3a, our statistical model will include a dummy variable for each of the two institutions. The new institution (University B) we incorporated into Study 3a has a high percentage of international students, whom we thought might behave differently in this study, given its locus in U.S. politics. The instrument for University B included one additional question asking whether the respondent was an international student, and the analyses below follow a pre-registered plan

sample that would help us assess the extent to which our conclusions might depend on a participant population—undergraduates—that are homogeneous with respect to age and education. We recruited participants (N=1,275) via Amazon Mechanical Turk.[18] As we report in the SI, this sample is diverse with respect to age. Additionally, we used an education quota to ensure that at least half our sample would be at a low level of education (lacking a college degree). Thus, Study 3b provides additional leverage via which to consider possible treatment effect heterogeneity. We pre-registered our expectations for both studies (See SI Section 9 and 10).

Results. Figure 5 reports the expected implicit and explicit attitudes toward Mike Harper, coded as in Study 1, as a function of the random assignment. There is no evidence that the inductions affected explicit attitudes in either study. (In Study 3a, the *F* statistic for the underlying model rounds to zero [$F(3, 435) < .01$, $p \approx 1.0$]. In Study 3b, $F(3, 1163) = 0.41$, $p = .75$. No pairwise comparisons between conditions are statistically distinguishable.) This is a predicted null result.[19]

However, familiarity with Mike Harper improves implicit attitudes toward him. In Study 3a, implicit attitudes in the Target condition are statistically distinguishable from the Control condition ($t = 3.46$, $p < .01$), though the contrast against the Non-target condition is only suggestive ($t = 1.50$, $p=.14$). In Study 3b, the Target condition is statistically distinguishable from all others (vs. Control, *t*

---

to exclude international respondents. However, results including international students are nearly identical. (We report these results in SI Section 5.) We also exclude 22 respondents (5%) identified as button-mashers by the same standard used in Study 1.

[18] Recent research has identified a concerning level of fraudulent responding on MTurk (Kennedy et al. 2020). We pre-registered and implemented several extra steps to ensure that the responses we collected would be high-quality, and we are confident that these measures were effective. We discuss these steps more fully in SI Section 12. The reported N is the number of respondents left after pre-registered exclusions.

[19] Per our pre-registration for Study3b: "We do not hypothesize any effects on explicit attitudes about Mike Harper."

= 3.07; vs. Non-target male, $t = 2.74$; vs. Non-target female, $t = 2.77$; all $p < .01$). Thus, induced familiarity with Harper improves implicit attitudes toward him, but induced familiarity with other politicians (including a demographically similar one) does not.

[Figure 5 here. Title: Treatment Effects on Explicit and Implicit Attitudes. Caption: Whiskers represent 95% confidence intervals. In Study 3a, the analysis sample is 439 for explicit attitudes and 417 for implicit attitudes. In study 3b, the sample size is 1,167 for explicit attitudes and 1,114 for implicit attitudes.]

Discussion. Studies 3a and 3b make three substantive contributions. First, they provide additional evidence that implicit and explicit attitudes are dissociated. They change via different processes, and one can change without the other: the familiarity induction moves implicit attitudes toward Mike Harper, but effects on explicit attitudes are convincingly null. Second, these studies show that merely increasing familiarity with a political candidate improves implicit attitudes about him—even in the absence of any substantive information. Third, they suggest that the channel via which familiarity improves implicit attitudes is reasonably encapsulated. Aside from a suggestive result in Study 3a that did not reemerge in 3b, implicit attitudes toward Harper improve in response to familiarity with Harper *only*.[20]

In the Supporting Information (Section 5), we report analyses stratified by age and education. We find that some evidence that age and education have *direct* effects on attitudes (with young and uneducated people having generally less favorable attitudes toward Harper), but no evidence that these factors moderate treatment effects. Such results constitute evidence that the

[20] We acknowledge that much more work must be done to understand just how encapsulated this channel is, but must leave that to future research.

effects we identify are reasonably homogeneous, and likely apply well across demographic groups (see Coppock et al. 2018).

While we think that this demonstration of mere exposure effects is instructive for the study of political messaging—we return to this below—we recognize that Studies 3a and 3b necessitated some tradeoffs. To isolate the effects of familiarity, we manipulated the concept in an unnaturalistic way. Furthermore, because we excised substantive information from the study entirely, we lost the ability to examine how familiarity effects might depend on the provision of substantive information. Thus, Study 4 seeks to reintroduce a substantive messaging. This inclusion allows us to test some limits of mere exposure effects, as well as to offer a fuller assessment of the APE Model in the context of political communication.

