Is it the Message or the Messenger? Examining Movement in Immigration Beliefs*

Hassan Afrouzi
Columbia University

Carolina Arteaga
University of Toronto

Emily Weisburst
UCLA Luskin

November 2022

[Link to Supplemental Materials]

Abstract

How do political leaders affect constituents' beliefs? Is it rhetoric, leader identity, or the interaction of the two that matters? Using a large-scale experiment we estimate treatments that decompose the relative importance of partisan messages vs leader sources, in the context of beliefs about immigration. Participants listen to anti-immigrant and pro-immigrant speeches from both Presidents Obama and Trump. These treatments are benchmarked to versions of the speeches recorded by an actor to control for message content, and to non-ideological presidential speeches to control for leader priming. Our findings show that political leader sources influence beliefs beyond the content of their messages in a special case: when leaders deliver unanticipated messages to individuals in their own party. This evidence supports the hypothesis that individuals will "follow their leader" to new policy positions.

JEL Classification: D83, C90

Keywords: Leaders, Political Beliefs, Partisan Identity, Polarization, Immigration

*We thank Jonathan Weisburst for sound editing and Michael Floyd for voice recording. We are also grateful to Marco Tabellini, Pablo Querubin, Tianyi Wang, Eric Merkley, Arthur Blouin, Kristen Cornelson, Sandra Black, Sarah Reber, and Felipe Goncalves for helpful feedback as well as seminar participants at UC Merced, UCLA, the University of Toronto, and the Canadian Economic Association (CEA). We gratefully acknowledge support from ISERP seed grant from Columbia University, SSHRC Institutional Grant and REG from UCLA. The experiment in this study was pre-registered with the American Economic Association with number AEARCTR-0006552. The study procedures were approved by the Institutional Review Boards of Columbia University, University of Toronto, and UCLA.

Afrouzi: Columbia Economics, 1022 International Affairs Building, 420 W 118th St, New York, NY, 10027, ha2475@columbia.edu. Arteaga: University of Toronto Economics, carolina.arteaga@utoronto.ca. Weisburst: UCLA Luskin School of Public Affairs, weisburst@ucla.edu.

Statements from political leaders directly influence their followers' beliefs and actions.[1] But, how much of this influence is due to the leaders' brand identity as opposed to the content of their political rhetoric, or to the interaction of the two? The answer to this question is important not only because it enhances our understanding of how political beliefs are formed, but also because it informs policies that aim to either (1) communicate new information effectively,[2] or (2) correct misconceptions in beliefs.[3]

Nonetheless, it is difficult to separately identify the role of political messages from the identity of their sources in the determination of beliefs because what leaders say is endogenous to their position. Leaders might choose to cater to the underlying views of their party constituents in order to increase their chances of getting elected, simply echoing and amplifying the views of their electors. Conversely, party constituents might choose to follow the political preferences of their elected leaders.[4] While these two cases are observationally equivalent, their implications for the roles of leader identity and the content of their political rhetoric are different. Consequently, the ideal experiment that would identify these two separate roles is one that independently varies the identity of the source and the content of a political message.

In this paper, we leverage a novel large-scale experiment with over 12,800 participants to isolate the impacts of political message content and political leader sources on beliefs about immigration policy. Immigration is an ideal context to study the determinants of leader influence because it is both a meaningful policy topic for voters and voters' views on immigration vary substantially across party lines.[5] In our experiment, we utilize audio recording treatments that are excerpts of *actual* speeches given by Presidents Barack Obama and Donald Trump and compare these treatments to exact anonymous replicas that were recorded by a voice actor. Our constructed speech treatments include an *anti-immigrant* and a *pro-immigrant* speech from each president. Further, we include additional treatments of audio segments of non-ideological speeches for each president (ceremonial "turkey pardon" speeches on Thanksgiving), for a total of ten speeches and eleven treatment arms.[6] We embed these treatments in an online survey using

[1] A large literature finds that leaders impact followers, see for example, Beaman et al. (2009), Beaman et al. (2012), Bidwell et al. (2020) and Fujiwara and Wantchekon (2013).

[2] See, for example, work on effective communication about the benefits of a new vaccine (Larsen et al., 2022).

[3] See, for example, research studying interventions relevant to the diffusion of fake news in online forums (Cinelli et al., 2021), beliefs about characteristics of migrants (Alesina et al., 2018), or support for female labor force participation (Bursztyn et al., 2018).

[4] Research in economics predicts that news media will cater to segmented consumer political preferences, potentially exacerbating polarization (Gentzkow and Shapiro, 2006; Baron, 2006; Mullainathan and Shleifer, 2005). Further, individuals who choose to voice opinions in opposition of their political party or group can face costly social sanctions (Bursztyn et al., 2022).

[5] Democrats are generally pro-immigrant while the majority of Republicans hold anti-immigrant views. 52% (70%) of voters in the 2020 (2016) presidential election characterized immigration as being "very important" to their vote (Doherty et al., 2020, 2016). Recent survey evidence shows that 50% of Democrats would like to see immigration levels increase, relative to only 13% of Republicans (Younis, 2020)

[6] Research in economics and psychology has found that cues about social identity can alter perception, beliefs,

a sample of Republican and Democrat participants, and stratify the treatment randomization within parties.

To inform our experimental design, we use a conceptual framework that models a partisan agent that starts with a prior belief about immigration and may update this belief when exposed to political statements. From this framework, we derive two alternative decompositions of the role of leader sources and partisan messages. In the first decomposition, we fix the president source and vary the message of a speech when the president's identity is known. This decomposition estimates a source priming effect, related to leader brand identity, and a source-specific message effect that is net of priming. For this exercise, we leverage variation in the exposure to a president delivering an immigration speech and a non-ideological "turkey pardon" speech. In the second decomposition, we fix the immigration speech message and vary the source of the statement. This decomposition estimates an anonymous message effect, which captures the message content effect, and the leader source's persuasion power for each message. Here, we leverage variation in exposure to the voice actor version of an immigration speech relative to the president version of the same speech. Given the stark differences in baseline views across political parties, we hypothesize that any results we find related to leader persuasion will be mirrored by party.

The conceptual framework also provides several useful predictions. First, the framework predicts that partisan messages will induce agents to update beliefs when they believe that the partisan message is associated with the true the state of the world regarding immigration, as opposed to an uninformative political signal. Likewise, the framework predicts that agents will be primed by the brand reputation of leaders to change their beliefs only if they subjectively link the leader's identity to the true state of the world regarding immigration. Lastly, the framework provides useful insights about when particular sources increase the persuasive power of a particular message type. Here, persuasiveness can be characterized by two multiplicative factors: 1) how unexpected a message is when it comes from a particular source and 2) the agent's subjective view of the reliability of that source. Thus, the framework predicts that leaders will have the most influence on beliefs when they express surprising or unexpected messages (e.g. a pro-immigrant message from President Trump) to an audience of supporters who are likely to find the leader to be reliable (e.g. Republican participants for President Trump messages).

Our decompositions of the determinants of leader influence illustrate that both political messages and sources matter. We first estimate the simple combined effect of political message and leader sources, by considering differences between participants who hear a speech from

and actions (e.g. Cohn and Maréchal, 2016). This last treatment arm allows us to explicitly control for the possibility that exposure to a particular leader primes participants about their political identity and subsequently alters their views on immigration.

a president's voice relative to participants who are not exposed to any audio message. We find that participants from both parties update their beliefs based on the total effect of presidential speeches, and these treatments move beliefs by $\approx 1-8\%$ in the direction of the message. Participants become more anti-immigrant (pro-immigrant) when they hear a president speech that is anti-immigrant (pro-immigrant), with effects that are larger for speeches that oppose the party prior. Next, we estimate our first decomposition by fixing the presidential source and varying the message between a non-ideological statement that contains no information about immigration (turkey pardon) and an immigration message. This decomposition does not display any independent source priming effect related to the brand identity of leaders; participants do not change their views on immigration based on hearing a non-ideological message from either president.

In our second decomposition, we separate the total effect of the president speeches between the effect of an anonymous message and the source persuasion effect. Here we fix the message of a speech and vary the messenger voice or source, either the actor or the original president. We estimate that in nearly all cases, the effect of the anonymous message drives a substantial portion of the total effect. Our findings on the importance of message content are notable for two reasons. First, the rhetoric used in the speeches for this experiment is emotional and political, as the speech statements focus on immigration views and policy proposals and do not include any factual information or substantive content about immigration. Ex ante, it is unclear that such statements would contain enough new information to move respondent beliefs, especially in a polarized political environment. The anonymous message effects show that individuals have the capacity to update beliefs when presented with new perspectives, even when those positions are not supported by rigorous evidence. These results suggest that the increasingly divergent political views we observe across parties may be partly a function of a lack of exposure to opposition views.

Turning to the role of leader persuasion, which captures the added effect of a specific source delivering a particular message, we find that this channel is only important in one symmetric case: when a party leader delivers a message against party lines to members of his own party. That is, Obama (Trump) has an additional source persuasion effect only for Democrats (Republicans) when giving a speech that is anti-immigrant (pro-immigrant). For these treatments, we find that the source persuasion effect comprises 44-58% of the total movement in beliefs. These effects serve to reduce the distance between party beliefs ("partisan polarization") in the treatment groups, relative to the control group, because the source persuasion effects are only present for speeches that oppose the party prior. The overall implication of the results is that *supporters of a leader are willing to follow their leaders to a new position.* In our context, this leads to a 30% reduction in polarization in beliefs about immigration. However, in our general framework, any surprising position taken by a trusted leader could serve to sway beliefs, which could have the effect of increasing political division in a different context. The source persuasion findings underscore the importance of particular messengers delivering partisan statements.

Our study builds on a large literature in behavioral economics and political economy about the nature of bias in information sources and the impact that new information can have on beliefs. When individuals are exposed to new information, they are aware of possible biases of the information source (Gentzkow et al., 2018), may be more skeptical of sources that do not align with their ideology (Chopra et al., 2022), can perceive bias from identical messages presented from different sources (Baum and Gussin, 2008), and can engage in motivated reasoning, which posits that people distort their inference process in the direction of states they find more attractive (Thaler, 2021). Our work contributes to this landscape by empirically isolating the message and source components of partisan information signals.

Our first contribution is to empirically test when partisan leaders can persuade individuals to change beliefs, separate from the content of a political message. Prior work by Chiang and Knight (2011) finds that newspaper endorsements of political candidates are most influential when they come from an unexpected source; for example, an endorsement of a Democratic candidate is most effective when it comes from a right-leaning newspaper. In contemporanous work by Larsen et al. (2022), exposure to a video of Donald Trump on Fox News encouraging the take up the Covid vaccine increases vaccination rates. These findings are consistent with our results and model framework. We exploit the richness of our experimental design, to extend their results and find that source and voter alignment or voter's trust, play a key role in explaining the influence of unexpected or counter-stereotypical information.

Our work on leader persuasion also relates to a large literature in political science. In this field, a number of papers have experimentally varied the source of a policy statement (*party cue*) within a survey and then tested how beliefs may change given a particular source.[7] This work generally finds that party cues are important, but finds mixed impacts. These papers typically test hypotheses using a brief policy statement or a single partisan figure and then use one summary question to measure belief outcomes. Our comprehensive experiment allows us to vary message ideology for two very different leaders and to compare these treatments to identical messages with an ambiguous source, as well as to non-ideological messages from these leaders. Further, we reduce noise in our outcome measure of immigration beliefs by combining information from a large number of post-treatment questions that ask about different aspects of immigration policy. This robust range of treatments and outcome questions allows us to credibly

[7] e.g. Merkley and Stecula (2021); Bakker et al. (2020); Bullock (2020); Barber and Pope (2019); Broockman and Butler (2017); Boudreau and MacKenzie (2014); Druckman et al. (2013); Nicholson (2012); Bullock (2011); Goren et al. (2009); Gilens and Murakawa (2002).

isolate the effect of both partisan messages and partisan leader sources on belief formation.

The second contribution of our paper is to test whether partisan rhetoric from leaders, rather than factual information, can sway beliefs. This question is unsettled in the economics literature on information signaling. Several studies have shown theoretically and empirically that partisan individuals often assign their own biases to messages, and can interpret identical neutral signals in differing ways in accordance with their existing priors.[8] Given that individuals can place so much of their own bias on information signals, it is ex ante unclear whether political messaging could serve to move beliefs. Moreover, a second strand of research finds mixed results of the ability of curated messages to move beliefs in an intended direction.[9] Work in this area has found that individuals have large misperceptions about facts relevant to policy issues, and that they may update their beliefs when provided with new factual information.[10] Some of this work finds meaningful impacts for *fact-based* information treatments on the policy topic of immigration (e.g. Alesina et al., 2018; Haaland and Roth, 2020), but this research may not translate to our research question testing the messaging effects of *emotion-based* partisan speeches from leaders. Our experiment uses actual audio from presidential speeches in lieu of listed statistics or text narratives, and attempts to approximate the biased, incomplete, and sometimes inaccurate political messages common in the real world.

Finally, we contribute to the literature on priming effects in behavioral economics. Studies in this literature have found that exposure to cues which remind individuals of their ethnic, religious, and/or cultural identity can meaningfully alter beliefs, economic decisions, and risk preferences.[11] In designing our experiment, we recognized that the mere signal of seeing a political leader could remind an individual of their own political affiliation and induce updating toward the known and expected beliefs of their political party. We test this priming hypothesis directly in our study by including non-ideological messages that do not contain any information on immigration from each president. Our null finding of this form of political priming is interesting in the context of a broad literature that highlights the importance of identity cues in signaling group norms (Cohn and Maréchal, 2016).

1 Experiment Design

Our experiment is embedded within an online survey, where participants are exposed to different audio segments of presidential speeches. Participants are asked a series of background questions

[8] e.g. Baysan (2021); Fryer Jr et al. (2019); Benoît and Dubra (2019); Andreoni and Mylovanov (2012).

[9] e.g. Levy (2021); Song (2021); Durante et al. (2019); Kalla and Broockman (2018); Martin and Yurukoglu (2017); Adena et al. (2015); Enikolopov et al. (2011); DellaVigna and Kaplan (2007); Gerber et al. (2009).

[10] e.g. Haaland and Roth (2020); Grigorieff et al. (2020); Alesina et al. (2018); Bursztyn et al. (2018); Cruces et al. (2013).

[11] e.g. Akerlof and Kranton (2000); Callen et al. (2014); Cohn et al. (2015); Benjamin et al. (2016); Cohn et al. (2015).

on demographics, political views, and news consumption prior to treatment. After treatment, participants are asked questions on their views on immigration. The experiment was pre-registered with the American Economic Association[12] and was conducted during the period of the 2020 presidential election between October 16 - November 10, 2020. We explicitly construct the experiment to be stratified by political party and recruited participants who identify as Republicans or Democrats, as we expect that respondents will interpret treatment through the lens of political identity and any results that we find should be mirrored across political parties.

The experiment contains 11 treatment arms for each political party, designed to test how partisan statements from political leaders move beliefs. These include 4 president immigration speech segments, consisting of one pro-immigrant and one anti-immigrant speech for *both* Presidents Donald Trump and Barack Obama, and replicate versions of these 4 speeches recorded by a voice actor. We additionally include 2 presidential speeches with no ideological or immigration content; these speeches are ceremonious "turkey pardon" addresses for Trump and Obama.[13] The last arm of the study is a control group that is not exposed to any audio speech treatment. Figure 1 depicts the experiment design.[14]

The source material for the treatments are actual speeches delivered by each president. The audio clips used in the survey are extracted excerpts from these speeches. It is important to note that there is *no deception* used in this study, as speeches are always introduced to study participants as "an excerpt of a presidential speech from President [Donald Trump or Barack Obama]." After the edited speeches were constructed, we hired a voice actor to record replicate versions of the speech segments. The actor versions of speeches are introduced to study participants as "an excerpt of a presidential speech read by an actor."

We use an address given by Barack Obama on November 20, 2014 and an address given by Donald Trump on January 19, 2019 as source material. The purpose of the Obama speech was to introduce new protections for undocumented immigrants, including the Deferred Action for Childhood Arrivals (DACA) program, which created new safeguards for individuals who arrived in the U.S. illegally as children. Coupled with these reforms, the Obama speech also outlined additional provisions for border security. The purpose of the Trump speech was to provide a proposal to end a government shutdown related to immigration policy negotiations

[12] The pre-registration number is AEARCTR-0006552.

[13] The presidential tradition of a ceremonial turkey pardon typically consists of an event where a president gives a speech on the day before Thanksgiving, which includes the presentation of a live turkey who the president spares from being killed for a Thanksgiving dinner.

[14] We note that our experiment includes only one president source from each party by design; as a result, we are not able to separate any potential differences between Presidents Obama or Trump from other partisan figures within their parties. We use the term "leader" as a general term associated with the experimental effect of Presidents Obama and Trump, and note that the distinction of these figures versus other figures within the same ideology is an interesting avenue for future research.

with Congress; this proposal included both new border security programs as well as a concession to protect DACA recipients. Both speeches were televised from the White House, prepared in advance, and delivered using a teleprompter. We are able to extract both pro-immigrant and anti-immigrant segments from each speech because each speech contains proposals to provide protections for immigrants as well as proposals to curb illegal immigration. Similarly, we use original "turkey pardon" speeches to compose a non-ideological treatment speech that contains no information about immigration for each president.[15] The full text and web links to audio for the original speeches and treatment excerpts are included in the Supplemental Materials Appendix S2.

We take several steps to minimize experimenter demand effects following recommendations outlined in Haaland et al. (2020). First, we present a neutral framing for recruitment and in the language used in the survey. Second, respondents complete the survey on their own technological devices without the physical presence of a researcher. This approach maximizes privacy as we do not collect names or any contact information, which provides anonymity and privacy to participants. These design choices potentially increase truthful reporting Ong and Weiss (2000). Third, in order to minimize demand effects, we also stratify treatments by party lines and expose participants to a single treatment type (particular speech segment from the president or actor), which cannot reveal the overarching variation in treatments of the study or research question to a particular participant.

How similar are the anti-immigrant (or pro-immigrant) speech treatments for the two presidents? While our experiment does not directly compare treatment effects across presidents, as the extracted segments are sourced from two separate original speeches, the Trump and Obama speeches are in fact similar in their rhetoric and ideology.[16] As a direct test of the similarity of the statements across presidents, we asked participants how anti-immigrant (or pro-immigrant) they felt each speech was after treatment, on a scale of 0 to 100. Figure 3 shows the perceived degree of sentiment *strength* for the speeches, or the participant perception that an anti-immigrant (pro-immigrant) speech is anti-immigrant (pro-immigrant). We plot perceptions from the actor speech groups so that responses are not colored by the president's reputation.[17] Strikingly, the Trump and Obama speeches are perceived nearly identically within message type, for both Republicans and Democrats. Further, the perceived strength of statements across parties is also similar, with both groups viewing the pro-immigrant speeches as somewhat stronger than

[15]We use the first Thanksgiving turkey pardon for each president, delivered on November 25, 2009 by Obama and on November 21, 2017 by Trump. These speeches are edited only for length.

[16]Appendix Section A1 shows the opening paragraphs of each treatment speech segment. The text of the speeches exhibit parallel sentiments within the anti-immigrant or pro-immigrant treatments, as well as several common phrases used by both presidents.

[17]Figure A2 shows corresponding estimates for all treatment groups.

the anti-immigrant speeches. Again, while the experiment design does not rely on any direct comparisons *across* president or *across* party, these results are reassuring in that they show that the constructed speeches are similar for both Presidents and message types.

2 Conceptual Framework

Any statement from a political leader is in essence a double treatment with a *message or content* and a *source*: in principle, both the identity of the leader and the content of the message interact to shape and alter the audience's beliefs. It is our goal to disentangle these effects in this study. In this section, we provide two parallel decompositions of the source and content effects that we then use as a blueprint to guide the design of our experimental treatments.

Setting. Let Ω denote a binary sample space for an individual's beliefs about immigration: whether immigration is favorable or unfavorable. A belief system about immigration is a probability space on events that contain messages as well as sources. Let the set M contain all the messages that an individual might expect to receive from a leader, and let S denote the set of all the leaders who could be the sources of such messages. We can then define the augmented outcome space $\bar{\Omega}$ as $\Omega \times M \times S$ and model an individual's subjective beliefs as a probability space on $\bar{\Omega}$. Assuming $\bar{\Omega}$ is finite, it suffices to assign a probability to every outcome on this set. Let $P : \bar{\Omega} \to \mathbb{R}_+$ denote such a probability measure. Then, for any triple $(\omega, m, s) \in \bar{\Omega}$, $P(\omega, m, s)$ captures the subjective beliefs of the individual about the joint probability of the outcome ω (for instance, immigration being favorable) with message/content m coming from leader s.

Double Treatments with Political Sources and Messages. Let us now consider a treatment group that hears a political message m from a certain political leader s. Then, this treatment group's belief about an outcome $\omega \in \Omega$ (e.g. favorability of immigration) relative to the control group that is not treated is given by

$$\frac{1 + P(\omega|m, s)}{1 + P(\omega)} \tag{1}$$

where $P(\omega) \equiv \sum_{s' \in S, m' \in M} P(\omega, m', s')$ is the unconditional belief of the control group about ω.[18] We now present two decompositions of this treatment effect that helps us disentangle the treatment effect of the message m from the source s.

2.1 Decomposition I: Fixing the Source

Fix a source $s \in S$ and consider a second treatment group that is treated with a message m^0 from the source s that is irrelevant to immigration and thus independent of the outcome $\omega \in \Omega$. In

[18] We have added 1 to both the denominator and numerator to avoid dividing by zero when $P(\omega) = 0$, i.e. when an individual finds immigration fully favorable (or unfavorable). This allows us to keep all participants in the analysis, regardless of the value of their belief outcome after treatment. The important notion here is that both denominator and numerator and transformed symmetrically so that when $P(\omega|m, s) = P(\omega)$, the relative effect $\frac{1+P(\omega|m,s)}{1+P(\omega|m)}$ is also 1.

our experiment, m^0 is a turkey pardon speech from a certain president. Then this group's belief corresponds to $P(\omega|s,m^0)$, which by the independence assumption is equal to $P(\omega|s)$.[19] This treatment, therefore, allows us to decompose the treatment effect in Equation 1 to:

$$\ln(\frac{1+P(\omega|m,s)}{1+P(\omega)}) = \underbrace{\ln(\frac{1+P(\omega|s)}{1+P(\omega)})}_{\beta_s \equiv \text{ source priming}} + \underbrace{\ln(\frac{1+P(\omega|m,s)}{1+P(\omega|s)})}_{\beta_{sm} \equiv \text{ identified message}} \tag{2}$$

Here, the first term in the right hand side captures the unconditional effect of the sources identity on the audience's belief about ω. This is the source effect for an irrelevant message, which is also equivalent to the isolated source effect across all possible messages, and corresponds to what is commonly referred to as the *source priming* effect. The second term, therefore, measures the effect of the immigration message m for the audience about ω, *knowing* that m was delivered by source s, which we call the *identified message* effect of m from source s.

Bayesian Interpretation of Source Priming Effect. The elicited belief from a treatment group about $\omega \in \Omega$ that hears a message m^0 which is irrelevant to immigration from a source s is given by $P(\omega, m^0|s) = P(\omega|s)$. To interpret how this belief compares to that of a control group, we can then apply Bayes' rule to write this in terms of the odds ratio for the source,

$$P(\omega|s) = P(\omega) \times \underbrace{\frac{P(s|\omega)}{P(s)}}_{\text{odds ratio}\equiv\Theta(s|\omega)} \tag{3}$$

where the left-hand side captures the beliefs of the treated individual, $P(\omega)$ on the right-hand side represents the belief of the control group (or prior), and the third term—the odds ratio of the source—captures the priming effect for the source. It is, however, more useful in this case to think about the treatment effect in terms of independence of outcome and source, rather than the odds ratio itself. In particular, notice that the odds ratio in Equation (3) can be written as:

$$\Theta(s|\omega) = \frac{P(s,\omega)}{P(\omega)P(s)}$$

This expression then delivers the following intuitive prediction:

Remark 1. The beliefs of the primed individual should be the same as the belief of a subject in the control group only if the identity of the source and the outcome ω are subjectively independent:

$$P(\omega,s) = P(\omega)P(s) \Leftrightarrow \Theta(s|\omega) = 1$$

Furthermore, to map this prediction to our formulation in terms of $\ln(1+P(\omega|s))$, we can

[19] Note that by independence we have $P(\omega,m'|s) = P(\omega|s)P(m'|s) \Rightarrow P(\omega|s,m') = P(\omega,m'|s)/P(m'|s) = P(\omega|s)$.

re-write the source priming effect, β_s, in Equation (2) as

$$\beta_s = \ln\left(\frac{1+P(\omega|s))}{1+P(\omega)}\right) = \underbrace{\ln\left(1+\frac{P(\omega)}{1+P(\omega)}\times(\Theta(s|\omega)-1)\right)}_{\text{priming effect of } s\equiv\Delta(s|\omega)} \tag{4}$$

Thus, according to Bayes' law, the term $\Delta(s|\omega)$ characterizes the priming effect of the source: it is increasing in the odds ratio $\Theta(s|\omega)$ and is equal to zero if ω and s are perceived to be independent, i.e. if $\Theta(s|\omega) = 1$. Alternatively, when $\Theta(s|\omega) > 1$, meaning that if the source s is more subjectively favorable conditional on the state of world ω, then the source priming effect is positive in the Bayesian benchmark.

Here, we can easily see the implications of Remark 1. Individuals who subjectively believe that the identity of a particular leader is linked to a particular state of the world will be subject to priming effects when they are exposed to *any* message from that leader. An example could be an individual who updates their beliefs to become more anti-immigrant when they are exposed to an uninformative message from President Trump, if this individual subjectively links Trump to the likelihood that immigration is unfavorable. The opposite could also hold. That is, exposure to an uninformative message from President Trump could increase the favorable view of immigration, perhaps for the Democrat participants in the opposing party. However, if the individual believes that the likelihood of being exposed to an uninformative message from a president is independent of the true state of the world regarding immigration, she will not update her beliefs as a result of the source priming effect.

Bayesian Interpretations of Identified Message Effect. Similarly, to interpret the identified message effect, we can use Bayes' law to write

$$P(\omega|m,s) = P(\omega|s) \times \underbrace{\frac{P(m|\omega,s)}{P(m|s)}}_{\text{odds ratio}\equiv\Theta(m|\omega,s)} \tag{5}$$

where the left-hand side captures the beliefs of an individual treated with message m from source s, $P(\omega|s)$ on the right-hand side represents the belief of the control group that was treated with an irrelevant message m^0 from the same source (so that their belief is $P(\omega|m^0,s) = P(\omega|s)$), and the third term—the odds ratio of the message m conditional on the source s—captures how the likelihood of the message is different in the particular state of the world ω. It is then easy to see that the beliefs of control and treatment groups in this experiment are different only if this likelihood is different in different states of the world, i.e. $P(m|\omega,s) \neq P(m|s) = \sum_{\omega'\in\Omega} P(\omega')P(m|\omega',s)$. In particular, we can derive the following remark.

Remark 2. For $\omega \in \Omega$, let $\neg\omega$ denote the event where the outcome is not ω, then the beliefs of the treatment and control groups in the experiment above are different if and only if the likelihood of the message m coming from source s is different conditional on ω versus $\neg\omega$:

$$P(m|s) \neq P(m|s,\omega) \Leftrightarrow P(m|s,\neg\omega) \neq P(m|s,\omega)$$

Finally, similar to our exercise above, we can show that the identified message effect, β_{sm}, is increasing in the likelihood ratio $\Theta(m|\omega,s)$ and is non-zero if and only if this likelihood ratio is different than one. IWe can interpret this effect as moving individuals when they believe that a president's statement provides new information about the true state of the world regarding immigration, relative to a benchmark of exposure to hearing that president's voice, or adjusted for any source priming effect.

2.1.1 Decomposition II: Fixing the Message

Alternatively, fix a message about immigration, $m \in M$, and consider a second treatment group that is treated with the political message m but from an anonymized source—i.e., a voice actor who reads the message m from a potential political source $s' \in S$ without revealing the exact identity of the source. Then, this group's belief corresponds to the joint beliefs about the outcome ω and sources s' summed over all possible recent presidents that could be the source of m: $P(\omega|m) \equiv \sum_{s' \in S} P(\omega, s'|m) = \sum_{s' \in S} P(s'|m)P(\omega|s', m)$. Thus, this treatment allows us to perform the following decomposition of Equation 1:

$$\ln\left(\frac{1+P(\omega|m,s)}{1+P(\omega)}\right) = \underbrace{\ln\left(\frac{1+P(\omega|m)}{1+P(\omega)}\right)}_{\beta_m \equiv \text{ anonymous message}} + \underbrace{\ln\left(\frac{1+P(\omega|m,s)}{1+P(\omega|m)}\right)}_{\beta_{ms} \equiv \text{ source persuasion}} \tag{6}$$

Here, the first term in the right hand side measures the pure effect of the statement's message or content when the source is anonymous or the *anonymous message* effect; while the second term, fixing the content of m, measures the contribution of knowing the source's identity s to the total treatment effect—i.e. how much knowing the source's identity further persuades or dissuades the audience to change their beliefs about immigration. Accordingly, we refer to the second term as the *source persuasion* effect.

Bayesian Interpretation of Anonymous Message Effect. Similar to the exercise for the identified message effect, we can use the Bayes' law for interpretation of the anonymous message effect:

$$P(\omega|m) = P(\omega) \times \underbrace{\frac{P(m|\omega)}{P(m)}}_{\text{odds ratio} \equiv \Theta(m|\omega)} \tag{7}$$

where the only difference is that these terms now capture beliefs that are aggregated across all possible sources, rather than a particular source s. In particular, the left-hand side captures the beliefs of an individual treated with message m from an anonymous source (so that $P(\omega|m) = \sum_{s \in S} P(s|m)P(\omega|s,m)$) sums over all possible sources), $P(\omega)$ on the right-hand side represents the belief of the control group that was not treated, and the third term—the odds ratio of the message m—captures how the likelihood of the message is different in the particular state of the world ω. It is then easy to see that the beliefs of control and treatment groups in this experiment are different only if this likelihood is different in different states of the world, i.e. $P(m|\omega) \neq P(m) = \sum_{\omega' \in \Omega} P(\omega')P(m|\omega')$. Formally, we can derive the following remark.

Remark 3. For $\omega \in \Omega$, let $\neg\omega$ denote the event where the outcome is not ω, then the beliefs of the treatment and control groups in the experiment above are different if and only if the likelihood of the message m is different conditional on ω versus $\neg\omega$:

$$P(m) \neq P(m|\omega) \Leftrightarrow P(m|\neg\omega) \neq P(m|\omega)$$

Finally, similar to our exercises above, we can show that the anonymous message effect, β_m, is increasing in the likelihood ratio $\Theta(m|\omega)$ and is non-zero if and only if this likelihood ratio is different than one. Here, the implications of the anonymous message effect are that individuals will update their beliefs if they subjectively believe that the information contained in the message is linked to the true state of the world on immigration, or they find the message content to be a substantive signal on this issue.

Bayesian Interpretation of Source Persuasion Effect. Consider now the belief of a treatment group who are treated with a message m coming from a source s. Then, as discussed in Decomposition II, the belief of this group for a given ω, i.e. $P(\omega|m,s)$, relative to the belief of a group that is treated with the same message but from an anonymized source, i.e. $P(\omega|m)$, gives us the persuasion effect of the source s. Now, we can use Bayes' law to write the relationship between these two in terms of the odds ratio of the source for the outcome ω, *conditional* on the fixed message m:

$$P(\omega|m,s) = P(\omega|m) \times \underbrace{\frac{P(s|\omega,m)}{P(s|m)}}_{\Theta(s|\omega,m)\equiv\text{odds ratio of } s|m} \tag{8}$$

Using this equation, we can characterize the conditions under which source persuasion exists—i.e., the odds ratio deviates from unity, $\Theta(s|\omega,m) \neq 0$:

Remark 4. For $s \in S$, let $\neg s$ denote the event that the source is not s. Then, there is source persuasion if (1) revealing the identity of the source is surprising *and* (2) the source is more reliable than other sources for the outcome ω:

$$\Theta(s|\omega,m) > 1 \Leftrightarrow \underbrace{(1 - P(s|m))}_{\text{surprise}} \times \underbrace{(P(\omega|s,m) - P(\omega|\neg s,m))}_{\text{(relative) reliability of source}} > 0 \tag{9}$$

Note that while the surprise term is always either positive or zero, the reliability term can either be negative or positive, and this term governs the direction of persuasion. Moreover, if either of these terms are zero, i.e. either revealing the source of a message is not surprising or the source is not more or less reliable than others, there is no persuasion in the Bayesian benchmark.

To map this prediction to our formulation in terms of $\ln(1+P(\omega|m,s))$, we can re-write the source persuasion effect, $\beta_{m,s}$ in Equation (6) as

$$\beta_{ms} = \ln(\frac{1+P(\omega|m,s)}{1+P(\omega|m)}) = \underbrace{\ln(1+\frac{P(\omega|m)}{1+P(\omega|m)} \times (\Theta(s|\omega,m)-1))}_{\text{persuasion effect of } s \equiv \Delta(s|m,\omega)} \tag{10}$$

where $\Delta(s|m,\omega)$ characterizes the source persuasion effect β_{ms}: it is increasing in the odds ratio $\Theta(s|m,\omega)$ and is equal to zero when there is no persuasion. More generally, $\Delta(s|m,\omega)$ allows the source to amplify the effect of the message m on beliefs about ω through the surprise and reliability channels as discussed above.

Again, we can characterize the prediction using an example. Source persuasion effects will only induce updating for surprising messages that come from sources that the individual views as reliable. A surprising message could be one where the immigration message contrasts with the partisan reputation of the leader, i.e. an anti-immigrant message that comes from President Obama, or a pro-immigrant message that comes from President Trump. An individual may find a partisan leader to be more subjectively reliable than an ambiguous source if that leader represents their political party, i.e. Republicans may find President Trump to be more reliable while Democrats may find President Obama to be more reliable.

3 Empirical Framework

This study uses a randomized controlled experiment, and the empirical approach will use simple comparisons of treatment arms that leverage this randomization.

Our first objective is to test the total impact of an anti-immigration or a pro-immigration statement from a party leader on participant beliefs about immigration. To do this, we compare the president recordings of a particular immigration message to the control group that received no audio treatment, separately by political party, p. The regression corresponds to:

$$ln(P(y_i)+1) = \beta_0 + \beta_t 1[Message_i = 1] \times 1[Source_i = 1] + \gamma X_i + \epsilon_i \tag{11}$$

The coefficient β_t comprises the *total* combined impact of a statement, including both the source and message effects. We denote $Message_i = 1$, as treatment arms which discuss immigration topics, and $Source_i = 1$; as treatment arms where we have the presidents delivering the speeches. Outcomes are constructed as $ln(P(y_i)+1)$, where the outcome, $P(y_i)$, is the index of probability anti-immigrant (pro-immigrant) for Republicans (Democrats). The intercept, β_0, is the average belief in the control group. In our preferred specifications, we also include a vector of pre-treatment control variables selected from a double lasso procedure, X_i, to increase precision of the estimates (see Online Appendix A4).[20] As discussed below and shown in Figure A5, the results are robust to excluding these controls.

Next, we decompose this total effect, β_t, in two alternative approaches. In the first, we decompose the effect between source priming, or the effect of a source when content is not relevant; and the effect of the immigration speech voiced by Obama or Trump, conditional on hearing the president. That is, we fix the president source as either Trump or Obama's voice, and vary the message between the turkey pardon and an immigration speech. Each regression includes a president immigration speech, a president turkey pardon speech, and the "no audio" control group. The relevant regression is:

$$ln(P(y_i)+1) = \beta_0 + \beta_s 1[Source_i = 1] + \beta_{sm} 1[Source_i = 1] \times 1[Message_i = 1] + \gamma X_i + \epsilon_i \quad (12)$$

Where the term $1[Source_i = 1]$, includes both treatment groups with a president's voice, or the turkey pardon and the immigration speech, and the second term, $1[Message_i = 1] \times 1[Source_i = 1]$, includes only the president's immigration speech. This approach decomposes β_t into β_s and β_{sm}. We interpret β_s as source priming, or the difference between beliefs after listening to the turkey pardon audio, relative to the control group; and β_{sm} as the effect of an immigration statement from an identified president, net of priming effects, or relative to the president turkey pardon message. We refer to this residual effect, β_{sm}, as the identified message effect.

In our second exercise, we turn to the central decomposition of the paper. Here, we decompose the total effect, β_t, between the anonymous message effect, β_m, and the source's persuasion power for that message, β_{ms}. For this decomposition, we fix the message of a speech, and vary the source between the actor and the president. Each regression includes a president immigration speech, the voice actor version of the same speech, and the "no audio" control group. The regression equation is:

$$ln(P(y_i)+1) = \beta_0 + \beta_m 1[Message_i = 1] + \beta_{ms} 1[Message_i = 1] \times 1[Source_i = 1] + \gamma X_i + \epsilon_i \quad (13)$$

Where $1[Message_i = 1]$, includes both treatment groups with a particular message, or the actor and president speech versions, and $1[Message_i = 1] \times 1[Source_i = 1]$, includes only the president's version of the immigration speech. Here β_m captures the effect of an anonymous message, that is an immigration effect delivered by the actor; and β_{ms} captures the source

[20] The procedure consists of first including all potential control variables from baseline and converting categorical variables into sets of indicators. Our set of potential controls is quite rich given the large number of survey questions we ask to participants prior to the treatment. We keep the controls selected from the LASSO minimization of a regression which predicts treatment assignment. We then repeat this exercise for a regression that predicts the outcome variable and take the union of controls selected from both procedures as our X_i controls. The list of controls included in the models is outlined in Online Appendix A4.

persuasion effect, or the difference in beliefs from listening to the same words, delivered by the president relative to the actor.

4 Experiment Setting and Structure

4.1 Survey Instrument

We recruited participants who are eligible voters and identify as either Republicans or Democrats through a survey aggregation company called Cloud Research. Cloud Research partners with a number of different online survey panels to compile a sample that targets particular demographic groups and ensures basic quality standards for participant responses. We targeted a participant pool that would mimic the demographic characteristics of the Democratic and Republican parties, such that our results would best represent responses of individuals in these parties.

Prior to treatment, we ask a series of demographic questions, including gender, race, age, education, employment status, party affiliation, and the candidate that the participant supported in the 2016 presidential election. We additionally ask a background question on four different political issues; immigration, gun control, abortion, healthcare, and taxes, as well as which of these issues is most important to the respondent's vote. We also ask how often participants receive their news from different modes (e.g. newspaper, TV, facebook), and different sources (e.g. Fox News, MSNBC). Lastly, we ask participants to state whether they are "fans of" a list of public figures and celebrities, which includes both Donald Trump and Barack Obama. Our questions about views on immigration and approval of Presidents Trump and Obama are embedded within larger lists of policy issues and public figures so as not to prime participants prior to treatment.

After treatment, we ask several questions about participant immigration views. We ask whether participants favor or oppose proposals to: expand construction of a wall on the U.S./Mexico border, hire more border patrol agents, require businesses to check the immigration status of workers, deport all immigrants living in the U.S. illegally, deport the subset of this population with a criminal record, allow immigrants living in the U.S. illegally to become citizens, or allow the subset of this population who came to the U.S. illegally as children to become citizens (DACA recipients). We also ask whether participants view the following groups as positively or negatively contributing to U.S. society: immigrants working legally, immigrants not working legally, immigrants from English-speaking countries, immigrants from Spanish-speaking countries, and "dreamers" or undocumented immigrants brought to the U.S. by their parents as children (DACA recipients). Lastly, we ask participants whether immigrants benefit the economy, commit a disproportionate share of crimes, and if they feel immigration should be increased, decreased, or kept at present levels, as well as their overall perception of the contribution of immigrants to the U.S.

Each of these questions have responses that are collapsed into either a pro-immigrant response, an anti-immigrant response, or a neutral response. We scale each question to have responses that range from 0 to 1, where 0 is a pro-immigrant answer, 1 is an anti-immigrant answer, and 0.5 is a neutral response. We then average across 16 questions to construct our anti-immigration views index. We use an outcome of an anti-immigrant index for Republicans and a pro-immigrant index for Democrats. The pro-immigrant index is simply calculated as 1 minus the anti-immigrant index.

Our post-treatment questions also include several questions about the treatment itself. We ask participants who heard an immigration treatment speech how anti-immigrant or pro-immigrant they perceived the speech to be. We also ask participants which president they thought gave the original speech out of a choice of the four most recent presidents (and an other option). For individuals in the ambiguous source group, or those that received the actor versions of the speeches, this question solicits the participants' best guess of the original source of the speech.[21]

4.2 Sample Restrictions

The survey begins with several screener questions. These include asking for consent to participate, verification that the participant is not a robot and is able to listen to audio content on their device, and assurance that the participant is a U.S. citizen who is eligible to vote. Further, we include a quality check that simulates comprehension of the audio speech treatment. This check asks participants to listen to an audio segment of a weather forecast and asks comprehension questions about what was discussed. For participants randomized to listen to an audio speech treatment in the experiment, we also ask comprehension questions about the subject matter of the treatment speech. The full survey instrument is available in Supplemental Materials Appendix S1.

Figure A1 depicts the sample restrictions that were used to arrive at the final sample used in analysis. There were 19,780 individuals who completed the survey and passed the initial recruitment criteria, which included consenting to the survey, having audio capability, being a U.S. citizen, and declaring in the survey that they were affiliated with the same party they said they were affiliated with in the recruitment advertisement.[22] Some individuals attempted the survey more than once, and we keep only the first attempts for these participants, dropping 1,459 observations. Next, we geocode the locations of the IP addresses of survey-takers and keep only the respondents located in the U.S., dropping 771 individuals. Within the survey, we also

[21]This question also serves as an attention check for individuals who received a speech with a revealed president source; and in fact, nearly all participants in these groups guess the president correctly in this case (Figure A2).

[22]A sub-set of individuals state that they are Independents (or neither Republican or Democrat) after starting the survey; we keep these individuals in the primary sample and exclude them in a robustness check.

drop individuals who fail the quality check related to the weather forecast audio clip, or 3,526 individuals. We also remove a small number of individuals who are treated and did not answer the treatment comprehension questions correctly, or 465 individuals. Next, we remove 593 individuals who took the survey exceptionally quickly or slowly, keeping those who completed the survey in 4 to 30 minutes. Lastly, we remove 93 individuals who did not answer all of the pre-treatment demographic questions. The final sample contains 12,873 individuals, of which 6,992 are Democrats and 5,881 are Republicans. This translates to treatment group sizes of approximately 635 individuals in the Democrat sample and 535 individuals in the Republican sample.

4.3 Sample Characteristics and Balance

Political party affiliation is an increasingly important group identifier in the U.S., where the political climate is highly polarized. We deliberately constructed our experiment to separately measure effects by party affiliation to account for these differences in ideology and group identity. Figure 2 plots the distribution of the key outcome, the anti-immigration beliefs index, for the control group that received no audio treatment. Republicans clearly hold views that are more anti-immigrant than Democrats in our sample, with the mass of each party separated along the distribution of this outcome.

Table 1 further summarizes the outcome index and each of the question components for the control group, or participants who were not assigned to any audio speech treatment. There is notable variation in the views across the component questions. This pattern highlights the strength of our survey in capturing multi-dimensional views on immigration; had we followed earlier literature and focused on a single outcome question, we would have missed potentially important variation in beliefs. We incorporate the views across questions by constructing an average index; however, we also explore patterns across questions in Section 7.

Figure A3 shows that our sample is quite similar to the make-up of the national Democratic and Republican parties. One of our pre-treatment questions is lifted from the national Gallup survey on immigration attitudes and asks whether participants think that immigration levels should be increased, decreased, or be kept at present levels. Panel A and B show that our sample is similar to but slightly more moderate than the national Gallup survey data for each party, with larger share of Republicans and Democrats who feel that immigration should be kept at present levels. Panels C and D show that the study sample is quite similar to the national parties in demographics, though the sample is slightly more likely to be young, white, female, and more educated. These differences may partly reflect the fact that the study pool was recruited for an online survey, which tends to attract individuals who have greater access and comfortability with technology and may also be more likely to be young and educated.

The summary statistics for both samples are shown in Table 2. As noted above, the Republican and Democrat samples are designed to reflect the demographics of their respective national parties. Likewise, the Republican sample has a higher share of respondents who are white, older, and from the southern U.S., while a smaller fraction are college educated. 36% of the Republican sample regularly watches FOX News, compared to 14.5% of the Democrat sample. The parties are predictably stratified in their support of Trump and Obama; 92% (9%) of Republicans (Democrats) voted for Trump in 2016, while 78% (16%) of Republicans (Democrats) are fans of Donald Trump. Likewise, 89% of Democrats are fans of Obama, as compared to only 17% of Republicans. Our sample has a high voter participation and is likely engaged in politics, as over 80% of both sample groups voted in the 2016 election. 15% of Republicans view immigration as their top political issue, compared to 5% of Democrats. Consistent with anticipated party views, 45% of Republicans believe that immigration should be decreased, compared with only 14% of Democrats.[23]

Table 2 also shows that the study sample is balanced across treatment groups. For each demographic characteristic and party sample, we regress the characteristic on indicators for the 11 treatment arms in the study and calculate the joint significance of these indicators. Successful randomization will be associated with a lack of joint significance for these treatment indicators. Nearly all tests pass balance and do not show statistical significance. Only 5 of 62 tests are significant at the 10% level and 3 at the 5% level, similar to the number we should expect to fail due to chance alone ($\approx$ 6 and 3 tests, respectively).

5 Results

5.1 Total Effect

Our first set of results relate to the total impact of partisan statements on immigration beliefs. Table 3 displays the differences between treatment groups that heard a president version of an anti-immigrant or pro-immigrant speech relative to the control group that did not hear any speech. Our results are consistent with findings from other work; partisan statements from political leaders do move beliefs. We measure immigration beliefs in the direction of each party's policy leaning. For Republicans we measure outcomes as anti-immigration beliefs, and for Democrats to pro-immigration beliefs. Of course, each measure is simply one minus the other one.

Overall, these results show that participants from both parties update in the *direction* of the message that they hear. A pro-immigrant message increases pro-immigrant beliefs, while an anti-immigrant message increases anti-immigrant beliefs. We find that movements *away*

[23] These views are taken from the single pre-treatment question on immigration views asked of all participants.

from the party prior appear to be larger and more significant; pro-immigrant effects are larger for Republicans and anti-immigrant effects are larger for Democrats, and we find effect sizes are symmetric across parties. For Republicans (Democrats), pro-immigrant (anti-immigrant) messages decrease their subjective probability that immigration is unfavorable (favorable) by 3-7%. In addition to estimating differences in the average effect across groups, we find similarly significant differences in the distributions of the immigration index across treatment groups, using Kolmogorov-Smirnov tests of equality.