Study 4: How Do Substantive and Non-Substantive Messaging Elements Interact?

Study 4 returns to, and builds on, the general approach used in Study 1. Participants are exposed to the same three advertisement conditions (high-quality ad, low-quality ad, or no ad). However, this assignment is fully crossed with a separate assignment that manipulates the substantive information participants have about Mike Harper: they are told he was implicated in a political scandal, or they are given neutral information. This procedure is similar to Study 2 (in the SI), though we made some operational improvements and filed a separate pre-registration to reflect our developing theoretical understanding.

Procedure. After viewing one of the ads used in Study 1, we asked all participants to report, in open-ended format, what they typically eat for breakfast—a small distractor task to serve as a buffer between survey sections. Next, respondents were randomly assigned to read one of the two news stories presented in Figure 6. As the figure shows, one clipping was a non-political control story

about cheese making. The other story, the Scandal Story, reveals that Mike Harper was implicated in serious misconduct.

Three aspects of the Scandal Story bear some emphasis. First, the story is plain text—a medium far less rich in affective content than our audio/visual advertisements. Second, the story is parsimonious—just 153 words. Third, the allegations levied against Harper are conclusive. We wanted to leave little doubt that positive characteristics one might have ascribed to Harper on the basis of his ad are, in fact false. Fourth, we chose an allegation (embezzling funds) that, although conclusively bad, we expected to be less affect-laden than plausible alternatives. (For instance, we decided against revealing evidence of sexual assault.) Thus, while the video advertisements are tailored to target *associational* mental processes, the news story is suited to target *propositional* mental processes. From the standpoint of the APE Model, the advertisement manipulation should lead to changes in implicit attitudes, while the scandal story should lead to changes in explicit attitudes.[21]

However, given the results we describe above, our pre-registered predictions explicitly departed from the APE. Most important, for participants assigned to the control story, we predicted a mere exposure effect: ad exposure should improve implicit attitudes toward Mike Harper. At the

[21] To see how this prediction relates to dual process models of attitude change, consider that Gregg et al. write that "Entertaining the idea, out of the blue, that a novel object is X or ~X or that an existing object known to be X is in fact ~X (or vice versa)" is a quintessential example of an occurrence that should affect propositional (or abstract) processes, and therefore induce sharp explicit attitude change (2006, 4). Our instrumentation is very much motivated by Gregg et al.'s Study 3, which proceeded as follows: the authors induced positive or negative attitudes toward fictional social groups (Niffites and Luupites) via an association-based priming task in which one group was consistently paired with positive or negative traits (e.g. benevolent or barbaric). Then, in an effort to influence propositional processes, some subjects were informed that, due to a clerical error, the positive and negative traits were associated with the opposite of the intended group. As the APE Model would predict, this treatment induced a change in explicit—but not implicit—attitudes. For a political science study that has several similar design elements, see Groenendyk (2019).

same time, because Studies 1 and 2 revealed several instances where stimuli intended to target explicit attitudes nonetheless affected implicit attitudes, we predicted that the scandal manipulation would harm implicit (not just explicit) attitudes about Harper—and indeed might neutralize mere familiarity effects.

[Figure 6 here. Title: News Stories Used in Study 4. No caption.]

Once again, experiment participants were undergraduates at two institutions. One of the two institutions expanded the research participation requirement applied to undergraduates, resulting in a considerably larger sample size. As in Study 3, we follow a pre-registered plan to exclude international students from one of the participating institutions, resulting in a useable N of 1,104.[22]

Results. A between-subjects ANOVA reveals significant effects for both manipulations.[23] Of course, our main interest is in the interdependence between the two random assignments. Thus, Figure 7 shows expected implicit and explicit attitudes for the six possible assignments.[24] Clearly, a tidy framework in which each kind of attitude (implicit and explicit) is affected via a different route is not supported. For instance, the ad quality manipulation is, according to the APE, better suited to influence implicit attitudes. But the high-quality ad significantly improved explicit attitudes—compared to control and low-quality conditions, and irrespective of the scandal condition (all $p<.01$). Moreover, the scandal manipulation is, again according to the APE, better suited to influence explicit attitudes. But the scandal story had a significant negative effect on implicit

[22] Results including the 104 excluded international students are nearly identical. Similar to previous studies, 76 respondents (7%) are excluded for button-mashing through the IAT.