5.2 Decomposition I: Fixing the Source

Table 4 and Figure 4 show the results that decompose the impact of a political statement between source priming (β_s) and an identified message (β_{sm}). We find there is no priming effect in any of the treatment groups. This null finding is present in both the average differences (coefficient significance) and the Kolmogorov-Smirnov tests of distributional equality. Neither party changes their immigration views after listening to a non-ideological (turkey pardon) message from either leader, when compared to the control group that does not hear any speech. This means that, for both parties, the bulk of the updating we observe from the total effect loads on the coefficient of the identified message (β_{sm}). Figure 6 plots decomposition between these two channels. The light blue and red bars, which represent the the identified message effect (β_{sm}), drive both the direction and the size of the total effect, whereas the darker bars, which represent the source priming effect are always negligible and insignificant.

As discussed in the introduction, *ex ante*, it is unclear the extent to which individuals may change their beliefs as a result of any priming effects related to leader reputation on immigration. Prior work has found that reminding individuals of their ethnic, religious, and/or cultural affiliation can meaningfully alter behavior (e.g. Akerlof and Kranton, 2000; Callen et al., 2014; Cohn et al., 2015; Benjamin et al., 2016). It is reasonable to think that cues about political affiliation and ideology through exposure to leaders could have similar effects. The fact that we find no source priming effect implies that individuals are not reminded of the favorability of a particular belief system about immigration when they hear the voice of a partisan president or leader. This decomposition allows us to rule out this channel empirically and focus our attention on the parallel decomposition that fixes the message content and varies the source of the statement.

5.3 Decomposition II: Fixing the Message

Our second decomposition separates the total effect of the partisan statements between the effect of an anonymous message (β_m) and the source persuasion effect (β_{ms}). In this case, in Table 5 and Figure 5, we estimate that in nearly all cases, the effect of the anonymous message β_m, drives a substantial portion of the total effect. Again, the results are strikingly parallel across parties. Both Republicans and Democrats update their priors on immigration in the direction of a message from an anonymous source.

The findings for the importance of message content, given by β_m, are notable for two reasons. First, the rhetoric used in the messages for this experiment is emotional and political, as the speech statements focus on immigration views and general policy proposals and do not include any factual information or substantive content about immigration. Again, *ex ante*, it is unclear that such statements would contain enough new information to move respondent beliefs, especially in a polarized political environment. Second, messages move individuals more when the messages oppose an individual's prior; Democrats are more moved by anti-immigration messages and Republicans are moved more by pro-immigrant messages. Further, speech segments that originated from an opposition leader have the capacity to move beliefs; Democrats (Republicans) find the message substance of the Trump (Obama) anti-immigrant (pro-immigrant) speech to be convincing when they are not aware of the source of the statement. Collectively, these results imply that individuals are capable of updating beliefs when presented with new viewpoints, even when these arguments are not supported by rigorous facts or documentation. These findings suggest that increasing partisan polarization may be partly caused by a lack of exposure to opposition views.

Turning to the role of source persuasion, which captures the added effect of a specific source delivering a particular message, we find that this channel is only important in one symmetric case: when a party leader delivers a message against party lines to members of his party. That is, Obama (Trump) has a persuasion effect only for Democrats (Republicans) when giving a speech that is anti-immigrant (pro-immigrant). In these cases, source persuasion effects explain a substantial share of the total updating of participants: 58% of the total 7% change in beliefs for Republicans, and 44% of the total 7.6% change in beliefs for Democrats (Figure 7). Figure A4 plots the sample group outcomes for the cases with source persuasion effects, and shows that the entire distribution shifts as a result of the leader source delivering these statements in both symmetrical cases.

The source persuasion findings underscore the importance of particular messengers delivering partisan statements. Our results show that party members will "follow their leader" and update their policy views when presented with new and surprising information from a leader that they support. The findings show that leaders have the power to affect the beliefs of followers, and that party affiliates do not simply support leaders when they mirror their pre-existing beliefs.

5.4 Decomposition II: Mechanisms for the Source Persuasion Effect

The conceptual framework predicts that the strength of leader persuasion will be a function of two factors, the surprise of the message coming from a particular source and the subjective reliability of the source. Surprise of the message is the participant's belief that the actual source was unlikely given the message, or whether the participant *does not expect* that a particular source would deliver a message of a certain type. Specifically, this is characterized as 1 minus the perceived probability of the actual or true source conditional on the message, or $(1 - P(s|m))$. Ex ante, it is not clear that participants will find any message to be surprising, as it could be the case that participants think that all politicians can easily change positions and any message type from any leader is equally likely.

In our setting, we can examine actual data on the surprise of the message using participant guesses about the identity of the true source when they heard an actor message. Panel A of Figure 3 shows our measure of message surprise: the share of respondents in the actor groups who guessed the true president *incorrectly* when asked which president they thought gave the speech (out of four recent presidents and an "other" option). The plot shows that the most surprising messages are those where a leader delivers a message that opposes their reputation on immigration. The anti-immigrant Obama speech and the pro-immigrant Trump speech have more surprising sources from the perspective of participants in the study.

The conceptual framework illustrates that the second factor that drives leader persuasion is the subjective reliability of the source. Again, prior to conducting this experiment, it is not clear that participants will view some leaders as more reliable messengers of information than others. If individuals are persuaded solely by the information in messages, it could be the case that they prefer particular leaders only because those leaders have platforms that reflect what the individuals already believe.

We are not able to directly measure subjective reliability in our data, but we are able to illustrate differences in the favorability of leaders across parties which are likely correlated with reliability. Panel B of Figure 3 summarizes pre-treatment support for presidents across parties. Republicans overwhelmingly voted for Trump in 2016, with a share of 91.5%, in contrast to only 9% of Democrats. Similarly, when participants are asked whether they are fans of Obama and Trump, the vast majority of Republicans state that they are fans of Trump (78%) and a minority are fans of Obama (17%). A similar story is present for Democrats, with 89% stating they are fans of Obama, versus only 6% that are fans of Trump. This pattern suggests that Republicans may find Trump to be more subjectively reliable than Obama, and vice versa for Democrats.

The source persuasion effects that we find are consistent with the predicted importance of these two factors in the conceptual framework. The persuasion effect is a joint or multiplicative function of surprise and subjective reliability; this implies that a statement must be both surprising (coming from the particular source) and the leader must be perceived as reliable for an effect to be present. We find this pattern in our results: participants are only persuaded by counter-reputational messages from a leader that they support, i.e. Trump's pro-immigrant speech for Republicans and Obama's anti-immigrant speech for Democrats. We see null effects in source persuasion when a statement is surprising, or counter-reputational, but is delivered by the opposition leader, who may be perceived as less reliable, i.e. Trump's pro-immigrant speech for *Democrats* or Obama's anti-immigrant speech for *Republicans*. Likewise, while it is likely that each party finds its own leader to be reliable, statements that are consistent with leader reputation are unsurprising and leaders do not persuade in this case, i.e. Trump's *anti-immigrant* speech for Republicans and Obama's *pro-immigrant* speech for Democrats.

5.5 Decomposition II: Implications for Polarization

Leveraging the randomization of the experiment, we can also explore how the distance between Republican's and Democrat's immigration beliefs, or partisan polarization, would be affected under different alternative scenarios.[24] We calculate the change in polarization by estimating a regression using the anti-immigration index as the outcome that includes both parties in the sample. Each regression includes a treatment group for Democrats and Republicans relative to a corresponding control group for both parties, and records the change in distance between the two groups in the treatment groups versus the control group. Figure 8 and Table A1 display the results of this exercise. Given our findings that people move in the direction of the messages that they hear, we show that polarization increases when participants hear messages consistent with their priors and decreases when participants hear messages that oppose their priors.

First, we consider a counterfactual that is most likely to arise in the real world, the case when party leaders deliver statements to their respective followers which are consistent with the priors of their followers (Own Leader, With Prior). These messages increase polarization by 10%, and this effect is the result of the anonymous message effect, or message content. Statements consistent with party priors that come from the opposition leader also increase polarization, but these point estimates are substantially smaller and not significant.

Partisan statements that are contrary to the priors of participants decrease polarization. The against-prior anonymous message effects from *either* the party leader or the opposition leader both decrease polarization by 12%. As discussed above, the source persuasion effect amplifies the impact of the against-prior messages only for party leaders, leading to an extra decrease in polarization of 14%, or a total decrease in polarization of ≈ 30%. In contrast, against-prior statements are not more persuasive when they come from the opposition leader, and here, the anonymous message effect dominates. In sum, these results imply that party leaders have the capacity to move their followers in a new alternative direction, and in this setting, yields the potential to reduce political polarization.

[24]Throughout this paper, we use the term "polarization" to refer specifically to partisan/party sorting or partisan polarization, or the distance in beliefs between individuals of different political parties.

6 Robustness

In this section, we probe the robustness of the baseline findings. We focus only on the results from Decomposition II for brevity, as this decomposition provides the central results of this paper. Figures A5, A6 and A7 show that the results are stable across various specifications and sample restrictions. The first specification in Figure A5 drops the demographic covariates that are included in the baseline model using the double Lasso procedure (See Appendix A4) and shows very similar results. In the second specification, we adjust the sample to exclude observations collected on election day or later (November 3, 2020) to adjust for idiosyncratic features of the political environment at this time, and again find similar results.[25] In our baseline sample, we include the first attempts of the survey for individuals who attempted to take the survey multiple times. As a third test, we check that the results are consistent when dropping the first attempts of all duplicate responders, and again find very similar results.

Next, we vary the construction of the immigration index outcome. Some of the questions in the survey have 5 option responses (e.g. Strongly Agree, Agree, Disagree, Strongly Disagree, No Opinion) and others have 3 option responses (e.g. Yes, No, No Opinion). In our baseline index, we translate all questions into 3 options; a pro-immigrant answer, an anti-immigrant answer, or a neutral response, collapsing the variation in the 5 option questions. In the fourth test in Figure A5, we show the results when we consider the full variation from 5 option questions and find very similar results. Lastly, when we initially pre-registered the experiment, we specified an index that would use only 14 of the 16 post-treatment questions in the survey, because we believed that some of the questions might produce ambiguous responses, given that they related to potentially more favored immigrant populations. These two questions asked whether respondents thought that immigrants from English-speaking countries or immigrants living in the U.S. legally contribute positively or negatively to society. In practice, these questions did not produce anomalous results and thus we added them to our baseline index to increase its information content. However, using only the 14 question version also produces similar results to the baseline estimates (final test in Figure A5).

In Figure A6, we consider alternate formulations of the outcome. In our baseline model, we utilize the $ln(P(y)+1)$ transformation in our baseline regressions so as not to exclude individuals with purely anti-immigrant or pro-immigrant views, where $P(y) = 0$. As a first check, we estimate the model using a simpler outcome of $ln(P(y))$ and find consistent results. As additional checks, we include a specification using the inverse hyperbolic sine of $P(y)$ and a simple linear regression using $P(y)$ as the dependent variable. The results of each of these alternative specifications

[25] We purposefully conducted our experiment during the period leading up to the national election in order to capture political beliefs and attitudes at this time, beginning on October 16, 2020. Due to a slower recruitment pace than anticipated, the study ran until November 10, 2020, and our baseline sample includes all survey dates.

produces results that are similar in direction and significance to the baseline estimates, given that each specification is a monotonic transformation of the key outcome. However, the scale of the values differs in these models as a result of the difference of arguments in the expressions.

Our baseline regressions consider the log of an immigration index that has a continuous support of [0,1], given that it is an average of multiple questions each of which can assume values of 0, 0.5, or 1. This construction has the benefit of incorporating gradations in beliefs about immigration for each participant. However, it is also useful to consider alternative specifications that discretize this outcome and employ logit or probit models (discretized using a value of 0.5). Figure A7 finds results that are quite similar in direction and significance to the baseline estimates using a logit and probit model, as well as when using a linear probability model.

7 By Question Results

Our central results rely on an index composed of 16 different questions to incorporate multiple dimensions of immigration attitudes. Here, we test the robustness of the index to its components, again focusing on the results from Decomposition II for brevity. First, we consider the sensitivity of the index by re-calculating the estimates leaving out one question at a time in Figure A8 and by estimating the models separately for each question in Figure A9. While the results are more disperse when separately estimated by question, across both sets of tests the findings are quite consistent with the baseline index outcome.

Figures A10 and A11 display the results for each individual question. The results are noisier when estimated at the individual question level, but suggestive patterns emerge. Focusing on the source persuasion effects, Republicans become more pro-immigrant (for the Trump pro-immigrant treatment) on questions related to their "overall view" of immigrants, providing a path to citizenship for immigrants residing in the U.S., especially for immigrants who came to the U.S. illegally as children ("Dreamers" or DACA recipients), and their views on whether all immigrants residing in the U.S. illegally should be deported. This pattern implies that Republicans are able to be persuaded to become more pro-immigrant on questions related to providing a legislative solution to citizenship, with a focus on providing this path to individuals who came to the U.S. as childrenArguably, these questions capture aspects of the immigration debate that are the most modest, or most open to support from right-leaning individuals.

For Democrats, we observe a somewhat symmetric pattern in the source persuasion effect (for the Obama anti-immigrant treatment) whereby partisans are more likely to become anti-immigrant on immigration issues that penalize immigrants who may be viewed as the most egregious rule-breakers. Here, we observe anti-immigrant movement on questions related to the normative societal contribution of immigrants living in the U.S. illegally, the economic contribution of immigrants, interest in deporting immigrants living in the U.S. illegally, especially those with a criminal record, the expansion of immigration background checks for workers, and increases in funding for U.S. border patrol agents. This pattern of findings suggests that Democrats may be prone to adopting anti-immigrant attitudes towards immigrants that may be commonly viewed as the most severe "bad actors."

The "by question" results also highlight a crucial methodological strength of the study: the fact that we are able to aggregate responses of multiple dimensions of immigration attitudes into a summary index. Immigration is a complex policy topic, and partisans may change their beliefs on some aspects of immigration and not others. Additionally, using information from multiple distinct questions increases the total precision of our outcome. Had we taken the approach of earlier work and considered only 1-2 outcome questions with a general focus, we very well could have missed the robust and symmetric effects that we observe for both parties in our study.

8 Heterogeneity

Our experiment was designed to specifically address differences in reactions to treatment along political party lines, which we hypothesized would be the critical dimension of heterogeneity for our study. In line with our hypothesis, the results are symmetrical and mirrored by party, as party affiliation appears to be the key characteristic that predicts both prior beliefs about immigration as well as responses to different leaders.

There are several additional dimensions of heterogeneity that are worth exploring within party group, and we examine these dimensions to the extent possible given our sample size and statistical power. Examining sub-groups of the data also permit insights about some of the decision models hypothesized to underly political belief formation that have been referenced in earlier work; namely, (1) motivated reasoning models, which predict that responses will vary by the strength of an individual's prior belief about immigration (or their party allegiance), and (2) dual-processing models, which predict that responses will vary by the attention or focus of the participant. Again, we focus on the results from Decomposition II for these sub-analyses.

First, we test the importance of the strength of party affiliation as well as the strength of prior beliefs about immigration in Figure A12. We do this by including results for alternate samples that exclude individuals who voted for the opposite party candidate in 2016 ("Flip Voters"), individuals who answered the pre-treatment immigration question by stating that immigration should be increased (decreased) who were Republicans (Democrats) ("Anti-Party Views"), individuals who were recruited as Democrats or Republicans but then identified as independents within the survey (Independents), individuals who stated that they were not a fan of their own party president (Non-Fan President), and lastly individuals who were predicted to have far right or left views on immigration using information on all pre-treatment questions in the survey ("Polarized").[26] In a motivated reasoning model, we might expect movement to the opposing position to be more likely among individuals with lower levels of attachment to their party prior. However, across each of these sub-groups of the data, the estimates are nearly identical to the baseline sample that includes all participants. An exception is that the message effects tend to move participants leftward for polarized Republicans, a feature that could be a result of these individuals having extreme right priors before treatment that do not permit them to move rightward to the same degree as other groups. Notably, this outlier result contrasts with the predictions of a motivated reasoning model which would anticipate larger leftward movements for moderate Republicans. Overall, the stability across groups suggests that neither moderates nor extreme members of each party are driving the results. Rather, consistent with the plots in Figure A4, it appears that there is a full shift of the distribution as a result of treatment that is not stronger at either tail.

Next we test whether differences in political engagement or news consumption matter in Figure A13. In a dual processing model, we might expect participants who are less engaged or attentive to be more likely to update their views in a direction that opposes their priors. Here, we re-estimate the models excluding individuals who consume news from at least two platforms (Newspaper, TV, Twitter, Facebook) and at least daily ("Multiple News Types"), individuals who consume news from both a left-leaning and a right-leaning source at least weekly ("Bipartisan News"), and individuals who did not vote in the 2016 election ("Non-Voters (2016)). Again, the estimates are remarkably stable across groups, suggesting that neither highly informed/engaged nor uniformed/unengaged participants are driving the findings of the study.

Lastly, we explore a number of demographic dimensions of heterogeneity in Figures A14 and A15. Overall, we find very few notable differences in the estimates according to gender, race, employment status, or educational attainment. The confidence intervals of nearly all of the comparison estimates overlap; and we are hesitant to draw meaningful conclusions from suggestive differences in point estimates given limitations due to sample size.

In sum, we find limited variation in our findings across several dimensions of heterogeneity *within* party. This result could be attributable to a lack of power to detect these kinds of sub-group differences, as we did not design our experiment to discern these effects given resource constraints. At the same time, while the standard errors do moderately increase with the smaller sub-groups, the point estimates for each cut of the data are quite consistent in practice. This stability stands in contrast to prior work in political science using either motivated reasoning models, which posits that individuals with stronger party affiliation will have stronger effects,

[26] "Polarized" is an estimated characteristic assigned to participants in the sample. We do this by predicting whether an individual has views in the top quartile of the party position (pro-immigrant for Democrats, anti-immigrant for Republicans) using only the control group and pre-treatment question responses. We use the coefficients from the estimation to extrapolate who would be polarized (without treatment) for the whole sample.

or dual processing models, which posits that individuals who are less informed about policy issues will have stronger effects (Bullock, 2020). Instead, it appears that the shifts in beliefs that we observe are present for all segments of the distribution within party. The pattern of findings supports the hypothesis that *party affiliation* is the most important factor that determines responsiveness to treatment.

9 Discussion

How do leaders change the beliefs of their constituents? We leverage a novel randomized experiment to investigate this question in the context of U.S. immigration policy, a topic that is both important to voters and is characterized by a high level of polarization. Our experiment uses audio segments of actual speeches from Presidents Barack Obama and Donald Trump, as well as replicate versions of these speeches recorded by an actor, to isolate: (1) the importance of partisan rhetoric or messages, (2) the priming effects of partisan leader identity, and (3) the persuasive power of leaders for particular messages, holding the content of a message fixed. The experiment is grounded in a Bayesian conceptual framework that provides predictions which inform the results.

Our first key finding is that political messages cause participants to change their views. We deliberately test messages that do not contain facts and are instead composed of emotion-based arguments and general policy proposals. Ex ante, it is not obvious that such messages will change beliefs, given that they contain no new substantive factual information about immigration. Both Republicans and Democrats are swayed by the pro-immigrant and anti-immigrant messages in this study, with effects that are stronger for messages that oppose party priors. An implication of these results is that the highly polarized nature of political beliefs in the U.S. could be related to the segmentation of information exposure by party. Information silos or "echo chambers" may be exacerbating polarization by shielding partisans from alternative information signals.

The aim to reduce partisan polarization is a potential policy goal in and of itself, as the increasingly divergent views across parties may serve to amplify societal divisions and create obstacles to policy progress (Boxell et al., 2020; Schultz, 1996). Researchers have found that multiple factors have contributed to a dramatic rise in polarization in recent decades, including the rise social media and the segmentation of media exposure (e.g. Di Tella et al., 2021; Levy, 2021; Guriev et al., 2021; Allcott et al., 2020; Jo, 2017; Barberá et al., 2015) and the introduction of widely available decentralized propaganda or "fake news" (Azzimonti and Fernandes, 2018). However, despite research on the causes of polarization, little is known about the ways in which polarization could be reversed, and our study contributes new insight on these questions.

We also find evidence that leaders can persuade individuals as messengers of information. While we do not find that participants change their beliefs through simple exposure to the identity of a leader (or source priming), we do find that certain messages are more persuasive when voiced by particular leaders. These persuasion effects are measured as the effect of a leader source, holding fixed the substance of a message. We find that a leader is most persuasive when expressing statements that are unexpected to an audience of individuals who find the leader to be credible. Specifically, President Obama is most persuasive when voicing an anti-immigrant speech to Democrats and President Trump is most persuasive when voicing a pro-immigrant speech to Republicans.

This pattern of results illustrates that supporters will "follow their leader" to new political positions. In our context, these effects lead to a reduction in the distance in beliefs about immigration across parties, or partisan polarization, suggesting that leaders have the potential to play an important role in reducing partisan division.

While we find that leader persuasion reduces polarization, the implications of our results are in fact more general. Our framework implies that surprising or new positions for leaders will be most persuasive to followers, which may or may not increase longer-term polarization depending on the circumstance. These impacts also have the potential to be meaningful in situations where a new issue arises and expectations about policy positions have yet to be set, such as during the early stages of a public health crisis like the recent COVID pandemic (Larsen et al., 2022). A more abstract implication of our general findings is that a leader has the potential to create a "cult of personality" and use his or her persuasive power to shape policy decisions.

Our work is the first to credibly and comprehensively isolate the determinants of the leader influence, focusing on the mechanisms that drive changes in beliefs resulting from exposure to statements from partisan leaders. Future research should continue to dissect the ways in which prominent partisan figures may shape, alter, or disrupt public opinion; as well as the persistence of a partisan leader's influence on constituent beliefs over time. Understanding these dynamics could provide new insights into the strengths and fragilities of democratic governments, which rely on their ability to build consensus among voters and their elected officials.

References

Adena, M., R. Enikolopov, M. Petrova, V. Santarosa, and E. Zhuravskaya (2015). Radio and the Rise of the Nazis in Prewar Germany. *The Quarterly Journal of Economics 130*(4), 1885–1939.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., A. Miano, and S. Stantcheva (2018). Immigration and Redistribution. *National Bureau of Economic Research.* Working Paper.

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). The Welfare Effects of Social Media. *American Economic Review 110*(3), 629–76.

Andreoni, J. and T. Mylovanov (2012). Diverging Opinions. *American Economic Journal: Microeconomics 4*(1), 209–32.

Azzimonti, M. and M. Fernandes (2018). Social Media Networks, Fake News, and Polarization. *National Bureau of Economic Research.* Working Paper.

Bakker, B. N., Y. Lelkes, and A. Malka (2020). Understanding partisan cue receptivity: Tests of predictions from the bounded rationality and expressive utility perspectives. *The Journal of Politics 82*(3), 1061–1077.

Barber, M. and J. C. Pope (2019). Does Party Trump Ideology? Disentangling Party and Ideology in America. *American Political Science Review 113*(1), 38–54.

Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science 26*(10), 1531–1542.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics 90*(1-2), 1–36.

Baum, M. A. and P. Gussin (2008). In the eye of the beholder: How information shortcuts shape individual perceptions of bias in the media. *Quarterly Journal of Political Science 3*(1), 1–31.

Baysan, C. (2021). Persistent Polarizing Effects of Persuasion: Experimental Evidence from Turkey. *American Economic Review.* Forthcoming.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: does exposure reduce bias? *The Quarterly journal of economics 124*(4), 1497–1540.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *science 335*(6068), 582–586.

Benjamin, D. J., J. J. Choi, and G. Fisher (2016). Religious identity and economic behavior. *Review of Economics and Statistics 98*(4), 617–637.

Benoît, J.-P. and J. Dubra (2019). Apparent Bias: What Does Attitude Polarization Show? *International Economic Review 60*(4), 1675–1703.

Bidwell, K., K. Casey, and R. Glennerster (2020). Debates: Voting and expenditure responses to political communication. *Journal of Political Economy 128*(8), 2880–2924.

Boudreau, C. and S. A. MacKenzie (2014). Informing the Electorate? How Party Cues and Policy Information Affect Public Opinion about Initiatives. *American Journal of Political Science 58*(1), 48–62.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2020). Cross-Country Trends in Affective Polarization. *National Bureau of Economic Research.* Working Paper.

Broockman, D. E. and D. M. Butler (2017). the Causal Effects of Elite Position-Taking on Voter Attitudes: Field Experiments with Elite Communication. *American Journal of Political Science 61*(1), 208–221.

Bullock, J. G. (2011). Elite influence on public opinion in an informed electorate. *American Political Science Review 105*(3), 496–515.

Bullock, J. G. (2020). Party cues. In *The Oxford handbook of electoral persuasion*, pp. 129. Oxford University Press, USA.

Bursztyn, L., G. Egorov, I. Haaland, A. Rao, and C. Roth (2022). Justifying dissent. *National Bureau of Economic Research.* Working Paper.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2018). Misperceived social norms: Female labor force participation in Saudi Arabia. *National Bureau of Economic Research.* Working Paper.

Callen, M., M. Isaqzadeh, J. D. Long, and C. Sprenger (2014). Violence and risk preference: Experimental evidence from Afghanistan. *American Economic Review 104*(1), 123–48.

Chiang, C.-F. and B. Knight (2011). Media Bias and Influence: Evidence from Newspaper Endorsements. *The Review of Economic Studies 78*(3), 795–820.

Chopra, F., I. Haaland, and C. Roth (2022). Do people demand fact-checked news? evidence from us democrats. *Journal of Public Economics 205*, 104549.

Cinelli, M., G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences 118*(9), e2023301118.

Cohn, A., J. Engelmann, E. Fehr, and M. A. Maréchal (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review 105*(2), 860–85.

Cohn, A. and M. A. Maréchal (2016). Priming in economics. *Current Opinion in Psychology 12*, 17–21.

Cohn, A., M. A. Maréchal, and T. Noll (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies 82*(4), 1289–1308.

Cruces, G., R. Perez-Truglia, and M. Tetaz (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics 98*, 100–112.

DellaVigna, S. and E. Kaplan (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics 122*(3), 1187–1234.

Di Tella, R., R. H. Gálvez, and E. Schargrodsky (2021). Does Social Media Cause Polarization? Evidence from Access to Twitter Echo Chambers during the 2019 Argentine Presidential Debate. *National Bureau of Economic Research.* Working Paper.

Doherty, C., J. Kiley, and N. Asheer (2020). In Changing U.S. Electorate, Race and Education Remain Stark Dividing Lines. *Pew Research Center*.

Doherty, C., J. Kiley, N. Asheer, and C. Jordan (2020). Election 2020: Voters are Highly Engaged, but Nearly Half Expect Difficulties to Voting. *Pew Research Center*.

Doherty, C., J. Kiley, and B. Johnson (2016). 2016 Campaign: Strong Interest, Widespread Dissatisfaction. *Pew Research Center*.

Druckman, J. N., E. Peterson, and R. Slothuus (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review 107*(1), 57–79.

Durante, R., P. Pinotti, and A. Tesei (2019). The Political Legacy of Entertainment TV. *American Economic Review 109*(7), 2497–2530.

Enikolopov, R., M. Petrova, and E. Zhuravskaya (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review 101*(7), 3253–85.

File, T. (2018). Characteristics of Voters in the Presidential Election of 2016. *U.S. Census.*

Fryer Jr, R. G., P. Harms, and M. O. Jackson (2019). Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. *Journal of the European Economic Association 17*(5), 1470–1501.

Fujiwara, T. and L. Wantchekon (2013). Can informed public deliberation overcome clientelism? experimental evidence from benin. *American Economic Journal: Applied Economics 5*(4), 241–55.

Gentzkow, M. and J. M. Shapiro (2006). Media Bias and Reputation. *Journal of Political Economy 114*(2), 280–316.

Gentzkow, M., M. B. Wong, and A. T. Zhang (2018). Ideological Bias and Trust in Information Sources. *Working Paper*.

Gerber, A. S., D. Karlan, and D. Bergan (2009). Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions. *American Economic Journal: Applied Economics 1*(2), 35–52.

Gilens, M. and N. Murakawa (2002). Elite cues and political decision-making. *Research in micropolitics 6*, 15–49.

Goren, P., C. M. Federico, and M. C. Kittilson (2009). Source cues, partisan identities, and political value expression. *American Journal of Political Science 53*(4), 805–820.

Grigorieff, A., C. Roth, and D. Ubfal (2020). Does information change attitudes toward immigrants? *Demography 57*(3), 1117–1143.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2021). 3g internet and confidence in government. *The Quarterly Journal of Economics 136*(4), 2533–2613.

Haaland, I. and C. Roth (2020). Labor market concerns and support for immigration. *Journal of Public Economics 191*, 104256.

Haaland, I., C. Roth, and J. Wohlfart (2020). Designing information provision experiments.

Igielnik, R. and A. Budiman (2020). The Changing Racial and Ethnic Composition of the U.S. Electorate. *Pew Research Center*.

Jo, D. (2017). Better the Devil You Know: An Online Field Experiment on News Consumption. *Working paper*.

Jones, J. (2019). New High in U.S. Say Immigration Most Important Problem. *Gallup*.

Kalla, J. L. and D. E. Broockman (2018). the Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments. *American Political Science Review 112*(1), 148–166.

Larsen, B., T. J. Ryan, S. Greene, M. J. Hetherington, R. Maxwell, and S. Tadelis (2022). Counter-stereotypical messaging and partisan cues: Moving the needle on vaccines in a polarized us. *National Bureau of Economic Research*. Working Paper.

Levy, R. (2021). Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review 111*(3), 831–70.

Martin, G. J. and A. Yurukoglu (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review 107*(9), 2565–99.

Merkley, E. and D. A. Stecula (2021). Party Cues in the News: Democratic Elites, Republican Backlash, and the Dynamics of Climate Skepticism. *British Journal of Political Science 51*(4), 1439–1456.

Mullainathan, S. and A. Shleifer (2005). The Market for News. *American Economic Review 95*(4), 1031–1053.

Nicholson, S. P. (2012). Polarizing cues. *American Journal of Political Science 56*(1), 52–66.

Ong, A. D. and D. J. Weiss (2000). The impact of anonymity on responses to sensitive questions 1. *Journal of Applied Social Psychology 30*(8), 1691–1708.

Schultz, C. (1996, 04). Polarization and Inefficient Policies. *The Review of Economic Studies 63*(2), 331–344.

Song, L. (2021). The Heterogeneous Effects of Social Media Content on Racial Attitudes. *Working Paper*.

Thaler, M. (2021). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *Available at SSRN 3717381*.

Younis, M. (2020). Americans Want More, Not Less, Immigration for First Time. *Gallup*.

Table 1: Immigration Index

	Republicans *Probability*		**Democrats** *Probability*	
	Anti-Immigrant		*Pro-Immigrant*	
Control Group (No Audio)	Mean	S.D.	Mean	S.D.
Immigration Index	0.579	(0.196)	0.729	(0.196)
Components of Index				
Overall View of Immigrants	0.375	(0.471)	0.828	(0.367)
Immigrant Crime Share	0.601	(0.420)	0.797	(0.349)
Immigrant Economic Impact	0.610	(0.414)	0.781	(0.351)
Ideal Level of Immigration	0.673	(0.346)	0.611	(0.333)
Contribution: Undocumented Immigrants	0.829	(0.298)	0.539	(0.386)
Contribution: Legal Immigrants	0.190	(0.327)	0.915	(0.230)
Contribution: English-speaking Immigrants	0.288	(0.312)	0.846	(0.253)
Contribution: Spanish-speaking Immigrants	0.449	(0.352)	0.792	(0.294)
Contribution: Dreamers	0.435	(0.358)	0.832	(0.278)
Path to Citizenship: Dreamers	0.227	(0.397)	0.896	(0.283)
Path to Citizenship: All Immigrants	0.399	(0.474)	0.861	(0.323)
Deport Immigrants with Crime Record	0.915	(0.255)	0.380	(0.452)
Deport All Immigrants	0.656	(0.442)	0.788	(0.383)
Check Worker Immigration Status	0.914	(0.256)	0.394	(0.451)
Border Patrol	0.878	(0.285)	0.607	(0.448)
Border Wall	0.823	(0.356)	0.800	(0.369)
N	555		661	

Notes: This table summarizes the components of the immigration index outcome for the "no audio" control group. 16 questions are used to construct the index and their topics are listed in the table. Each question typically has 5 option responses (e.g. whether a respondent strongly agrees, agrees, disagrees, strongly disagrees or has no opinion about a statement). Each response is coded as either "pro-immigrant" (e.g. groups strongly agree and agree for a particular statement), "anti-immigrant" (e.g. groups strongly disagree and disagree for a particular statement), or neutral (e.g. "no opinion" for a particular statement). We construct the probability anti-immigrant for each question as "1" for an anti-immigrant response, "0" for a pro-immigrant response, and "0.5" for a neutral response. The complete survey and question wording is available in the Supplemental Materials for this paper. The total anti-immigrant index is the average across all question responses. In regressions using Democrat respondents we use the outcome of a "pro-immigrant index"; the pro-immigrant index is simply 1 minus the "anti-immigrant" index, or $P(Pro-Immigrant) = 1 - P(Anti-Immigrant)$.

Table 2: Summary Statistics and Balance Tests

	Republicans				**Democrats**			
	Mean	S.D.	F-Test	P-Value	Mean	S.D.	F-Test	P-Value
Female	0.469	(0.499)	1.002	0.439	0.511	(0.500)	0.581	0.831
White	0.892	(0.311)	0.701	0.724	0.649	(0.477)	0.743	0.685
Black	0.029	(0.169)	0.868	0.563	0.186	(0.389)	1.140	0.328
Hispanic	0.040	(0.195)	1.261	0.247	0.083	(0.276)	0.620	0.798
Asian	0.029	(0.169)	0.415	0.940	0.066	(0.247)	0.488	0.899
18-24 years	0.081	(0.272)	0.732	0.695	0.154	(0.361)	0.906	0.527
25-34 years	0.184	(0.388)	1.193	0.290	0.257	(0.437)	1.239	0.260
35-44 years	0.219	(0.414)	0.904	0.528	0.218	(0.413)	0.830	0.600
45-64 years	0.324	(0.468)	1.004	0.437	0.235	(0.424)	0.804	0.625
65+ years	0.192	(0.394)	0.233	0.993	0.136	(0.343)	0.500	0.891
Northeast	0.166	(0.372)	0.678	0.746	0.220	(0.414)	0.530	0.871
Midwest	0.241	(0.427)	0.650	0.772	0.213	(0.410)	1.379	0.183
South	0.432	(0.495)	0.883	0.548	0.364	(0.481)	2.111	0.021
West	0.161	(0.368)	0.582	0.830	0.204	(0.403)	0.989	0.450
College or More	0.460	(0.498)	0.597	0.817	0.520	(0.500)	1.101	0.357
Full-time Employed	0.425	(0.494)	1.397	0.175	0.446	(0.497)	1.012	0.430
News (Weekly+): Facebook	0.397	(0.489)	0.992	0.448	0.386	(0.487)	0.576	0.835
News (Weekly+): Twitter	0.175	(0.380)	0.666	0.757	0.263	(0.440)	1.713	0.072
News (Weekly+): TV	0.584	(0.493)	0.781	0.647	0.601	(0.490)	0.752	0.675
News (Weekly+): Newspaper	0.250	(0.433)	1.974	0.032	0.320	(0.466)	1.274	0.239
News (Weekly+): FOX News	0.363	(0.481)	1.060	0.390	0.146	(0.353)	0.834	0.596
News (Weekly+): MSNBC	0.083	(0.276)	0.842	0.588	0.193	(0.394)	0.766	0.662
Party: Independent	0.077	(0.266)	1.672	0.081	0.054	(0.226)	2.174	0.017
Polarized (Estimated)	0.247	(0.431)	1.446	0.153	0.358	(0.479)	1.143	0.325
Voted (2016)	0.842	(0.365)	0.724	0.703	0.836	(0.371)	0.778	0.650
Voted for Trump (2016)	0.916	(0.277)	1.388	0.179	0.086	(0.280)	0.546	0.858
Fan of Trump	0.780	(0.414)	1.479	0.140	0.060	(0.237)	1.143	0.325
Fan of Obama	0.168	(0.374)	0.779	0.649	0.889	(0.314)	0.846	0.584
Immigration: Top Issue	0.156	(0.363)	1.139	0.328	0.051	(0.219)	1.190	0.292
Immigration: Should Increase	0.114	(0.318)	1.032	0.413	0.365	(0.482)	0.876	0.554
Immigration: Should Decrease	0.452	(0.498)	0.898	0.534	0.142	(0.349)	1.431	0.160
N	5881				6992			

Notes: Table summarizes demographic characteristics of sample from questions asked prior to intervention. Balance tests use separate regressions of each demographic characteristic on the full set of treatment group indicators, F-tests refer to joint significance of treatment assignment. "Party: Independent" refers to individuals recruited to the survey as Democrats or Republicans who indicate within the survey that they are Independents. "Polarized (Estimated)" is a probability of being in the top 25th percentile of anti-immigrant (pro-immigrant) views for Republicans (Democrats) constructed in the following way: predict this outcome in the post-treatment immigration index in the no audio control group using only pre-treatment characteristics and then use the coefficients to predict this outcome for the full sample. "Fan of Obama/Trump" comes from a question where we ask whether participants are a fan of these presidents. Prior to treatment, we ask participants to identify the top issue relevant to their vote ("Immigration: Top Issue") and whether they think immigration should be increased, decreased or remain constant ("Immigration: Should Increase (Decrease)").

Table 3: **Total Effect:** Combined Message and Source

		Republicans				
		ln(Probability Anti-Immigrant + 1)				
Message	**Source**	β_t	(S.E.)	%Diff.	*P(Dist.)*	N
Anti	Trump	0.012**	(0.006)	3.31%	*0.070*	1103
Anti	Obama	0.015**	(0.006)	4.18%	*0.004*	1061
Pro	Trump	-0.026***	(0.006)	-7.00%	*0.000*	1106
Pro	Obama	-0.014**	(0.006)	-3.80%	*0.045*	1080

		Democrats				
		ln(Probability Pro-Immigrant + 1)				
Message	**Source**	β_t	(S.E.)	%Diff.	*P(Dist.)*	N
Anti	Trump	-0.010**	(0.005)	-2.47%	*0.106*	1267
Anti	Obama	-0.032***	(0.005)	-7.65%	*0.000*	1283
Pro	Trump	-0.004	(0.005)	-0.97%	*0.305*	1302
Pro	Obama	0.008*	(0.005)	1.88%	*0.004*	1336

Notes: *p<0.1, **p<0.05, ***p<0.01. This table presents the estimates of the different president speech treatments, relative to the "no audio" control group. The relevant regression is $ln(P(y)+1) = \beta_0 + \beta_t 1[Message_i = 1] \times 1[Source_i = 1] + \gamma X_i + \epsilon$. The effects shown comprise the total combined impact of a president message, including both source and message effects. Outcomes are constructed as $ln(P(y)+1)$, where the outcome is the index probability anti-immigrant (pro-immigrant) for Republicans (Democrats). All regressions additionally include covariates selected by the double lasso procedure to improve precision (See Appendix A4). "% Diff." calculates the implied change in outcome *probability*, or the untransformed probability index, due to treatment, relative to the mean for the control group. "$P(Dist.)$" is the p-value from a Kolmogorov-Smirnov test of equality of the distributions of the control and treatment groups.

Table 4: **Decomposition I:** Impact of Message within Fixed Source

					Republicans					
					ln(Probability Anti-Immigrant + 1)					
			Source Priming				Identified Message			
Message	**Source**	β_s	(S.E.)	%Diff.	*P(Dist.)*	β_{sm}	(S.E.)	%Diff.	*P(Dist.)*	N
Anti	Trump	0.008	(0.006)	2.10%	*0.549*	0.005	(0.006)	1.29%	*0.389*	1634
Anti	Obama	0.007	(0.006)	1.96%	*0.187*	0.008	(0.006)	2.29%	*0.615*	1562
Pro	Trump	0.007	(0.006)	1.99%	*0.549*	-0.033***	(0.006)	-8.87%	*0.000*	1637
Pro	Obama	0.007	(0.006)	1.86%	*0.187*	-0.021***	(0.006)	-5.54%	*0.001*	1581
					Democrats					
					ln(Probability Pro-Immigrant + 1)					
			Source Priming				Identified Message			
Message	**Source**	β_s	(S.E.)	%Diff.	*P(Dist.)*	β_{sm}	(S.E.)	%Diff.	*P(Dist.)*	N
Anti	Trump	-0.002	(0.005)	-0.52%	*0.608*	-0.008	(0.005)	-1.84%	*0.576*	1942
Anti	Obama	-0.002	(0.005)	-0.43%	*0.879*	-0.030***	(0.005)	-7.03%	*0.000*	1903
Pro	Trump	-0.003	(0.005)	-0.61%	*0.608*	-0.001	(0.005)	-0.33%	*0.833*	1977
Pro	Obama	-0.002	(0.005)	-0.43%	*0.879*	0.009*	(0.005)	2.24%	*0.025*	1956

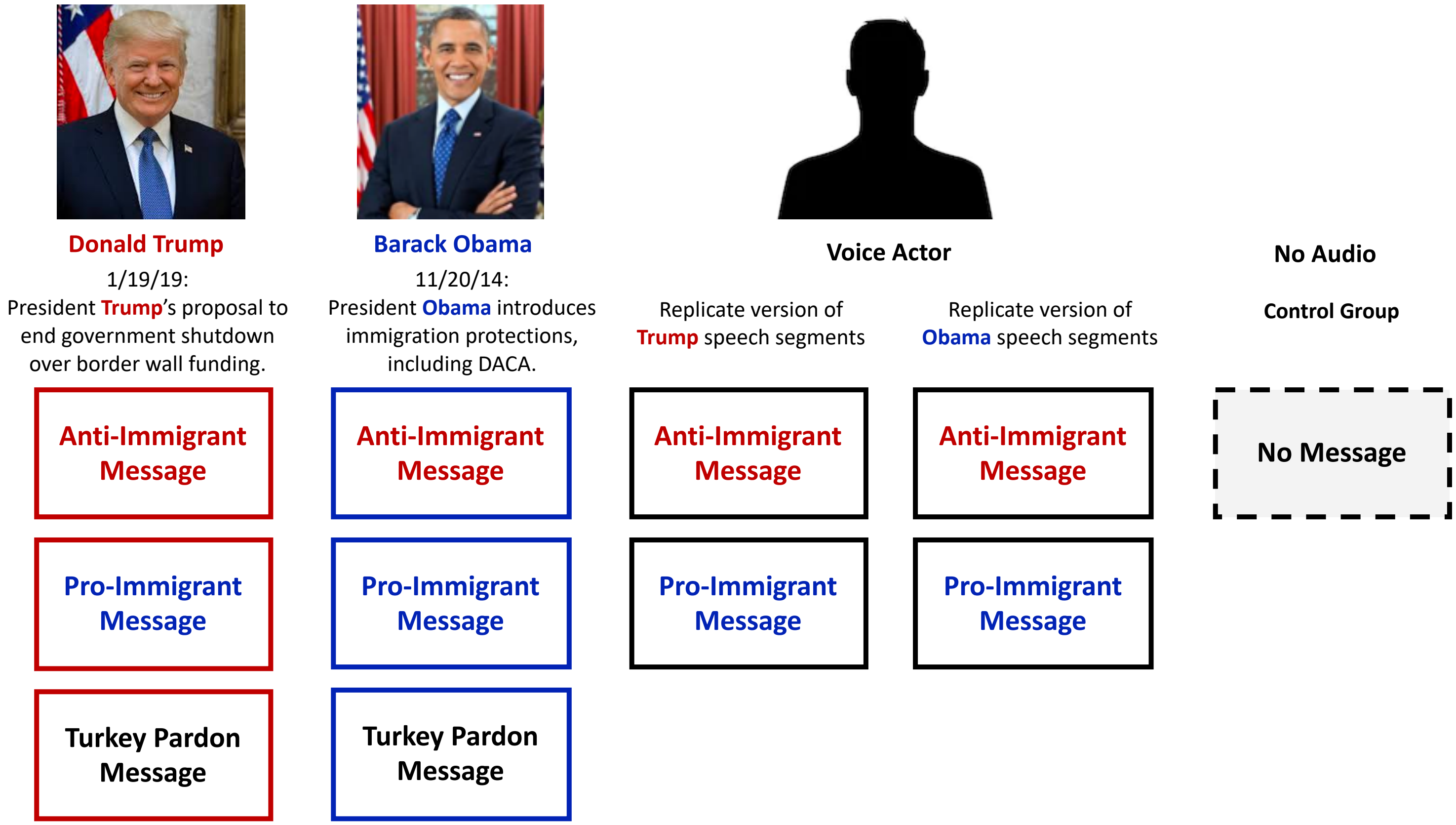
Notes: *p<0.1, **p<0.05, ***p<0.01. This table presents the estimates of the first decomposition of message and source effects of a speech treatment. This decomposition fixes the president source as either Trump or Obama's voice, and varies the message between the turkey pardon and an immigration speech. Each regression includes a president immigration speech, a president turkey pardon speech, and the "no audio" control group. The relevant regression is $ln(P(y)+1) = \beta_0 + \beta_s 1[Source_i = 1] + \beta_{sm} 1[Source_i = 1] \times 1[Message_i = 1] + \gamma X_i + \epsilon$, where the first regressor, $1[Source_i = 1]$, includes both treatment groups with a president's voice, or the turkey pardon and the immigration speech, and the second regressor, $1[Message_i = 1] \times 1[Source_i = 1]$, includes only the president's immigration speech. Outcomes are constructed as $ln(P(y)+1)$, where the outcome is the index probability anti-immigrant (pro-immigrant) for Republicans (Democrats). All regressions additionally include covariates selected by the double lasso procedure to improve precision (See Appendix A4). "% Diff." calculates the implied change in outcome *probability*, or the untransformed probability index, due to treatment, relative to the mean of the comparison group; where the comparison group is the "no audio" group for β_s and the turkey pardon group for β_{sm}. "$P(Dist.)$" is the p-value from a Kolmogorov-Smirnov test of equality of the distributions of the treatment and comparison groups.