[23] For explicit attitudes, $F_{ad}(2, 1{,}001) = 44.32$, $p<.001$; $F_{Scandal}(1, 1{,}002) = 63.99$ $p<.001$. For implicit attitudes, $F_{ad}(2, 1{,}025) = 6.35$, $p<.002$; $F_{Scandal}(1, 1{,}025)= 21.78$, $p<.001$.

[24] The underlying model is Attitude = $\alpha + \beta_{1\text{-}3}$(Low-quality × Scandal interactions) + $\beta_{4\text{-}5}$*(Low-quality × Scandal interactions) + β_6*Institution + ε.

attitudes in two of the three ad conditions ($p<.01$, $p=.17$, $p<.01$). Thus, both manipulations affected attitudes in both domains.

On the other hand, the results are generally consistent with a mere exposure effect for implicit attitudes. In the control condition, exposure to any advertisement (low- and high-quality assignments pooled) improved implicit attitudes ($p<.02$), though as the figure shows this effect was driven more by the high-quality assignment ($p<.01$ compared to control) than the low-quality assignment ($p=.15$). In the scandal condition, exposure to any advertisement improved implicit attitudes ($p<.02$), though here the effect was driven more by the low-quality ad ($p<.02$ compared to control) than the high-quality ad ($p=.10$). No high-quality vs. low-quality contrasts were significant for implicit attitudes.

[Figure 7 here. Title: Treatment Effects in Study 4. Caption: Whiskers represent 95% confidence intervals.]

Discussion. Study 4 does not support the APE Model. However, it does accumulate further evidence that mere exposure to an attitude object improves implicit attitudes. Interestingly, mere exposure effects emerge irrespective of the scandal condition. (As our pre-registration reflects, we thought the scandal condition might neutralize mere exposure effects.) More generally, Study 4 provides additional examples of implicit/explicit attitude dissociation. For instance, among respondents assigned to the scandal condition, a shift from a low-quality to a high-quality ad significantly improves explicit attitudes ($p<.01$), but has a (non-significant) negative effect on implicit attitudes ($p=.36$).

Although our primary focus lies in comparing implicit and explicit attitudes, the explicit attitude results in Study 4 bear some discussion in their own right. The Scandal story had a large negative effect on respondents who saw no ad (difference = -0.22, SE = 0.018, $p<.0001$). But the Scandal/Control contrast is comparatively muted among respondents who saw either the low- or high-quality ads.[25] In this sense, both ads served to inoculate Mike Harper from the negative effects of a scandal. Although we are cautious about *post-hoc* speculation for why this result occurred, positive ads (such as those in this case) do help candidates build trust with voters (Ridout and Holland 2010), which may have led to the muted effect. Moreover, the combination of the positive

[25] In the Low-quality condition, difference = -0.035, SE = 0.027, $p=.19$. For the High-quality condition, difference = -0.071, SE = 0.026, $p<.001$. The effect sizes in both advertisement conditions are distinguishable from the effect size in the No ad condition. For the relevant interaction terms, $\beta_{\text{Low-quality} \times \text{Scandal}} = 0.190$, SE = 0.033, $p<.001$; $\beta_{\text{High-quality} \times \text{Scandal}} = 0.154$, SE = 0.033, $p<.01$. However, effect sizes in the two ad conditions are not distinguishable from each other. For the relevant interaction, $\beta = -0.036$, SE = 0.037, $p=.34$.

ad with the negative story of the scandal may have led the participants to weigh both more carefully, affecting explicit evaluations (Chong and Druckman 2007b).

Discussion and Conclusion

Zajonc's landmark essay laying out the case for distinct mental systems governing cognition and affect contains a passage worth quoting at length:

> The dismal failure in achieving substantial attitude change through various forms of communication or persuasion is another indication that affect is fairly independent and often impervious to cognition… If a person believes that Candidate A is honest, we can simply give the person information proving that A is not honest… Yet this approach has been the least successful in attitude change. Even the most convincing arguments on the merits of spinach won't reduce a child's aversion to this vegetable. (Zajonc 1980, 158-59)

Quite. Zajonc's words represent an insight the broad strokes of which political psychologists have *partly* embraced, such as in studies that emphasize the paramount influence of social identities (e.g. Huddy et al. 2015) or citizens' motivations to bring political judgments in line with gut reactions (e.g. Lodge & Taber 2012). But as concerns attitude *change*, the field still has some ways to go. As we note at the outset, there are precious few studies of factors that cause implicit attitudes—arguably the most direct measure of raw affect—to change. And the preeminent concept in the study of persuasion—framing effects (e.g. Druckman 2010)—is, at least to the extent it refers to an actor emphasizing different pieces of *information* relevant to some judgment, entirely cognitive in nature.