Table 5: **Decomposition II:** Impact of Source within Fixed Message

					Republicans					
					ln(Probability Anti-Immigrant + 1)					
		Anonymous Message				Source Persuasion				
Message	**Source**	β_m	(S.E.)	%Diff.	*P(Dist.)*	β_{ms}	(S.E.)	%Diff.	*P(Dist.)*	N
Anti	Trump	0.010*	(0.006)	2.80%	*0.082*	0.003	(0.006)	0.72%	*0.665*	1624
Anti	Obama	0.009	(0.006)	2.48%	*0.144*	0.007	(0.006)	1.81%	*0.567*	1588
Pro	Trump	-0.011*	(0.006)	-2.93%	*0.047*	-0.015**	(0.006)	-4.05%	*0.073*	1646
Pro	Obama	-0.020***	(0.006)	-5.43%	*0.007*	0.007	(0.006)	1.91%	*0.275*	1656
					Democrats					
					ln(Probability Pro-Immigrant + 1)					
		Anonymous Message				Source Persuasion				
Message	**Source**	β_m	(S.E.)	%Diff.	*P(Dist.)*	β_{ms}	(S.E.)	%Diff.	*P(Dist.)*	N
Anti	Trump	-0.017***	(0.005)	-3.96%	*0.015*	0.006	(0.005)	1.41%	*0.363*	1880
Anti	Obama	-0.018***	(0.005)	-4.21%	*0.005*	-0.014***	(0.005)	-3.41%	*0.001*	1877
Pro	Trump	0.003	(0.005)	0.65%	*0.187*	-0.007	(0.005)	-1.59%	*0.329*	1935
Pro	Obama	0.011**	(0.005)	2.67%	*0.097*	-0.004	(0.005)	-0.90%	*0.141*	1988

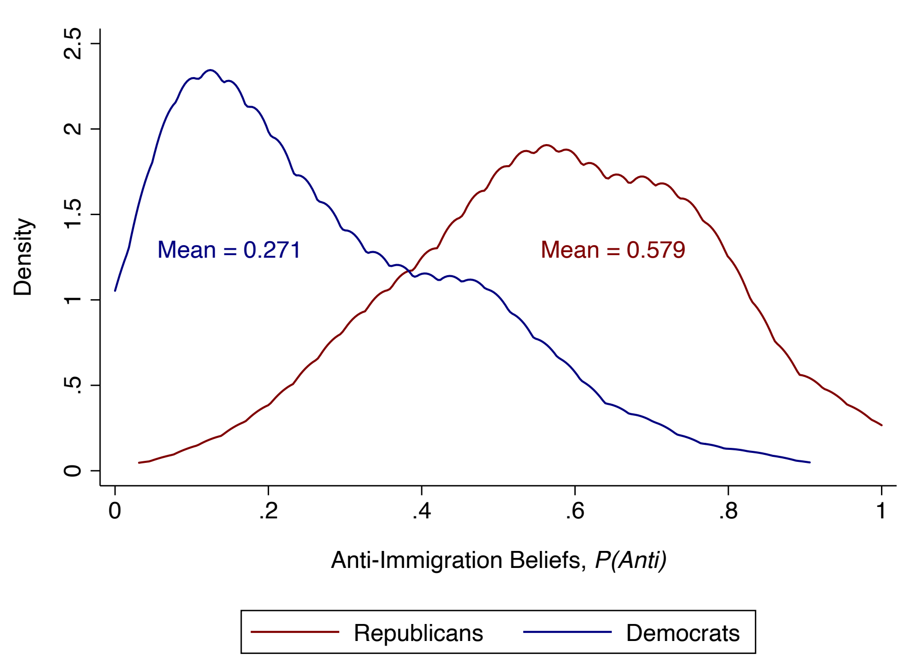
Notes: *p<0.1, **p<0.05, ***p<0.01. This table presents the estimates of the second decomposition of message and source effects of a speech treatment. This decomposition fixes the message of a speech, and varies the source between the actor and the president. Each regression includes a president immigration speech, the voice actor version of the speech, and the "no audio" control group. The relevant regression is $ln(P(y)+1) = \beta_0 + \beta_m 1[Message_i = 1] + \beta_{ms} 1[Message_i = 1] \times 1[Source_i = 1] + \gamma X_i + \epsilon$, where the first regressor, $1[Message_i = 1]$, includes both treatment groups with a particular message, or the actor and president speech versions, and the second regressor, $1[Message_i = 1] \times 1[Source_i = 1]$, includes only the president's version of the immigration speech. Outcomes are constructed as $ln(P(y)+1)$, where the outcome is the index probability anti-immigrant (pro-immigrant) for Republicans (Democrats). All regressions additionally include covariates selected by the double lasso procedure to improve precision (See Appendix A4). "% Diff." calculates the implied change in outcome *probability*, or the untransformed probability index, due to treatment, relative to the mean of the comparison group; where the comparison group is the "no audio" group for β_m and the actor speech group for β_{ms}. "$P(Dist.)$" is the p-value from a Kolmogorov-Smirnov test of equality of the distributions of the treatment and comparison groups.

Figure 1: Experiment Design



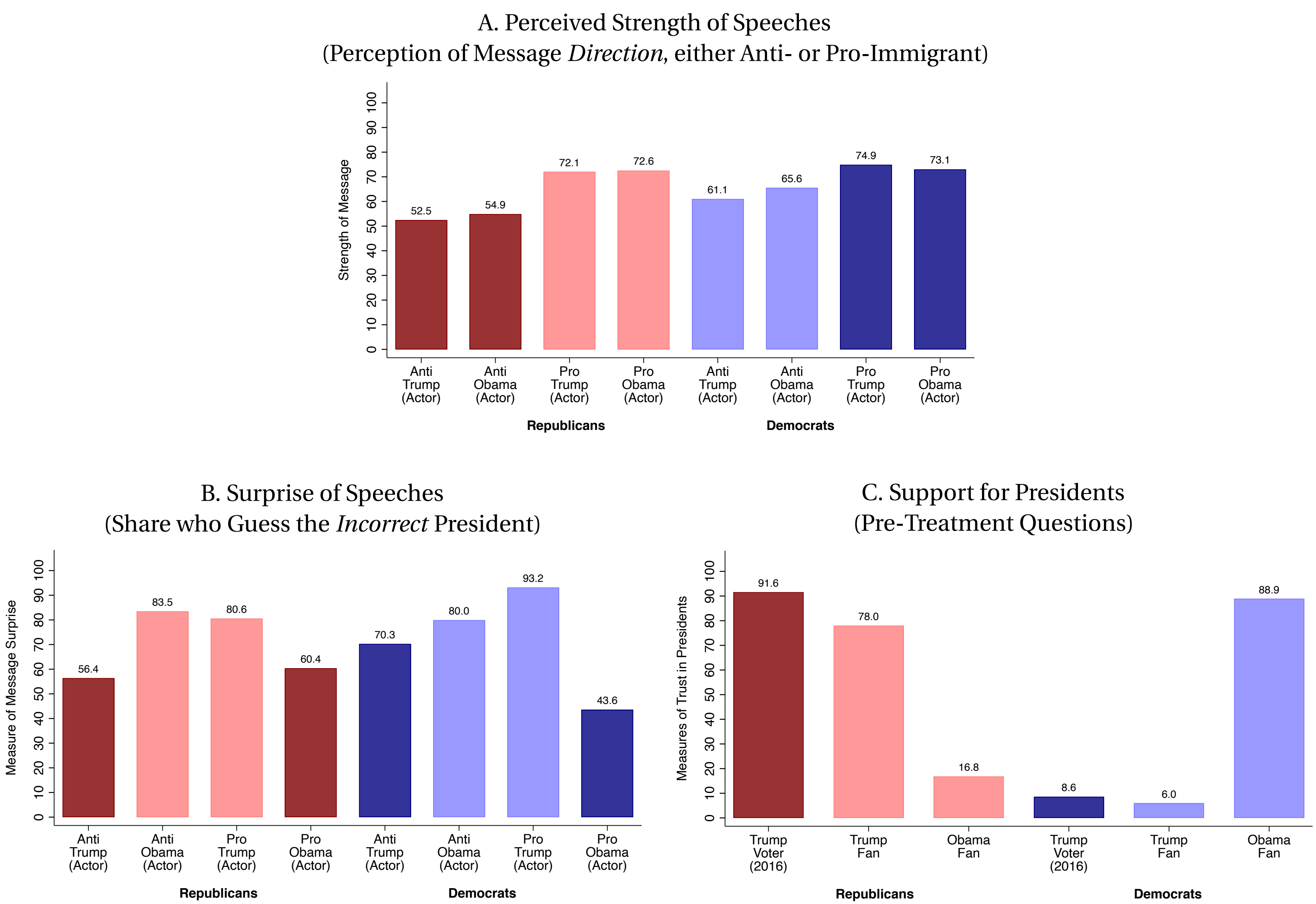
Notes: This figure depicts the treatment arms of the experiment. The sample is stratified by party, Republican or Democrat, and then within party all participants are randomized into the 11 treatment arms shown above.

Figure 2: Anti-Immigration Beliefs in No Audio Control Group



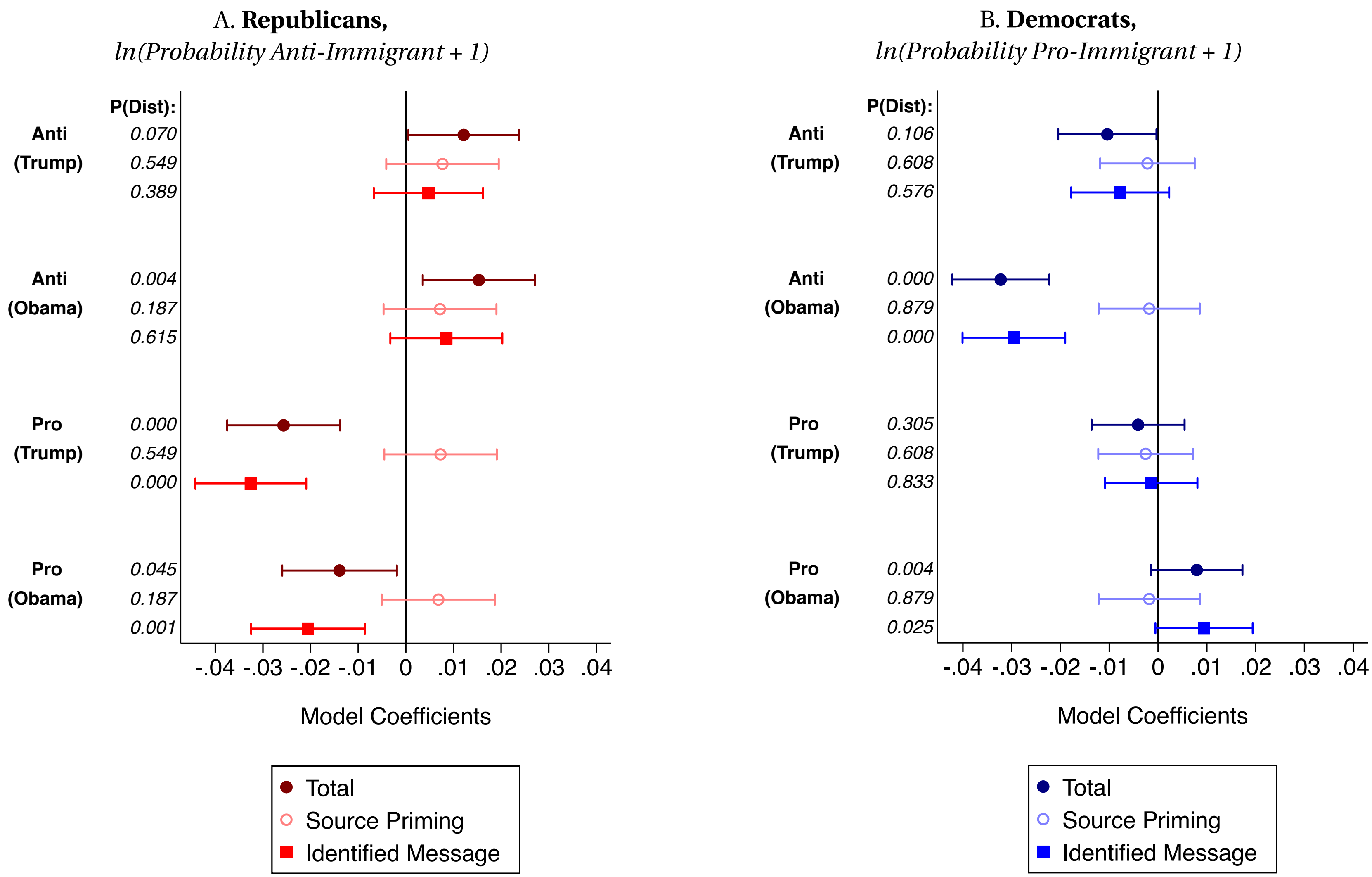
Notes: This plot shows the baseline distribution of the Anti-Immigration index or $P(Anti - Immigration)$ in the no audio control group. This index is composed of 16 questions about immigration beliefs asked in the second part of the survey, described in Table 1.

Figure 3: Measures of Message Similarity, Surprise of Messages, and Source Favorability



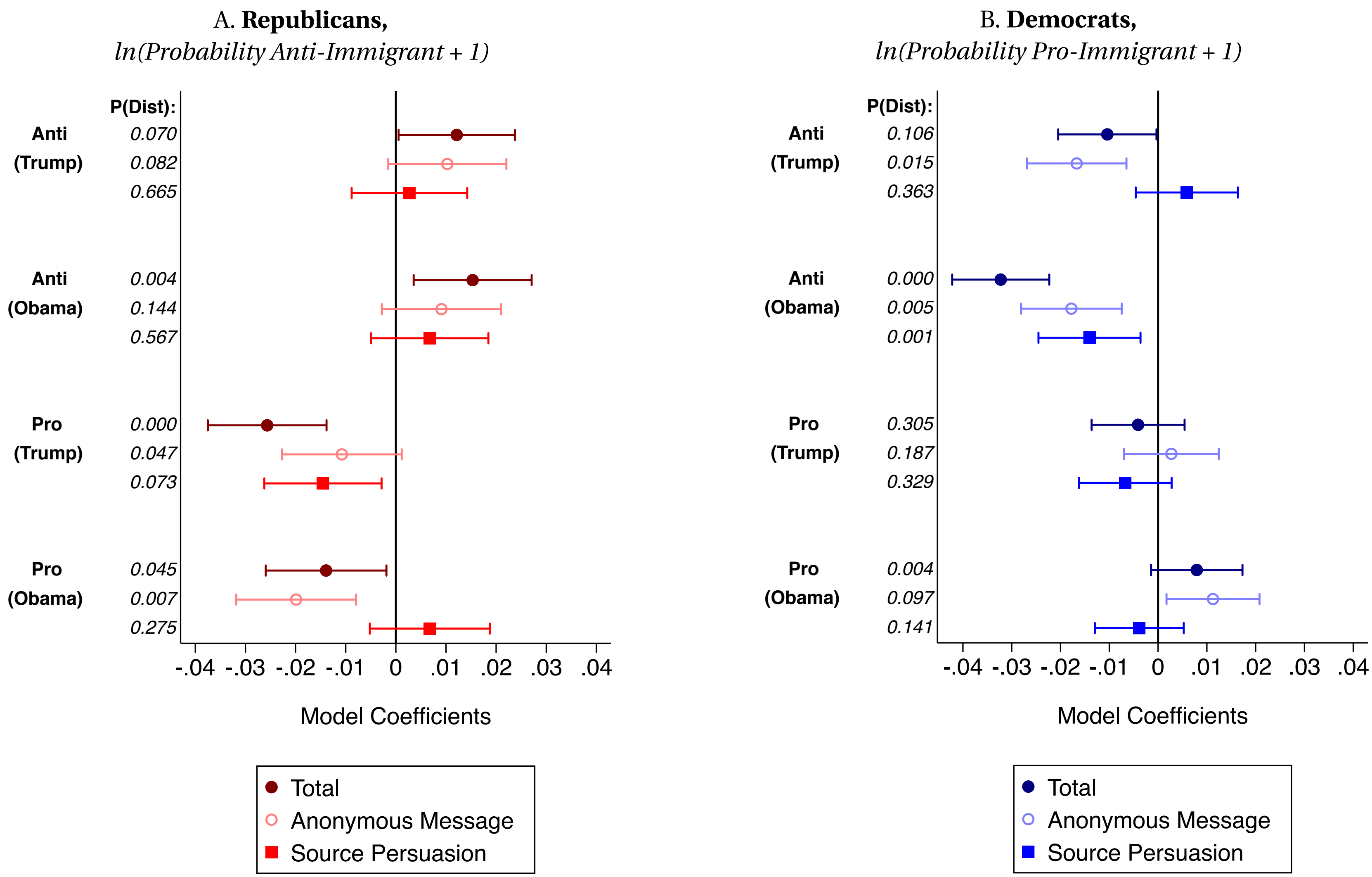
Notes: Panel A compiles participant responses to a question that asks participants how pro- or anti- immigrant they think the speech was after treatment. Specifically, Panel A plots the degree pro-immigrant for pro-immigrant speeches and the degree anti-immigrant for anti-immigrant speeches. The actor speeches shown in this figure do not reveal a stated source and can be compared across presidents to measure the "strength" of different speech messages. Panel B contains responses from a multiple choice question asked post-intervention about which president the participant thought was the source of the speech, from a list of the four most recent presidents (and an "Other" option). This plot shows the share of participants who guessed the *incorrect* president for each actor version of the treatments. This measure is the share of participants who would be surprised by the source of the speech. Panel C includes pre-treatment responses from questions asking whether participants voted for Trump in 2016 and whether participants are a fan of Trump and Obama.

Figure 4: **Decomposition I:** Impact of Message within Fixed Source



Notes: This figure presents the coefficient estimates and 95% confidence intervals of the first decomposition of message and source effects of a speech treatment (analogous to Table 4). This decomposition fixes the president source as either Trump or Obama's voice, and varies the message between the turkey pardon and an immigration speech. The total estimate corresponds to estimates from Table 3, and the Source Priming and Identified Message effects correspond to β_s and β_{sm} from Table 4, respectively. Outcomes are constructed as $ln(P(y)+1)$, where the outcome is the index probability anti-immigrant (pro-immigrant) for Republicans (Democrats). All regressions additionally include covariates selected by the double lasso procedure to improve precision (See Appendix A4). "% Diff." calculates the implied change in outcome *probability*, or the untransformed probability index, due to treatment, relative to the mean of the comparison group; where the comparison group is the "no audio" group for β_s and the turkey pardon group for β_{sm}. "$P(Dist.)$" is the p-value from a Kolmogorov-Smirnov test of equality of the distributions of the treatment and comparison groups.

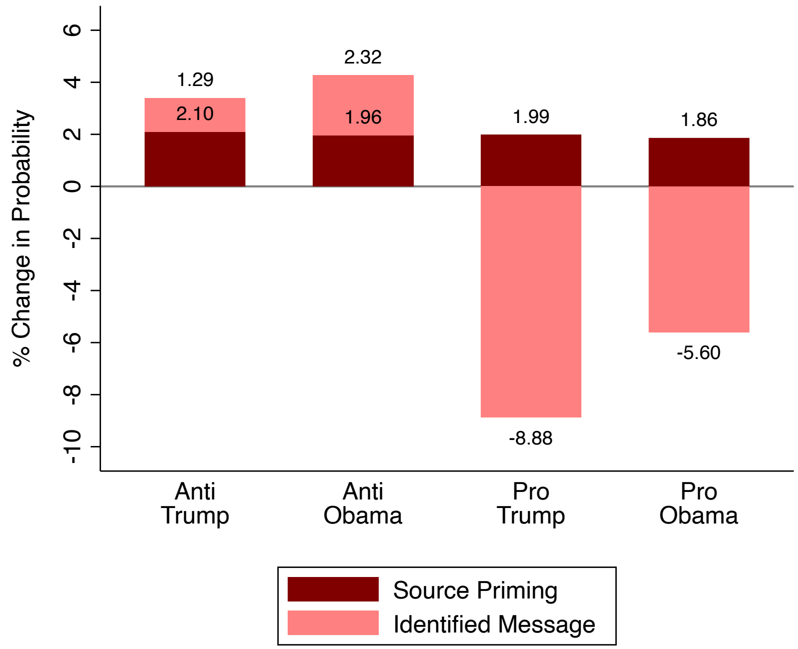
Figure 5: **Decomposition II:** Impact of Source within Fixed Message



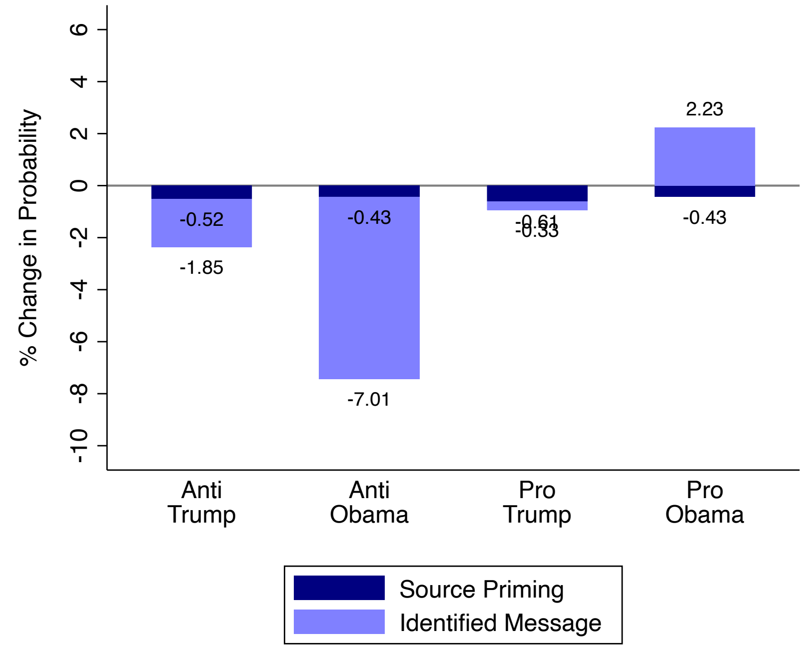
Notes: This figure presents the coefficient estimates and 95% confidence intervals of the first decomposition of message and source effects of a speech treatment (analogous to Table 5). This decomposition fixes the speech message, and varies the speaker source as either the actor or the president. The total estimate corresponds to estimates from Table 3, and the Anonymous Message and Source Persuasion effects correspond to β_m and β_{ms} from Table 5, respectively. Outcomes are constructed as $ln(P(y)+1)$, where the outcome is the index probability anti-immigrant (pro-immigrant) for Republicans (Democrats). All regressions additionally include covariates selected by the double lasso procedure to improve precision (See Appendix A4). "% Diff." calculates the implied change in outcome *probability*, or the untransformed probability index, due to treatment, relative to the mean of the comparison group; where the comparison group is the "no audio" group for β_m and the corresponding actor speech group for β_{ms}. "$P(Dist.)$" is the p-value from a Kolmogorov-Smirnov test of equality of the distributions of the treatment and comparison groups.

Figure 6: **Decomposition I:** Impact of Message within Fixed Source
(%Change in Probability)

A. **Republicans,**
Probability Anti-Immigrant

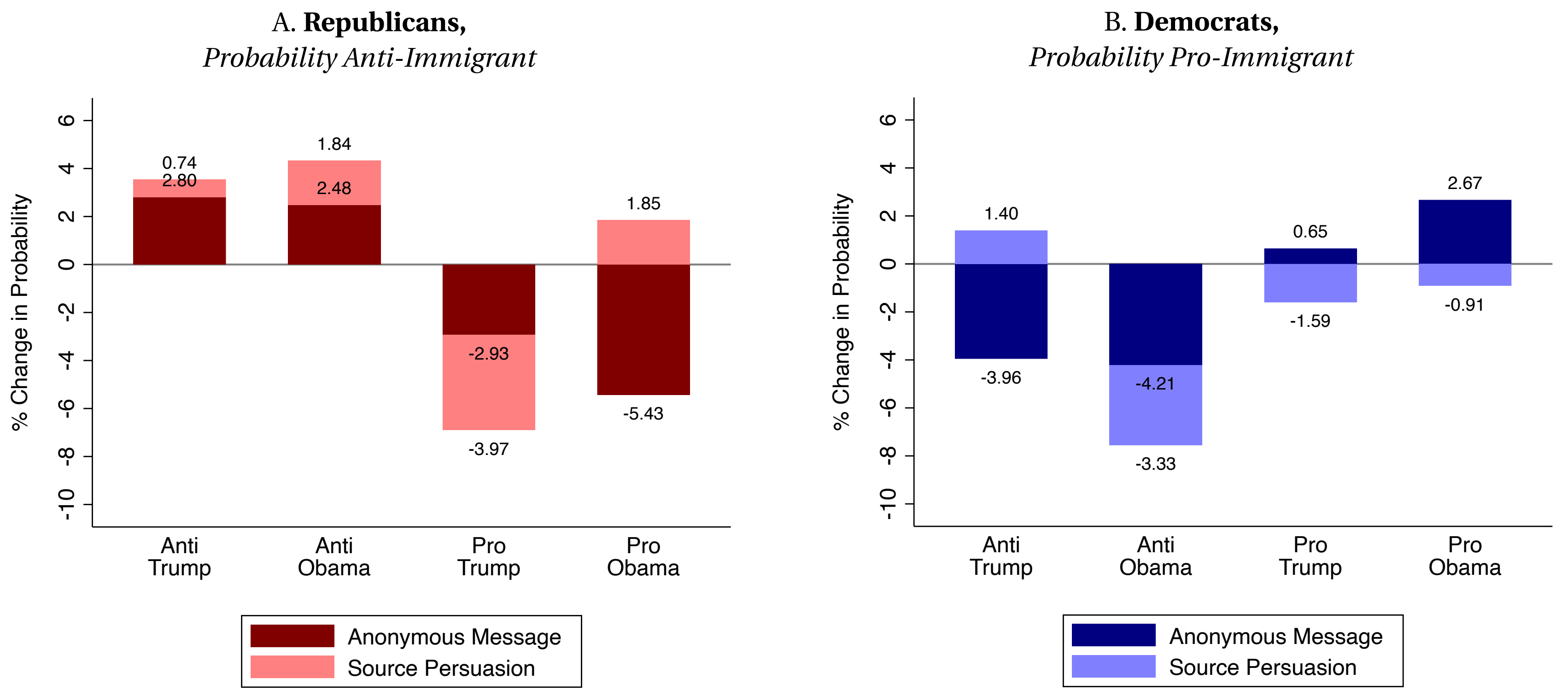


B. **Democrats,**
Probability Pro-Immigrant



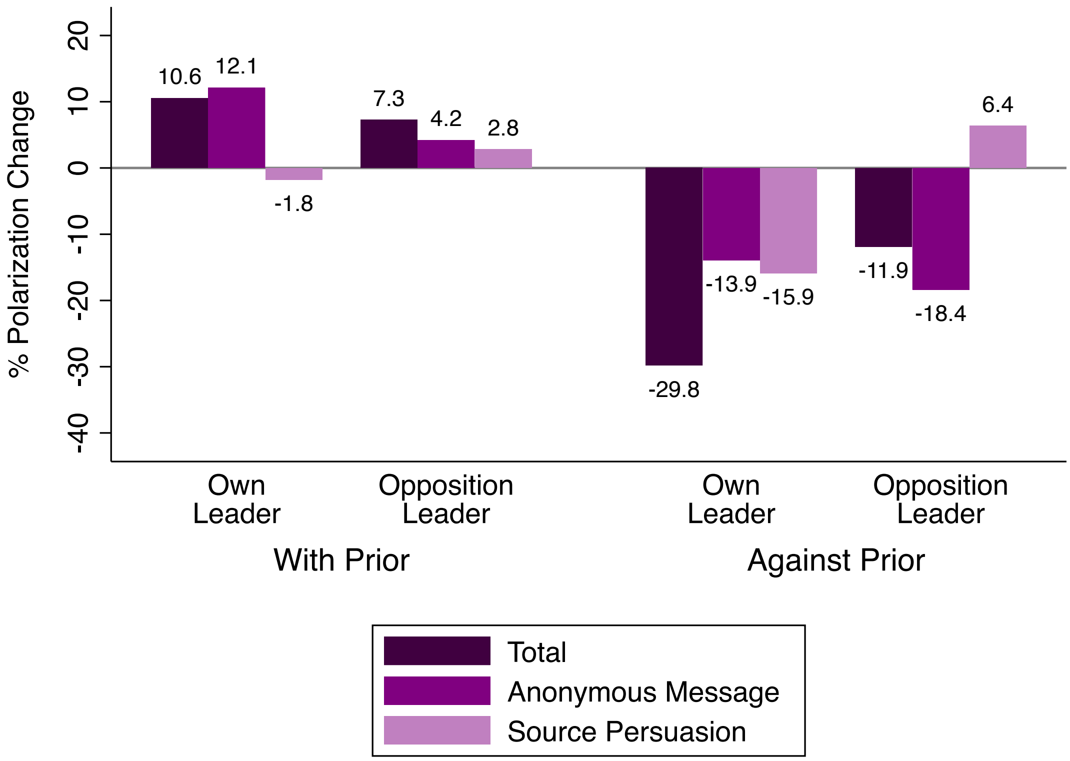
Notes: This figure presents the implied percent changes in probability anti-immigrant or pro-immigrant, given the first decomposition, or the estimates from Table 4. This decomposition fixes the president source as either Trump or Obama's voice, and varies the message between the turkey pardon and an immigration speech. The Source Priming and Identified Message effects correspond to β_s and β_{sm} from Table 4, respectively. Adding the two effects corresponds to the total change in probability for a particular president speech.

Figure 7: **Decomposition II:** Impact of Source within Fixed Message
(%Change in Probability)



Notes: This figure presents the implied percent changes in probability anti-immigrant or pro-immigrant, given the first decomposition, or the estimates from Table 5. This decomposition fixes the speech message and varies the speech source as the voice actor or the president's voice. The Anonymous Message and Source Persuasion effects correspond to β_m and β_{ms} from Table 5, respectively. Adding the two effects corresponds to the total change in probability for a particular president speech.

Figure 8: **Decomposition II:** Polarization Change in Counterfactual Scenarios (%Change in Party Difference in *Probability Anti-Immigrant*)



Notes: This plot corresponds to the coefficient estimates in Table A1, and displays implied changes in differences in *Probability Anti-Immigrant*. The baseline party Difference in $P(Anti) = 0.308$. The estimates represent percent changes in the distance in the anti-immigrant index probability between Republicans and Democrats in the treatment versus control group. The treatments for each party vary according to constructed counterfactuals. "Own Leaders" are treatments that are Trump for Republicans and Obama for Democrats, while "Opposition Leaders" are the converse. "With Prior" messages are anti-immigrant speeches for Republicans and pro-immigrant speeches for Democrats, and "Against Prior" messages are the converse. Likewise, actor control groups correspond to replicate messages for a given treatment. Specifically, the corresponding regression is $ln(P(Anti)+1) = \beta_0 + \beta_{m1}1[Message_i = 1] + \beta_{m2}1[Message_i = 1] \times 1[Republican_i = 1] + \beta_{ms1}1[Message_i = 1] \times 1[Source_i = 1] + \beta_{ms2}1[Message_i = 1] \times 1[Source_i = 1] \times 1[Republican_i = 1] + \gamma X + \epsilon$, where the interaction coefficient β_{m2} represents the change in polarization from the Anonymous Message effect and β_{ms2} represents the change in polarization from the Source Persuasion effect. The covariates X correspond to those in Appendix A4 for both party groups. A positive estimate is an increase in polarization.

ONLINE APPENDIX

A1 Opening Paragraphs of Immigration Speech Treatments

Trump Anti-Immigrant Speech "We believe in a safe and lawful system of immigration, one that upholds our laws, our traditions and our most cherished values. Unfortunately, our immigration system has been badly broken for a very long time. Over the decades, many presidents and many lawmakers have come and gone, and no real progress has been made on immigration. We are now living with the consequences – and they are tragic – brought about by decades of political stalemate, partisan gridlock, and national neglect. Illegal immigration reduces wages and strains public services. The lack of border control provides a gateway, and a very wide and open gateway, for criminals and gang members to enter the United States. I want this to end; it's got to end now. These are not talking points. These are the heartbreaking realities that are hurting innocent, precious human beings every single day on both sides of the border."

Obama Anti-Immigrant Speech "Today, our immigration system is broken, and everybody knows it. Families who enter our country the right way and play by the rules watch others flout the rules. Business owners who offer their workers good wages and benefits see the competition exploit undocumented immigrants by paying them far less. All of us take offense to anyone who reaps the rewards of living in America without taking on the responsibilities of living in America. Millions of us, myself included, go back generations in this country, with ancestors who put in the painstaking work to become citizens. So we don't like the notion that anyone might get a free pass to American citizenship. I know that some worry immigration will change the very fabric of who we are, or take our jobs, or stick it to middle-class families at a time when they already feel like they've gotten the raw end of the deal for over a decade. I hear these concerns."

Trump Pro-Immigrant Speech "Just a short time ago, I had the honor of presiding over the swearing in of five new great American citizens. It was a beautiful ceremony and a moving reminder of our nation's proud history of welcoming immigrants from all over the world into our national family. I told them that the beauty and majesty of citizenship is that it draws no distinctions of race or class or faith or gender or background. All Americans, whether first generation or tenth generation, are bound together in love and loyalty, friendship and affection. We're all equal. We are one team and one people proudly saluting one great American flag. Unfortunately, our immigration system has been badly broken for a very long time. Over the decades, many presidents and many lawmakers have come and gone, and no real progress has been made on immigration. The good news is these problems can all be solved, but only if we have the political courage to do what is just and what is right. Both sides in Washington must simply come together, listen to each other, put down their armor, build trust, reach across the aisle and find solutions."

Obama Pro-Immigrant Speech "For more than 200 years, our tradition of welcoming immigrants from around the world has given us a tremendous advantage over other nations. It's kept us youthful, dynamic, and entrepreneurial. It has shaped our character as a people with limitless possibilities – people not trapped by our past, but able to remake ourselves as we choose. But today, our immigration system is broken, and everybody knows it. It's been this way for decades. And for decades, we haven't done much about it. When I took office, I committed to fixing this broken immigration system. We need more than politics as usual when it comes to immigration; we need reasoned, thoughtful, compassionate debate that focuses on our hopes, not our fears."

Full speech transcripts are viewable in Supplemental Materials S2.

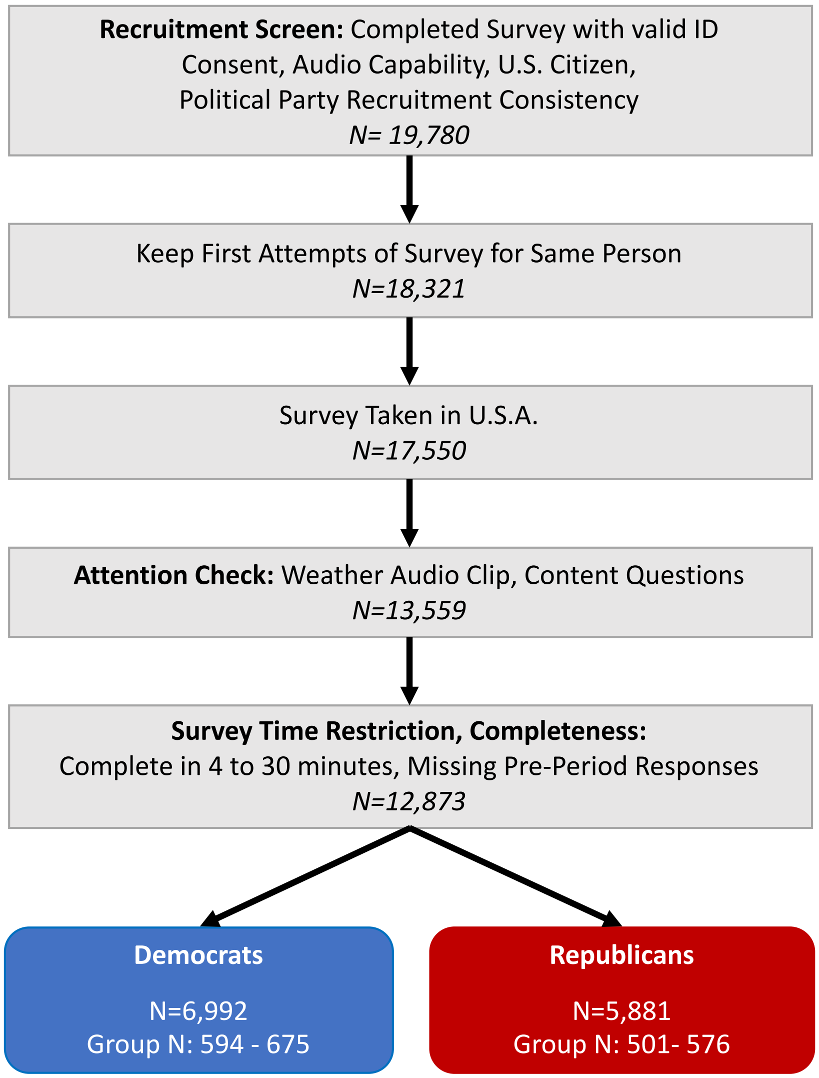
A2 Appendix Tables & Figures

Table A1: **Decomposition II:** Polarization Change in Counterfactual Scenarios

		Polarization Change *ln(Probability Anti)* *Change Diff.: Republican-Democrat*			
Leader	**Effect**	β	(S.E.)	%Diff.	N
Message With Prior					
Own Leader	Total	0.022**	(0.009)	10.55%	2439
Own Leader	Anonymous Message	0.026***	(0.009)	12.12%	
Own Leader	Source Persuasion	-0.005	(0.009)	-1.76%	3612
Opposing Leader	Total	0.013	(0.009)	7.28%	2363
Opposing Leader	Anonymous Message	0.009	(0.009)	4.21%	
Opposing Leader	Source Persuasion	0.004	(0.009)	2.82%	3523
Message Against Prior					
Own Leader	Total	-0.067***	(0.009)	-29.82%	2389
Own Leader	Anonymous Message	-0.032***	(0.009)	-13.90%	
Own Leader	Source Persuasion	-0.035***	(0.009)	-15.90%	3523
Opposing Leader	Total	-0.025***	(0.009)	-11.90%	2347
Opposing Leader	Anonymous Message	-0.040***	(0.009)	-18.36%	
Opposing Leader	Source Persuasion	0.014	(0.009)	6.38%	3536

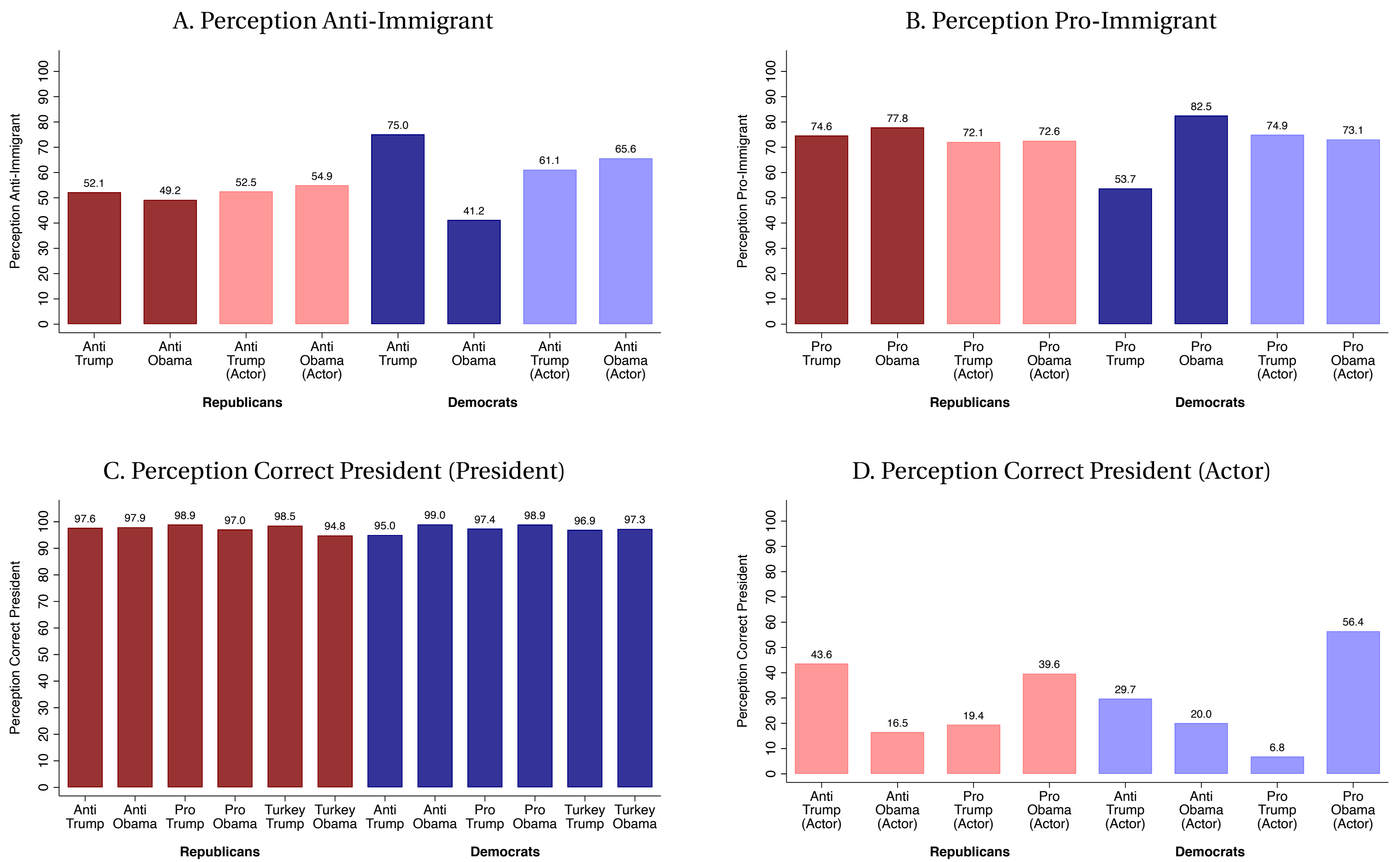
Notes: *p<0.1, **p<0.05, ***p<0.01. This table corresponds to the probability changes in Figure 8, and displays coefficient estimates for differences in $ln(ProbabilityAnti-Immigrant+1)$. The baseline party Difference in $P(Anti) = 0.308$. The estimates represent percent changes in the distance in the anti-immigrant index probability between Republicans and Democrats in the treatment versus control group. The treatments for each party vary according to constructed counterfactuals. "Own Leaders" are treatments that are Trump for Republicans and Obama for Democrats, while "Opposition Leaders" are the converse. "With Prior" messages are anti-immigrant speeches for Republicans and pro-immigrant speeches for Democrats, and "Against Prior" messages are the converse. Likewise, actor control groups correspond to replicate messages for a given treatment. Specifically, the corresponding regression is $ln(P(Anti)+1) = \beta_0 + \beta_{m1} 1[Message_i = 1] + \beta_{m2} 1[Message_i = 1] \times 1[Republican_i = 1] + \beta_{ms1} 1[Message_i = 1] \times 1[Source_i = 1] + \beta_{ms2} 1[Message_i = 1] \times 1[Source_i = 1] \times 1[Republican_i = 1] + \gamma X + \epsilon$, where the interaction coefficient β_{m2} represents the change in polarization from the Anonymous Message effect and β_{ms2} represents the change in polarization from the Source Persuasion effect. The covariates X correspond to those in Appendix A4 for both party groups. A positive estimate is an increase in polarization. All regressions additionally include covariates selected by the double lasso procedure to improve precision for both party groups (See Appendix A4)."% Diff." calculates the implied change in the difference of outcome *probability* across party, or the untransformed probability index, due to treatment, relative to the baseline difference across party in the control group.

Figure A1: Sample Restrictions



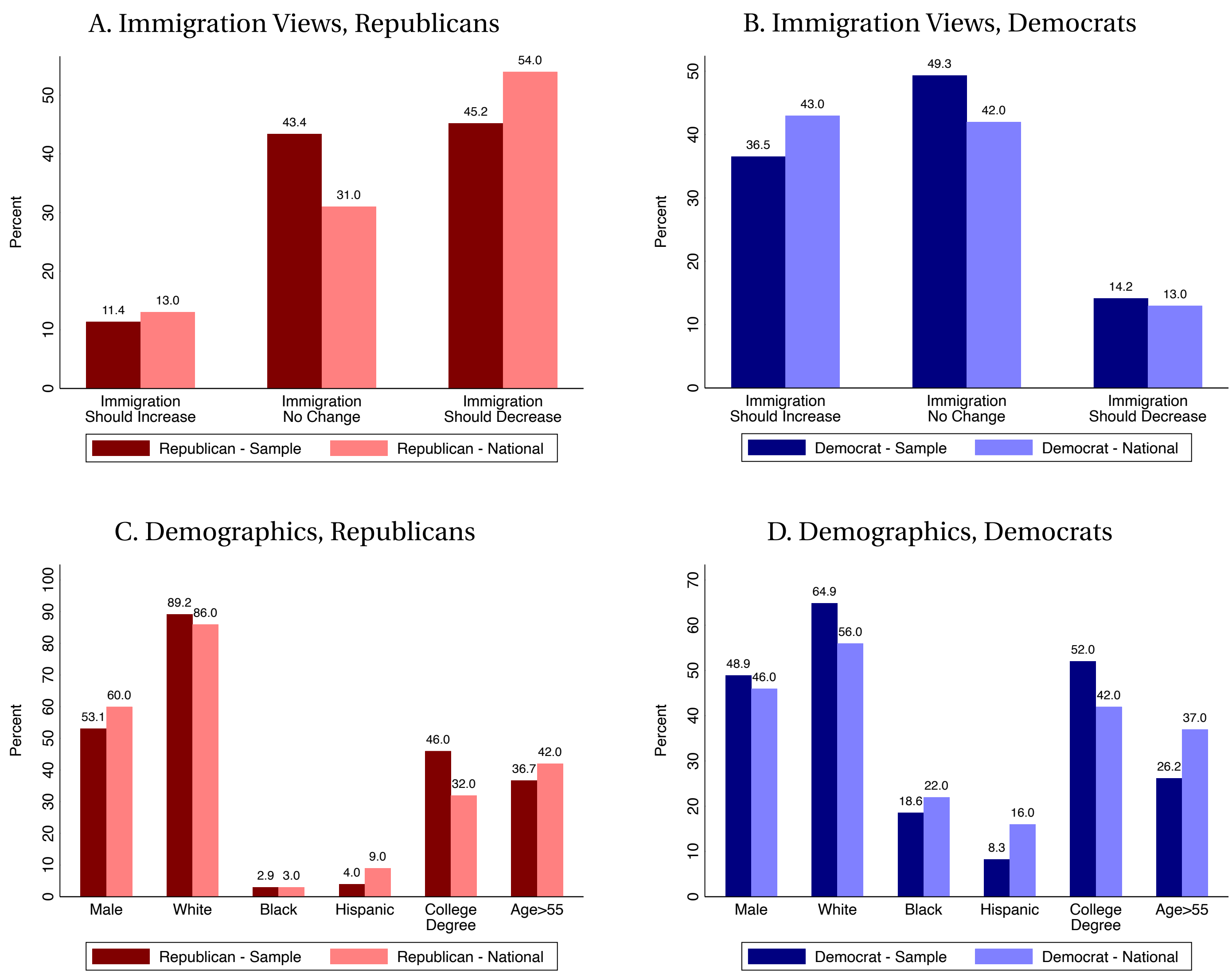
Notes: This figure displays the sample restrictions used to construct the study data set. The first restriction identifies individuals who took the survey more than once using IP addresses and retains only the first attempt. Next, we geocode the survey taker location and remove respondents who took the survey outside of the U.S. The third restriction cuts people who did not consent to the study, stated that their device did not have audio capability, were not U.S. citizens, or were recruited to the survey as a member of particular political party and then stated that their party affiliation was the opposite party within the survey. Next, we include two sets of attention checks to screen individuals for quality. First, we have all participants listen to an audio clip of a weather forecast and answer comprehension related questions. Second, for participants who listen to a treatment audio clip (immigration or turkey pardon speech) we also ask comprehension related questions. Lastly, we exclude individuals who take the survey exceptionally fast or slow, restricting the time in survey to 6 to 25 minutes.