Here, by directly comparing paths for implicit versus explicit attitude change, we have attempted to open a new front in the study of political persuasion. Our studies lead us to three main conclusions.

The first conclusion is simple, but bears emphasis. We find that implicit and explicit attitudes are distinct. Over and over again, they respond differently to the treatments we administered, in some cases even seeming to move in opposite directions. This result implies that political actors might—knowingly or not—employ different tools to engage in different kinds of persuasion. Alongside the finding that implicit and explicit attitudes have distinct consequences for behavior (e.g. Albertson 2011; Quillian 2006; Ryan 2017), this conclusion underlines that the tactics used to influence voters are even more complex and multifaceted than has heretofore been appreciated.

Second, we find little evidence that the APE Model works as a framework to understand implicit and explicit attitude change in the context of political communication. We word this conclusion narrowly, and we do so because we wish not to be construed as launching a frontal assault on the APE. We regard the evidence that psychology journals have accumulated in the model's favor as sound, which is why we made it the starting point for our inquiries. We conjecture that the main reason APE falls short in our research pertains to the different priorities that emerge in traditional cognitive psychology (APE's provenance) compared to psychologically-informed political science (our focus). Where studies in pure Psychology often sacrifice realism, the better to isolate and validly manipulate the constructs of interest, political scientists strive to understand how psychological principles apply in real communicative contexts.

As an example of how these different priorities can lead to different conclusions, compare our work to one of the studies that informed our understanding of the APE Model. Moran & Bar-Anan (2013) created negative associations with an alien creature by having participants listen to horrified screaming while viewing pictures of the creature on a computer screen. In contrast, our studies attempted to create associations with a candidate via plausible advertisements that varied in their production quality. Our manipulation, more than the screams, is susceptible to spill over to areas not intended. For instance, although our ads intended to provide equivalent information (by

having identical narration), perhaps respondents made different substantive inferences about Mike Harper's competence depending on whether or not they saw a low- or high-quality ad. We acknowledge this possibility. However, we see it not as a mistake, but rather as a data point that illustrates a limitation of the APE: the model does not transition away from a lab and into the field so seamlessly.[26]

Although we do not corroborate the APE, our studies accumulate evidence for an alternative. Our third conclusion is that political messages can improve implicit attitudes simply by increasing familiarity with a target object. This pattern arises in all four studies we conducted, including two (Study 3a and 3b) with a pure familiarity manipulation. To be sure, there is more to learn about this pattern: does it arise when familiarity is induced in a negative way (such as an attack ad)? How does the magnitude and perseverance of the effect relate to the treatment dosage? And what interventions beyond familiarity successfully move implicit attitudes? (Because we have found one route does not preclude the existence of others.)

Although we must leave these questions to future studies, the familiarity patterns we discuss above already help elucidate real political phenomena. We close with two examples.

Low-information campaign tools. Scholars have long noticed that politicians seek to influence voters with messages that convey little or no substantive information. Long-standing examples

[26] Gawronski & Bodenhausen have a discussion ("Is the APE Model Falsifiable?": 2018, 25) that is helpful for understanding how the results we report relate to the APE Model. There, the authors acknowledge that the model is difficult to falsify. Per the authors, because the model allows for interactions between associative and propositional processes it is possible to explain almost any empirical outcome in a post-hoc fashion. Thus, we see our studies as falsifying *one formulation* of the APE—the version that might have been especially revealing as concerns political communication. Other formulations can remain, though they would be far more complex and difficult to apply to politics.