Figure A2: Perception Treatment Content and Source



Notes: Panels A and B compile participant responses to a question that asks participants how pro- or anti- immigrant they think the speech was after treatment. Specifically, Panels A and B plot the degree pro-immigrant for pro-immigrant speeches and the degree anti-immigrant for anti-immigrant speeches. The actor speeches shown in this figure do not reveal a stated source and can be compared across presidents to measure the "strength" of different speech messages. Panels C and D plot responses from a multiple choice question asked post-intervention about which president the participant thought was the source of the speech, from a list of the four most recent presidents (and an "Other" option). This plot shows the share of participants who guessed the *correct* president for each of the treatments. Panel C shows the results for the revealed source president versions of speeches and Panel D shows the results for the actor versions of the treatments, without the revealed president source. This is the inverse of the share of participants who would be surprised by the source (share that guess the *incorrect* president) in Panel B of Figure 3.

Figure A3: Sample Comparison to Party Demographics

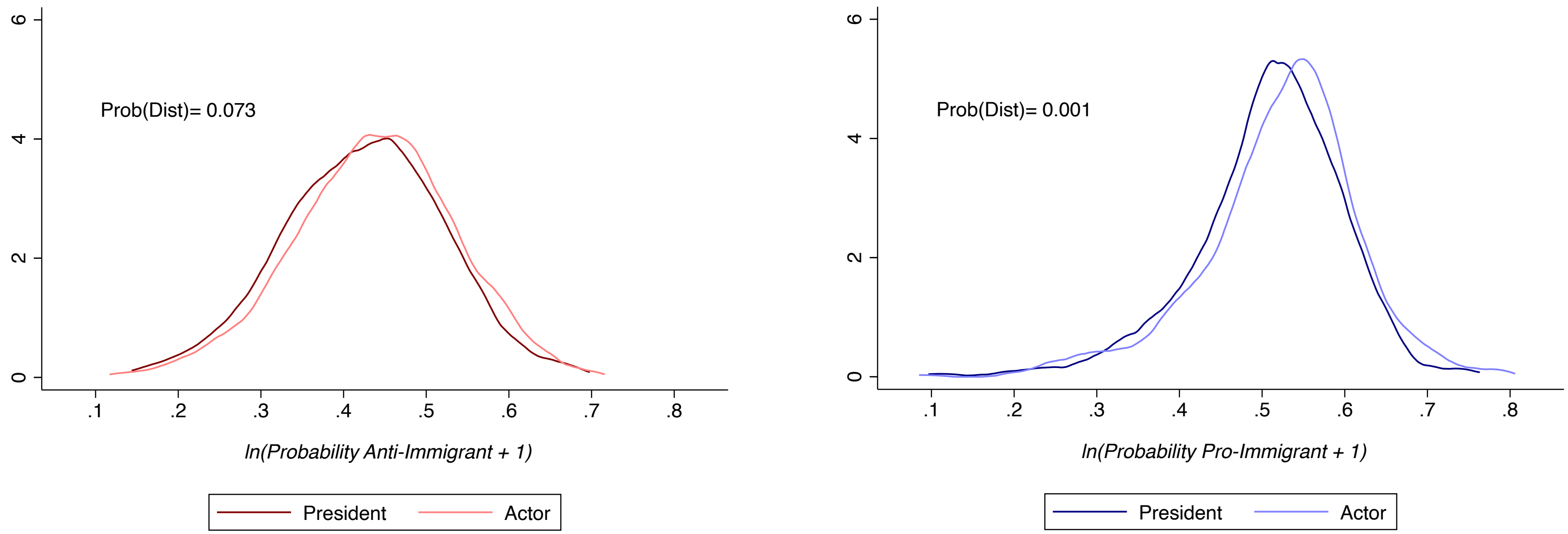


Notes: This set of plots compares characteristics of the study sample to national data for political parties. The top panel compares pre-treatment survey responses to a Gallup Survey question about whether immigration should be increased, decreased or stay the same (Jones, 2019). The national comparison is compiled from survey evidence and data on eligible voters from the U.S. Census, and Pew Research (File, 2018; Igielnik and Budiman, 2020; Doherty et al., 2020).

Figure A4: **Decomposition II:** Source Persuasion Effect
Distribution of Outcome by Treatment Group (Corresponds to β_{ms})

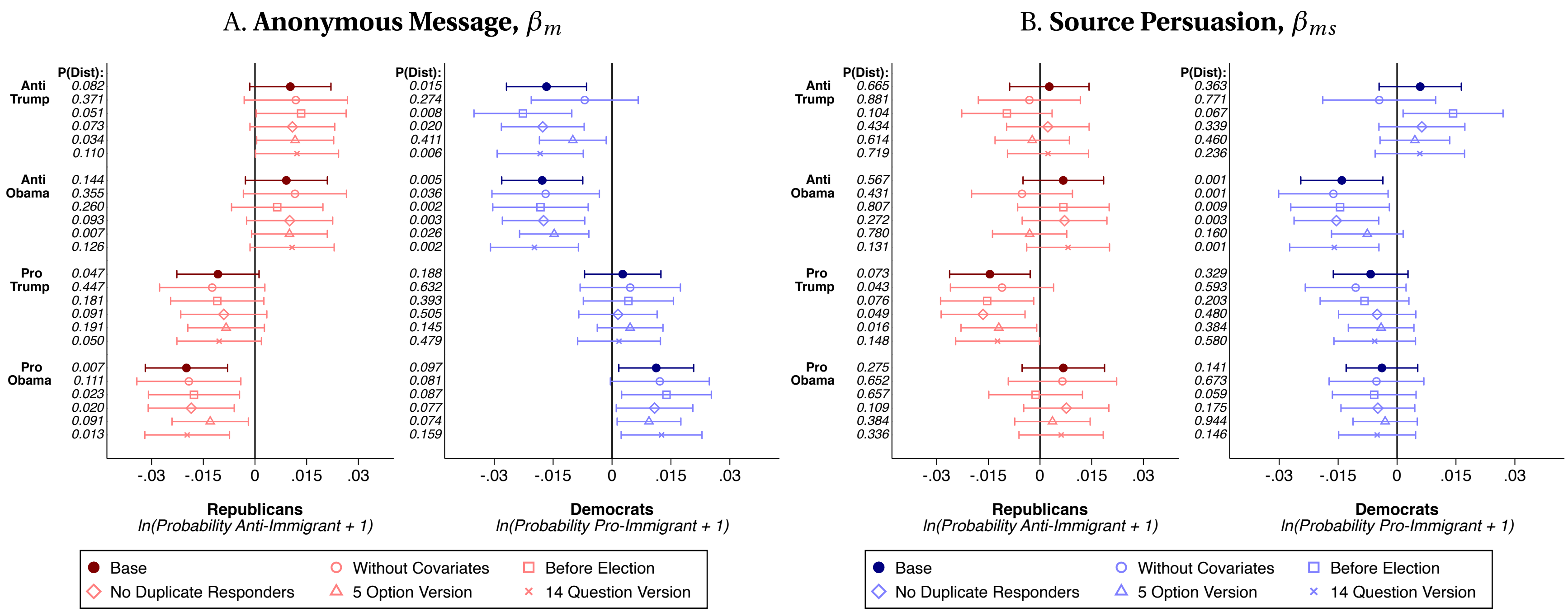
A. **Republicans,**
Pro-immigrant Message, Trump
ln(Probability Anti-Immigrant + 1)

B. **Democrats,**
Anti-immigrant Message, Obama
ln(Probability Anti-Immigrant + 1)



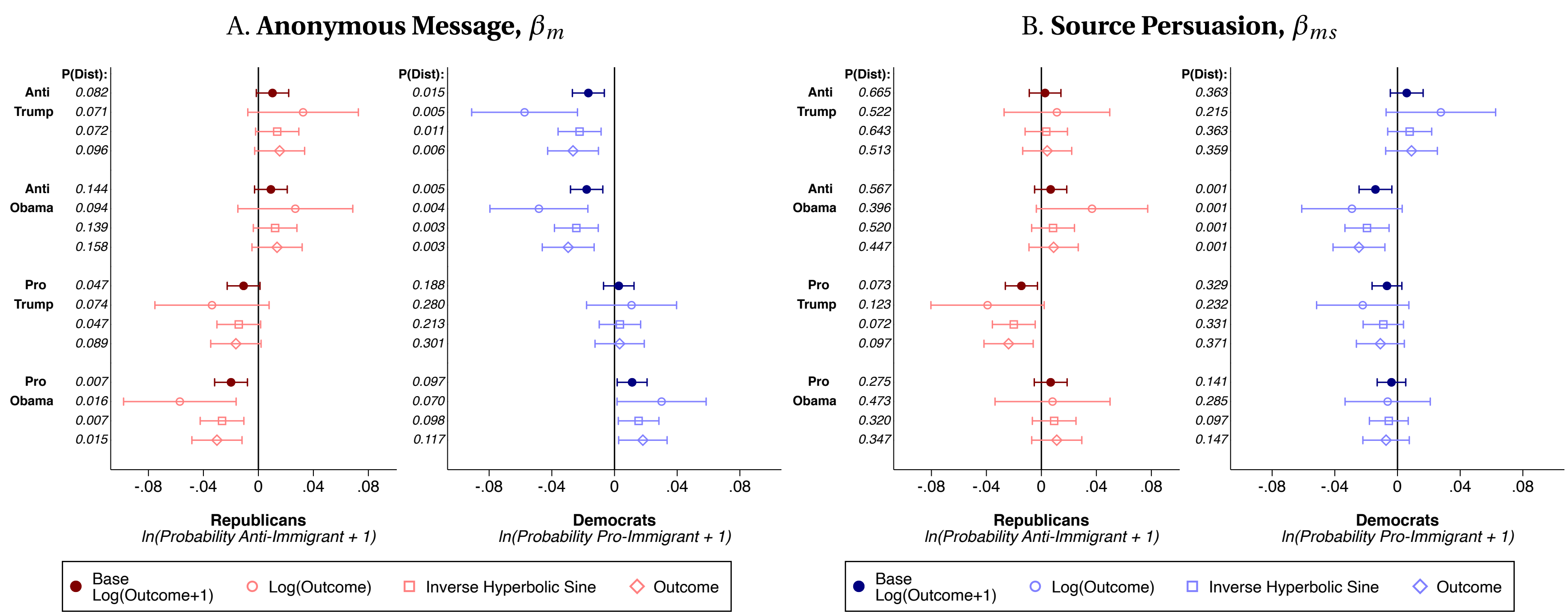
Notes: This figure plots the distribution of the outcomes for the two president treatments with Source Persuasion effects, the Pro-Immigrant Trump message for Republicans and the Anti-Immigrant Obama message for Democrats, relative to the replicate versions recorded by the actor. Outcomes are measured as $ln(y+1)$, and are first residualized by the set of control variables from the double lasso procedure described in Appendix A4 before being plotted. "Prob(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and control groups for each estimate.

Figure A5: **Decomposition II:** Robustness Specifications



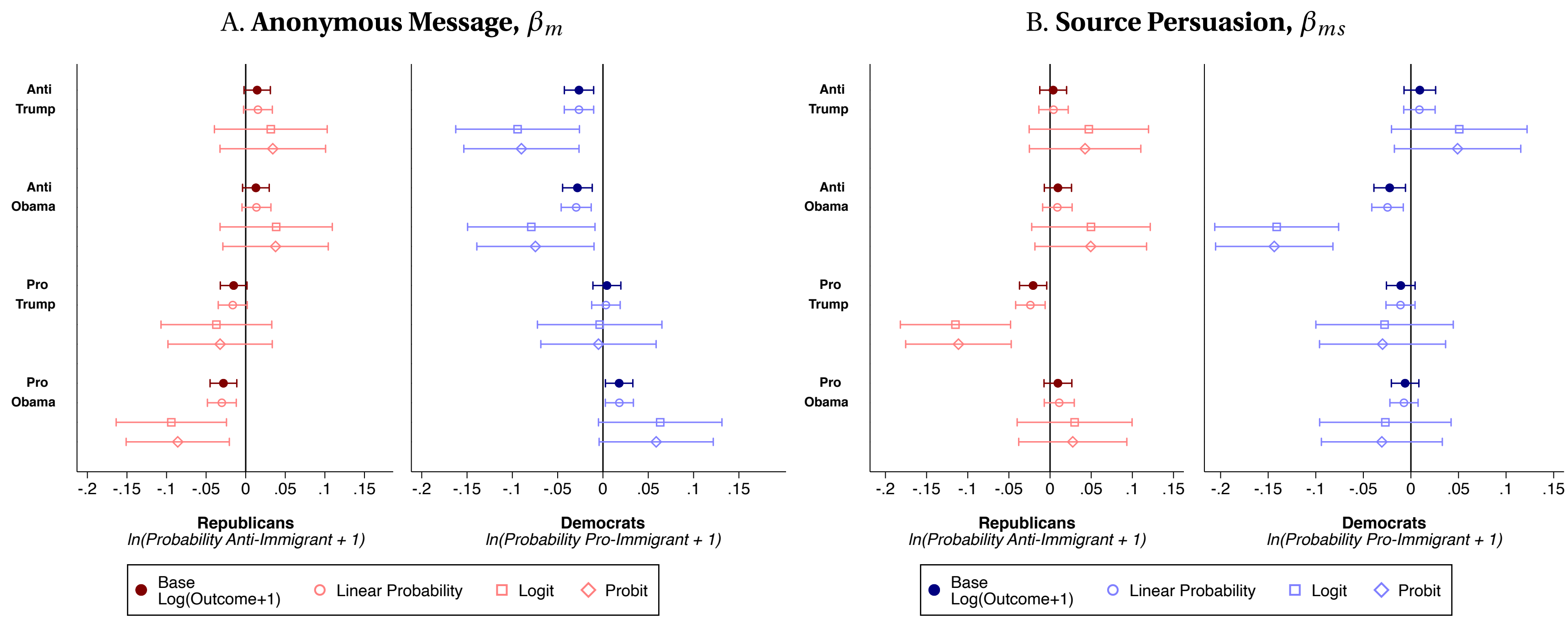
Notes: These figures replicate the baseline specification for Decomposition II for alternative sample restrictions and specifications. All outcomes are measured using $ln(y+1)$. "Without Covariates" is a specification that drops the covariate controls described in Appendix A4. "No Duplicate Responders" excludes any individuals who attempted the survey more than once; the baseline sample includes first attempts of the survey for these individuals. "5 Option Version" constructs the index using the full set of 5 options for questions that have 5 options (e.g. Strongly Agree, Agree, Neutral, Disagree, Strongly Agree), rather than the baseline version that collapses all answers to have 3 options to match 3 option questions (e.g. Agree, Neutral, Disagree). "14 Question Version" constructs the index to match the pre-registered version, which excludes questions on the societal contribution of legal immigrants or immigrants from English-speaking countries. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

Figure A6: **Decomposition II:** Robustness of Log Specification



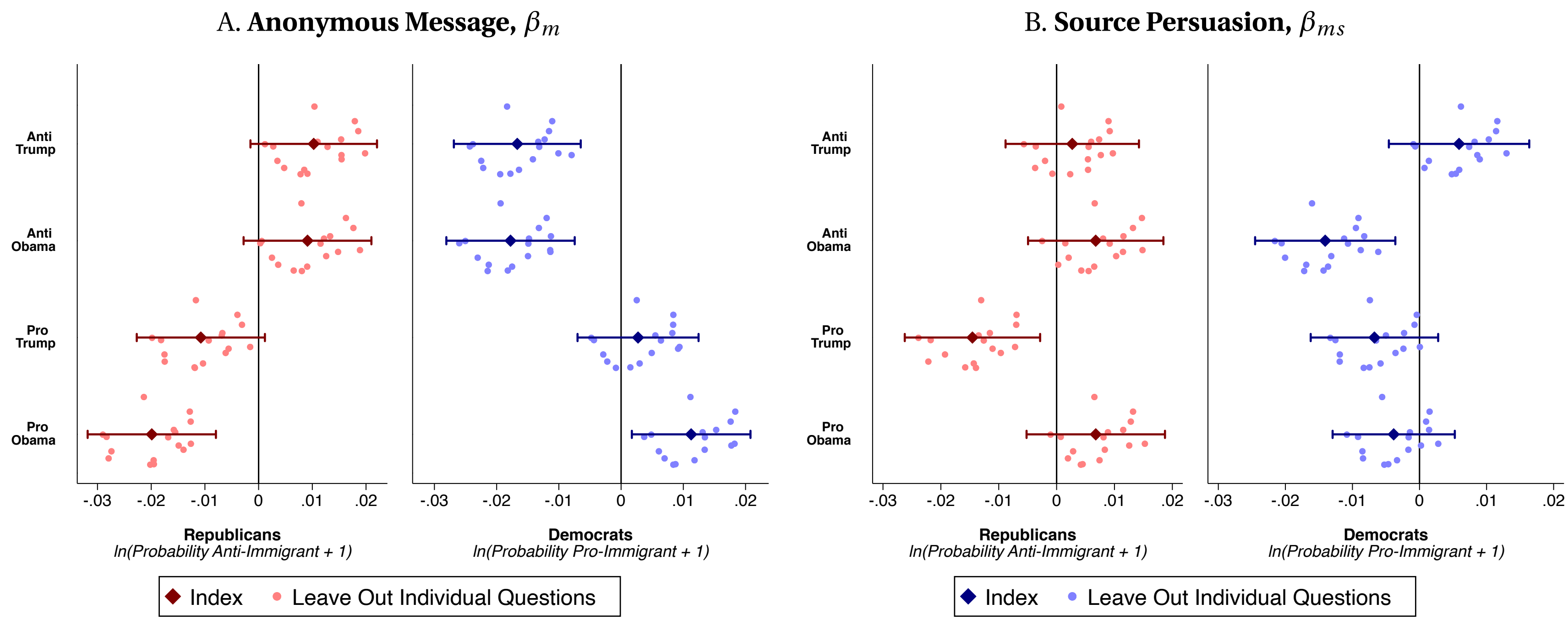
Notes: These figures replicate the baseline specification for alternative sample restrictions and specifications. The specifications compare the baseline transformation, $ln(P(y)+1)$, with the alternative transformation, $ln(P(y))$, where any observations with $P(y) = 0$ are dropped. The figure also includes the unlogged outcome $P(y)$ as the dependent variable. The standardized version of the unlogged outcome is not included in this plot given the difference in scale of this outcome; however, the results using this transformation are identical in significance and direction as the unlogged outcome, given that standardization is a linear transformation. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

Figure A7: **Decomposition II:** Change in Probability: Log, Linear Probability, Logit & Probit



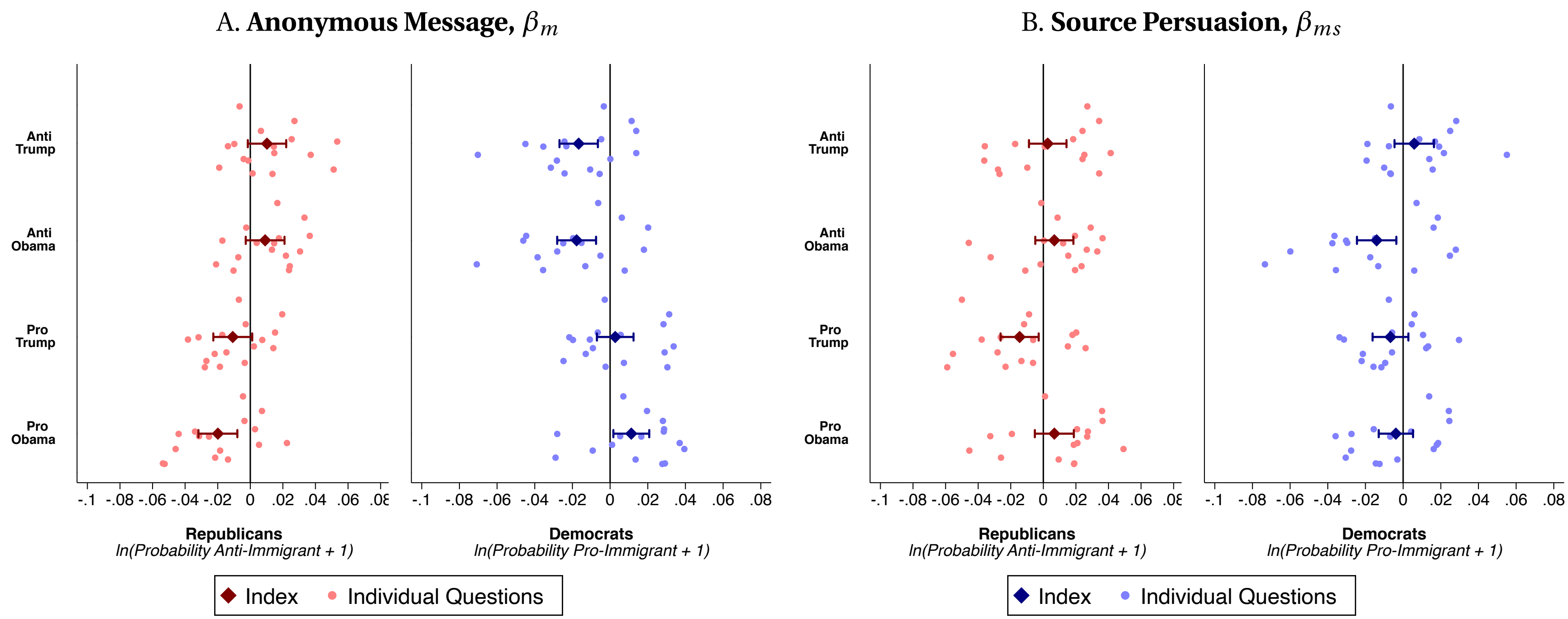
Notes: These figures replicate the baseline specification for alternative sample restrictions and specifications. Unlike the preceding tables, all coefficients from specifications are converted to corresponding changes in underlying probability indexes using the delta method, or estimates of marginal effects (at means) from logit or probit models. The estimates are drawn from the baseline transformation, $ln(P(y)+1)$, a linear probability model with $P(y)$ as the dependent variable, a logit model, and a probit model. To estimate the probit and logit models, the continuous $P(y)$ measures are discretized at 0.5, or a "neutral" response probability. All regressions are adjusted for covariates in Appendix A4.