include lawn signs, billboards, placards, and vacuous television spots. Many ads on modern advertising platforms—Facebook, YouTube, and Instagram, for instance—undoubtedly fit the bill, too (e.g. Fowler et al 2021). The most common explanations for such efforts are almost entirely *cognitive* in nature:[27] these ads can build name recognition, which is electorally beneficial (e.g. Abramowitz 1975; Levy & Squire 2000; Makse, Minkoff, and Sohkey 2019). Or, they might influence thoughts about a candidate's viability, incumbency, or experience (e.g. Green et al. 2015; Krasno 1994). Our results suggest that something as simple as a lawn sign may have effects that are more powerful than a purely cognitive model would predict: familiarity with a candidate can affect implicit (but not necessarily explicit) attitudes without necessarily changing cognitions about him or her. Note that a result we saw in Study 1—the greater decay of treatment effects on implicit attitudes after a delay of two weeks—fits in well with this literature. Because implicit attitude change stems from changing positive and negative associations with an attitude object, it unfolds in a more cumulative and incremental way than explicit attitude change (Gregg et al. 2006). Consistent with this, research in communication finds that the success of low-information political messages hinges in no small part on massive repetition (Fernandes 2013).

The content of positive vs. negative messages. There is considerable evidence that negative political ads are more informative than positive ads. In particular, negative ads tend to contain more information about political issues (Fowler & Ridout 2010; Koch 2008). Why would this be so? The results we present point to one possibility. If negative ads focused primarily on positive/negative associations, there are two reasons they might backfire: they might accidentally increase familiarity with the opponent politician, or the negative associations might, against intentions, become

[27] An important exception is Kam & Zechmeister (2013), who much like us, explicitly contemplate a "direct" effect of familiarity on attitudes. Although these authors examine the effects of implicit (i.e. subliminal) priming, they do not measure implicit attitudes as a dependent variable.

associated with the sponsoring politician (especially given the legal requirement that the sponsor verbally approve each ad). For these reasons, negative ads might need to primarily target *propositional* thinking. Positive ads, in contrast focus solely on the sponsoring candidate (often without any mention of the opponent), so it is more propitious for them to target associations and, as a result, implicit attitudes.

To close, we note that this series of investigations illustrates one underappreciated benefit of the disciplinary trend toward experimental pre-registration (Franco et al. 2014). One common concern about expecting researchers to pre-register theoretical expectations is that the requirement might hinder exploratory research, or suppress publication of studies with unexpected results (Laitin 2013). Privately, we have even heard some concern that if researchers' expectations are on-record, they might have even bigger incentives to torture results into alignment.

We think these concerns reveal a limited view of pre-registration. As Platt (1964) argues, scientific inference is like climbing a many-branched tree. At each fork, empirical tests exclude some paths forward while leaving others open. In this framework, pre-registrations serve a role similar to a chemist's lab book. They do not serve to handcuff researchers to any particular theoretical view. They simply create a clear record of what ideas a particular test was designed to evaluate. Rather than leading to incentives to hide unexpected results, this record makes it more appealing to lay them bare, as doing so makes it easier for readers to share in the same learning process. It is an approach we found refreshing and hope will continue.

The authors declare the human subjects research in this article was deemed exempt from review by the University of North Carolina at Chapel Hill and Stony Brook University. The authors declare no ethical issues or conflicts of interest in this research. This research was funded by the Junior Faculty Development Award at the University of North Carolina Chapel Hill as well as by Stony Brook University's Political Science Department. Research documentation and/or data that support the findings of this study are openly available in the APSR Dataverse at https://doi.org/10.7910/DVN/IK4NUZ.

References

Abramowitz, Alan I. 1975. "Name Familiarity, Reputation, and the Incumbency Effect in a Congressional Election." *Western Political Quarterly* 28(4): 668–84.

Albertson, Bethany L. 2011. "Religious Appeals and Implicit Attitudes." *Political Psychology* 32(1): 109–30.

Albertson, Bethany, and Shana Kushner Gadarian. 2015. *Anxious Politics: Democratic Citizenship in a Threatening World*. New York: Cambridge University Press.

Banaji, Mahzarin R, and Larisa Heiphetz. 2010. "Attitudes." In *Handbook of Social Psychology*, eds. Susan T Fiske, Daniel T Gilbert, and Gardner Lindzey. New York: Wiley, 348–88.

Bargh, John A, and Tanya L Chartrand. 1999. "The Unbearable Automaticity of Being." *American Psychologist* 54(7): 462–79.

Borah, Porismita, Erika Fowler and Travis Ridout. 2018. "Television vs. YouTube: Political Advertising in the 2012 Presidential Election" *Journal of Information Technology & Politics*. 15(3): 230-244.

Brader, Ted. 2006. *Campaigning for Hearts and Minds*. Chicago: University of Chicago Press.

Brader, Ted, and George E. Marcus. 2013. "Emotion and Political Psychology." In *Oxford Handbook of Political Psychology*, eds. Leonie Huddy, David O Sears, and Jack S Levy. Oxford University Press.

Chong, Dennis, and James N Druckman. 2007a. "Framing Theory." *Annual Review of Political Science* 10(1): 103–26.

Chong, Dennis and James N. Druckman. 2007b. "Framing Public Opinion in Competitive Democracies" *American Political Science Review* 101 (4): 637-655.

Coppock, Alexander. 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7(3): 613–28.

Coppock, Alexander, Thomas J Leeper, and Kevin J Mullinix. 2018. "Generalizability of Heterogeneous Treatment Effect Estimates across Samples." *Proceedings of the National Academies of Science* 115(49): 12441–46.

Dovidio, John F, Kerry Kawakami, and Samuel L Gaertner. 2002. "Implicit and Explicit Prejudice and Interracial Interaction." *Journal of Personality and Social Psychology* 82(1): 62–68.

Druckman, James N. 2010. "What's It All About." In *Perspectives on Framing*, ed. Gideon Keren. New York: Psychology Press.

Druckman, James N, and Cindy D Kam. 2011. "Students as Experimental Participants: a Defense of the 'Narrow Data Base'." In eds. James N Druckman, Donald P. Green, James H Kuklinski, and Arthur Lupia. Cambridge Handbook of Experimental Political Science.

Fang, Xiang, Surendra Singh, and Rohini Ahluwalia. 2007. "An Examination of Different Explanations for the Mere Exposure Effect." *Journal of Consumer Research* 34(1): 97–103.

Fazio, Russell H, and Michael A Olson. 2003. "Implicit Measures in Social Cognition Research: Their Meaning and Use." *Annual Review of Psychology* 54(1): 297–327.

Fowler, Erika Franklin and Ridout, Travis N. 2010. "Advertising Trends in 2010," *The Forum*. 8(4), Article 4.

Fowler, Erika Franklin, Michael Franz, Gregory J. Martin, Zachary Peskowitz and Travis N. Ridout (2021) "Political Advertising Online and Offline" *American Political Science Review*. 115(1): 130-149.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2017. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345(6203): 1502–5.

Gawronski, Bertram, and Galen V Bodenhausen. 2006. "Associative and Propositional Processes in Evaluation: an Integrative Review of Implicit and Explicit Attitude Change." *Psychological Bulletin* 132(5): 692–731.

Gawronski, Bertram, and Galen V Bodenhausen. 2007. "Unraveling the Processes Underlying Evaluation: Attitudes From the Perspective of the APE Model." *Social Cognition* 25(5): 687–717.

Gawronski, Bertram, and Galen V Bodenhausen. 2014. "Implicit and Explicit Evaluation: a Brief Review of the Associative-Propositional Evaluation Model." *Social and Personality Psychology Compass* 8(8): 448–62.

Gawronski, Bertram, and Galen V. Bodenhausen. 2018. "Evaluative Conditioning From the Perspective of the Associative-Propositional Evaluation Model." *Social Psychological Bulletin* 13(3): e28024.

Geer, John G. 2006. *In Defense of Negativity: Attack Ads in Presidential Campaigns*. Chicago: University of Chicago Press.

Green, Donald P. et al. 2015. "The Effects of Lawn Signs on Vote Outcomes: Results From Four Randomized Field Experiments." *Electoral Studies*: 1–22.

Gregg, Aiden P, Beate Seibt, and Mahzarin R Banaji. 2006. "Easier Done Than Undone: the Asymmetric Malleability of Automatic Preferences." *Journal of Personality and Social Psychology* 90(1): 1–20.

Greenwald, Anthony G, and Brian A. Nosek. 2008. "Attitudinal Dissociation: What Does It Mean?." In *Attitudes: Insights From the New Implicit Measures*, eds. Richard E. Petty, Russell H Fazio, and Pablo Briñol. Hillsdale, NJ: Lawrence Erlbaum Associates, 65–82.

Greenwald, Anthony G, Debbie E McGhee, and Jordan L K Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: the Implicit Association Test.." *Journal of Personality and Social Psychology* 74(6): 1464–80.

Greenwald, Anthony G, T Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity." *Journal of Personality and Social Psychology* 97(1): 17–41.

Groenendyk, Eric. 2019. "Of Two Minds, but One Heart: a Good 'Gut' Feeling Moderates the Effect of Ambivalence on Attitude Formation and Turnout." *American Journal of Political Science* 63(2): 368–84.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley & Sons.

Kaid, Lynda Lee, and Anne Johnston. 2001. *Videostyle in Presidential Campaigns: Style and Content of Televised Political Advertising*. Westport, CT: Praeger.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Macmillan.

Kam, Cindy D, and Elizabeth J Zechmeister. 2013. "Name Recognition and Candidate Support." *American Journal of Political Science* 57(4): 971–86.

Karpinski, Andrew, and James L Hilton. 2001. "Attitudes and the Implicit Association Test." *Journal of Personality and Social Psychology* 81(5): 774–88.

Kennedy, Ryan et al. 2020. "The Shape of and Solutions to the MTurk Quality Crisis." *Political Science Research and Methods* 8(4): 614–29.

Koch, Jeffrey. 2008. "Campaign Advertisements' Impact on Voter Certainty and Knowledge of House Candidates' Ideological Positions" *Political Research Quarterly*. 61(4): 609-621.

Krasno, Jonathan S. 1997. *Challengers, Competition, and Reelection: Comparing Senate and House Elections*. New Haven: Yale University Press.

Laitin, David D. 2013. "Fisheries Management." *Political Analysis* 21(1): 42–47.

Levy, Dena, and Peverill Squire. 2000. "Television Markets and the Competitiveness of US House Elections." *Legislative Studies Quarterly*: 313–25.

Lundberg, Kristjen B, and B Keith Payne. 2014. "Decisions Among the Undecided: Implicit Attitudes Predict Future Voting Behavior of Undecided Voters." *PLoS ONE* 9(1): e85680.

Makse, Todd, Scott L. Minkoff and Anand E. Sokhey. 2019. *Politics on Display: Yard Signs and the Politicization of Social Spaces*. New York: Oxford University Press.

Moran, Tal, and Yoav Bar-Anan. 2013. "The Effect of Object–Valence Relations on Automatic Evaluation." *Cognition and Emotion* 27(4): 743–52.

Pérez, Efrén O. 2016. *Unspoken Politics: Implicit Attitudes and Political Thinking*. Cambridge University Press.

Petty, Richard E., and John T. Cacioppo. 1986. "The Elaboration Likelihood Model of Persuasion." In *Advances in Experimental Social Psychology*, Advances in Experimental Social Psychology, Elsevier, 123–205.

Petty, Richard E., Russell H Fazio, and Pablo Briñol, eds. 2008. *Attitudes: Insights From the New Implicit Measures*. Washington DC: Psychology Press.

Platt, John R. 1964. "Strong Inference." *Science* 146(3642): 347–53.

Quillian, Lincoln. 2006. "New Approaches to Understanding Racial Prejudice and Discrimination." *Annual Review of Sociology* 32(1): 299–328.

Ridout, Travis N, Erika Franklin Fowler, John Branstetter and Porismita Borah. 2015. "Politics as Usual? When and Why Traditional Actors Often Dominate YouTube Campaigning" *Journal of Information Technology & Politics*. 12(3): 237-251.

Ridout, Travis N and Jenny L Holland (2010) "Candidate Strategies in the Presidential Nomination Campaign" Presidential Studies Quarterly 40(4): 611-630.

Ryan, Timothy J. 2017. "How Do Indifferent Voters Decide? the Political Importance of Implicit Attitudes." *American Journal of Political Science* 61(4): 892–907.

Schatz, Carla. 1992. "The Developing Brain." *Scientific American* 267(3): 60–67.

Sherman, Jeffrey W, Bertram Gawronski, and Yaacov Trope, eds. 2014. *Dual-Process Theories of the Social Mind*. New York: Guilford Press.

Vavreck, Lynn. 2016. "The Ad That Moved People the Most: Bernie Sanders's 'America.' *The New York Times*. Available online at www.nytimes.com

Zajonc, Robert B. 1968. "Attitudinal Effects of Mere Exposure." *Journal of Personality and Social Psychology Monograph Supplement* 9(2): 1–27.

Zajonc, Robert B. 1980. "Feeling and Thinking: Preferences Need No Inferences." *American Psychologist* 35(2): 151–75.