Figure A8: **Decomposition II:** Leave Out Question Distribution



Notes: This figure overlays the baseline estimates of each test using the full index on top of 16 separate versions of each outcome that each leave out a single question from the index. All outcomes are measured using $ln(P(y)+1)$ and coefficients represent changes in this transformed outcome. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

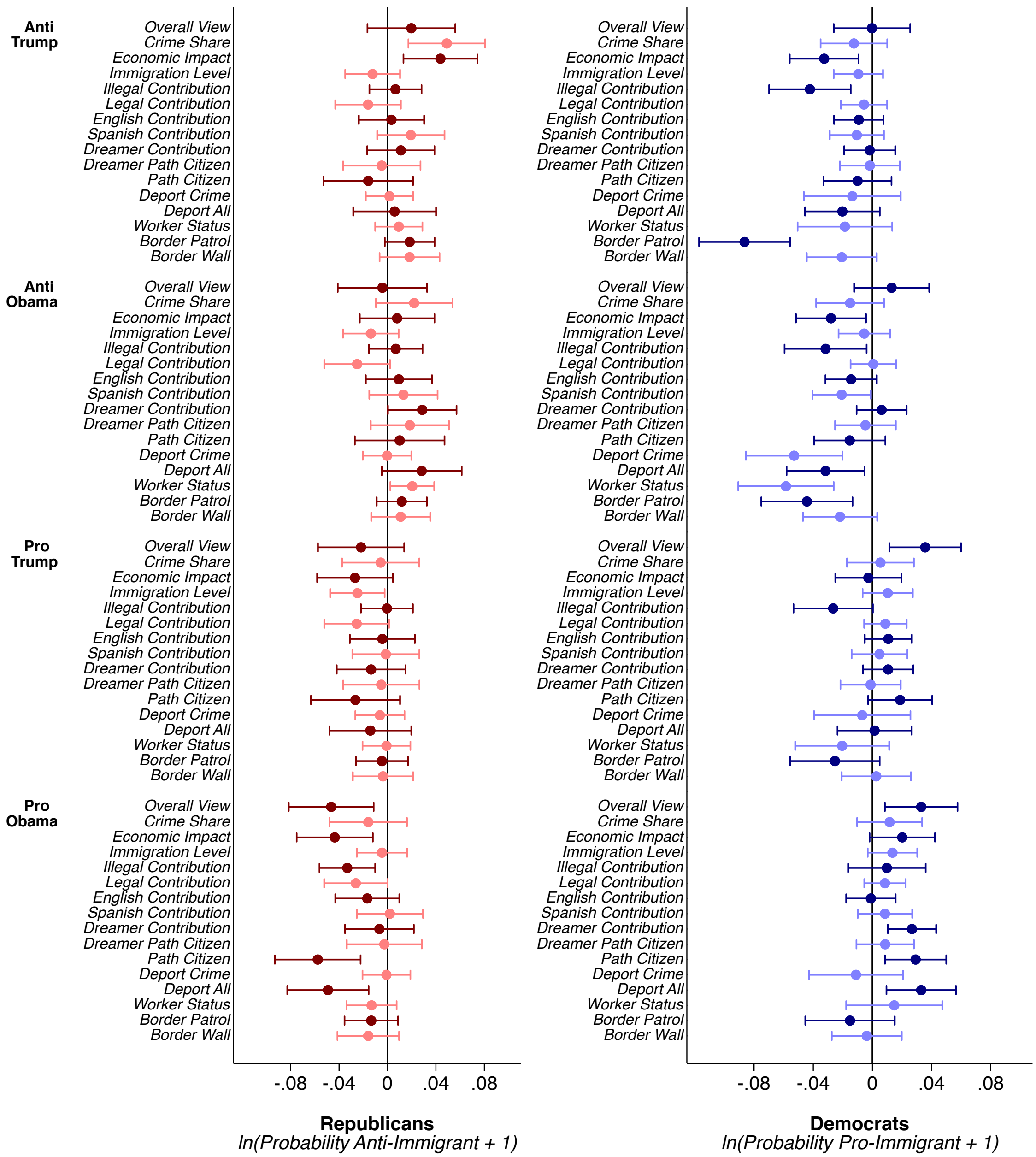
Figure A9: **Decomposition II:** By Question Distribution



Notes: This figure overlays the baseline estimates of each test using the full index on top of the outcome for each of the 16 questions that comprises the index. All outcomes are measured using $ln(P(y)+1)$ and coefficients represent changes in this transformed outcome. Each individual question has the same scale as the index and can assume values of $P(y) \in \{0, 0.5, 1\}$. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

Figure A10: **Decomposition II:** Results by Question

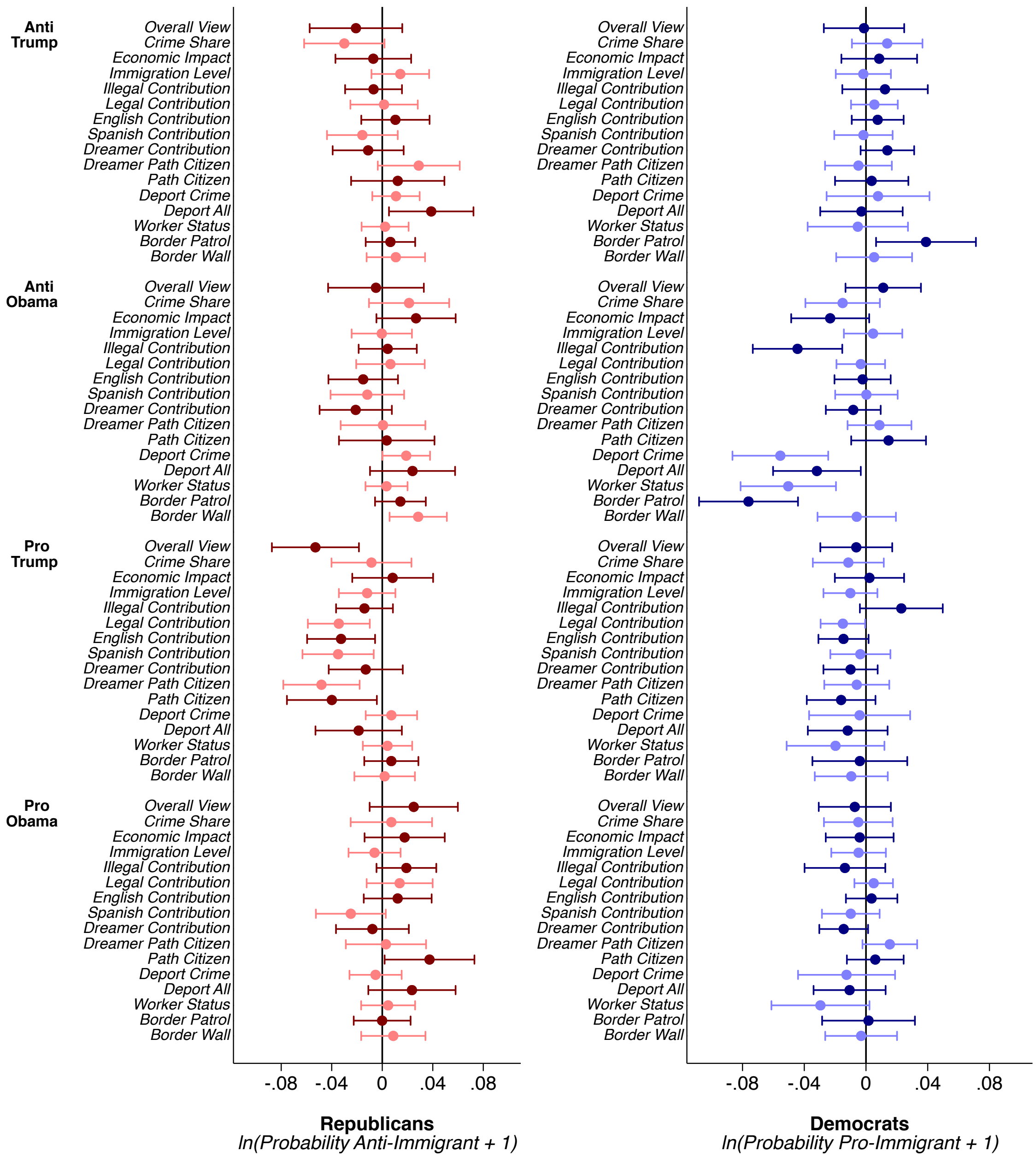
A. **Anonymous Message,** β_m



Notes: This figure displays the baseline estimates of each test using the full index along with results for each of the 16 questions that comprises the index. All outcomes are measured using $ln(P(y)+1)$ and coefficients represent changes in this transformed outcome. Each individual question has the same scale as the index and can assume values of $P(y) \in \{0, 0.5, 1\}$. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

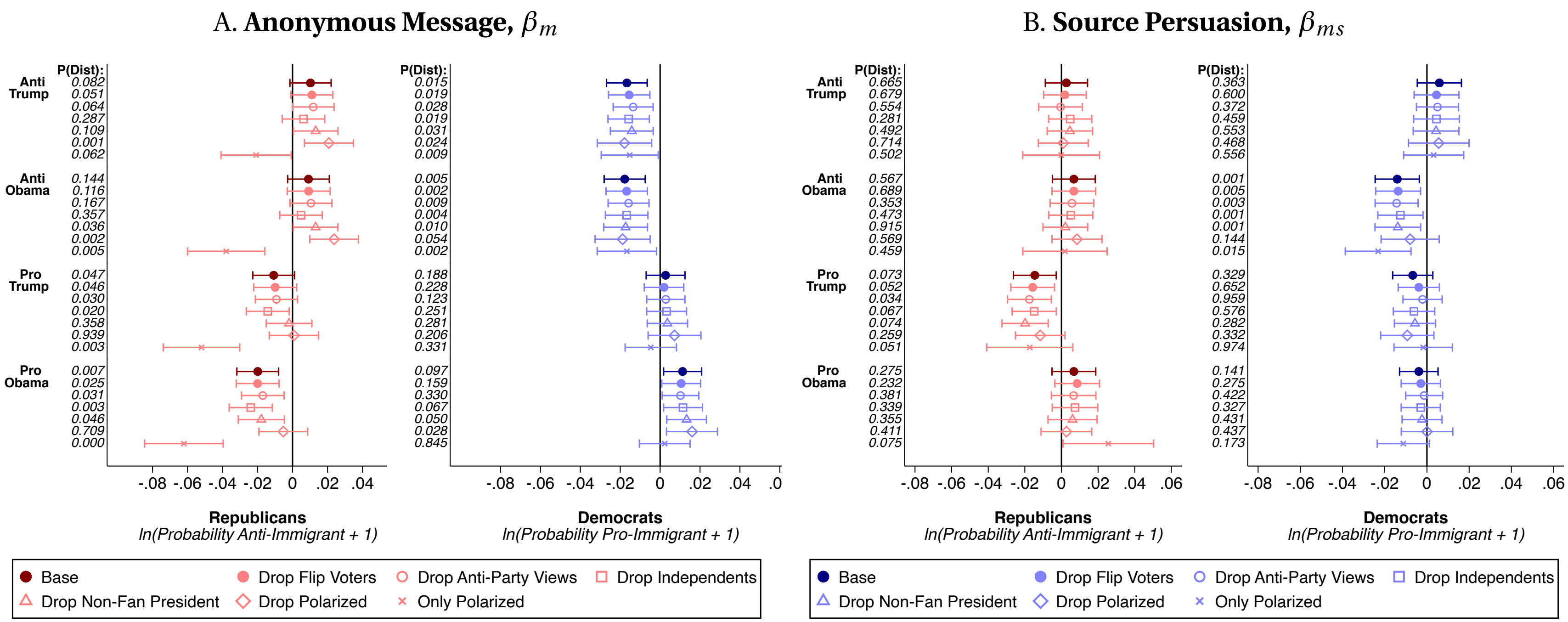
Figure A11: **Decomposition II:** Results by Question (Continued)

B. **Source Persuasion,** β_{ms}



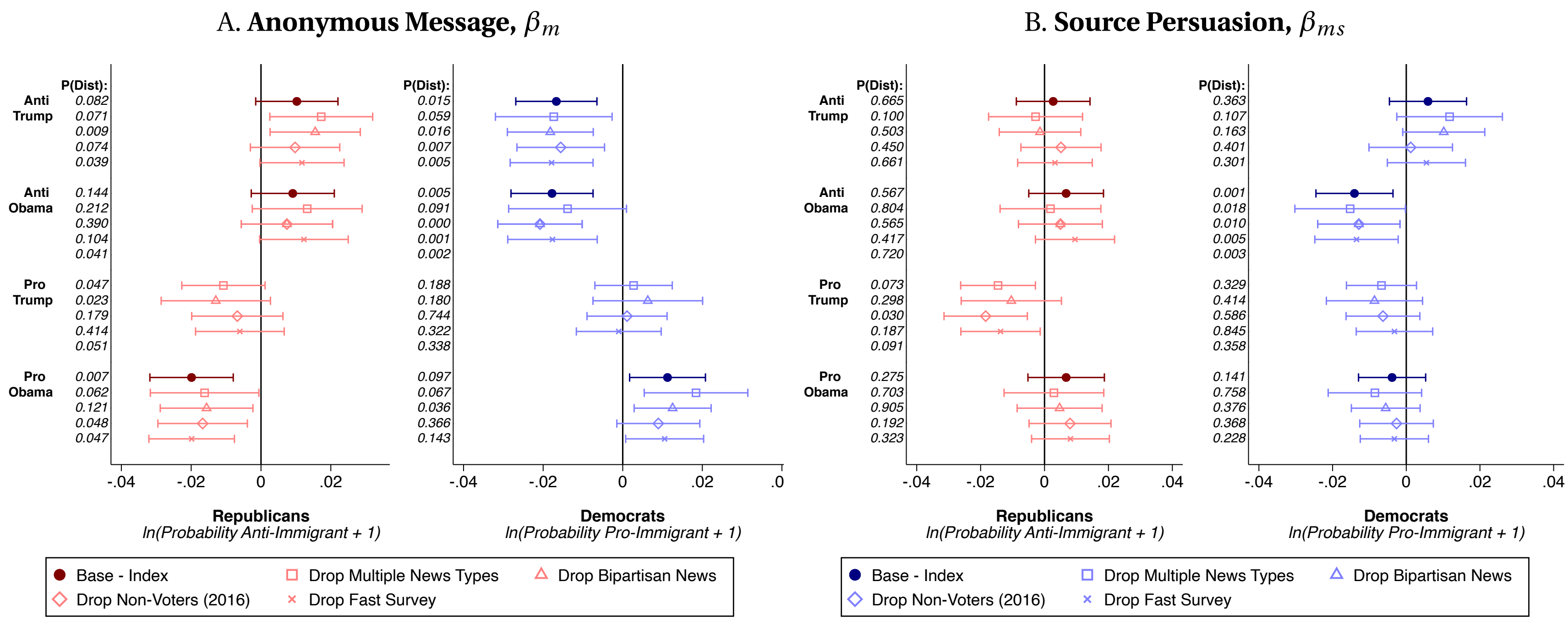
Notes: This figure displays the baseline estimates of each test using the full index along with results for each of the 16 questions that comprises the index. All outcomes are measured using $ln(P(y)+1)$ and coefficients represent changes in this transformed outcome. Each individual question has the same scale as the index and can assume values of $P(y) \in \{0, 0.5, 1\}$. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

Figure A12: **Decomposition II:** Heterogeneity, Moderates and Partisan Respondents



Notes: This plot shows the estimates for the baseline outcomes as compared to samples that remove either moderates or extremists. The outcome is $ln(P(Outcome+1)$. "Drop Flip Voters" removes Republicans who voted for Hillary Clinton in 2016 and Democrats who voted for Donald Trump in 2016. "Drop Anti-Party Views" removes Republicans (Democrats) who answered a pre-treatment question saying that they think immigration should be expanded (restricted). "Drop Independents" drops individuals who were recruited as either Democrats or Republicans and then later stated that they were Independents within the survey. "Drop Non-Fan President" excludes individuals who are not fans of the president from their party. "Drop Polarized" excludes those who are predicted to have extreme immigration views using a probit predictive model with pre-treatment characteristics estimated on the control group and extrapolated to the rest of the sample; "Only Polarized" isolates this group. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

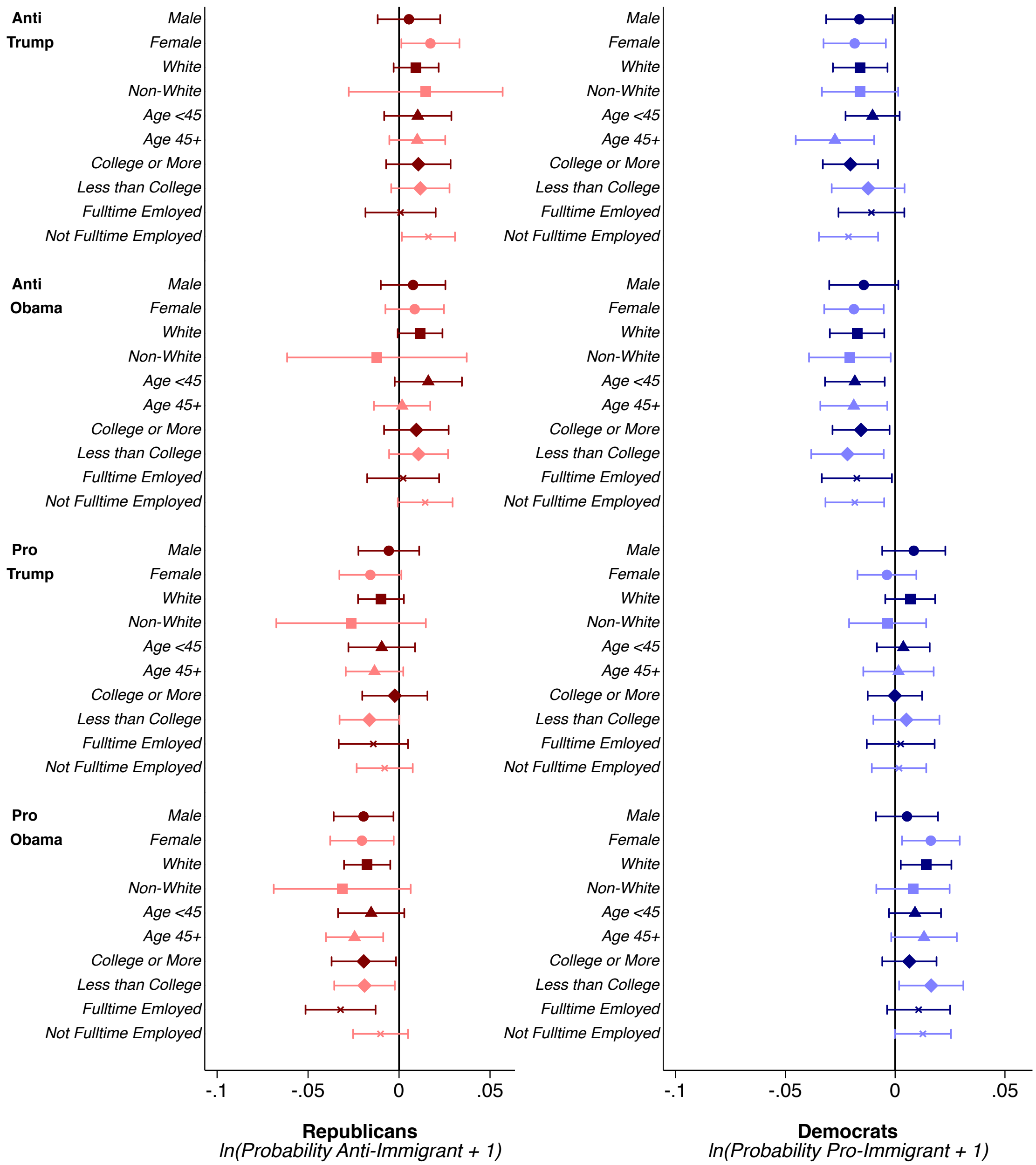
Figure A13: **Decomposition II:** Heterogeneity, Informed and Engaged Respondents



Notes: This plot shows the estimates for the baseline outcomes as compared to samples that remove engaged/informed or non-engaged/non-informed participants. The outcome is $ln(P(Outcome+1)$. "Drop Multiple News Types" removes participants who consume news at least daily through more than one mode: Newspaper, TV, Twitter and/or Facebook. "Drop Bi-Partisan News" removes participants who consume news from both a right-leaning and left-leaning news outlet at least weekly. "Drop Non-Voters (2016)" removes individuals who did not vote in the 2016 presidential election. "Drop Fast Survey" removes individuals who took the survey in 5 minutes or less, adjusted for the time span of audio treatments within the survey. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

Figure A14: **Decomposition II:** Heterogeneity, Demographics

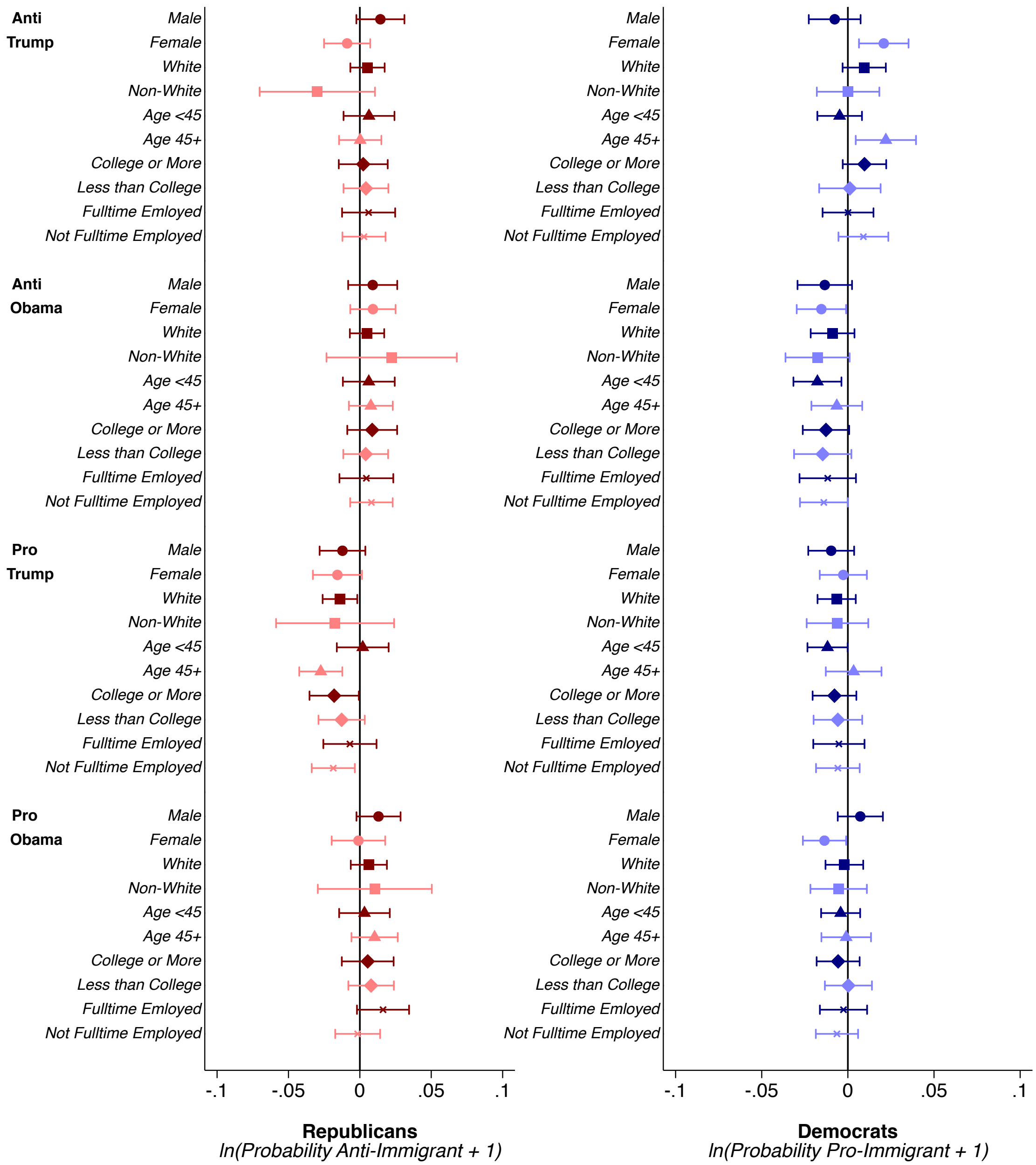
A. **Anonymous Message,** β_m



Notes: This plot shows the estimates for demographic sub-groups of the party samples. Note that the demographics of the party groups differs by design, to approximate the demographics of each political party. The outcome is $ln(P(Outcome+1)$. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

Figure A15: **Decomposition II:** Heterogeneity, Demographics (Continued)

B. **Source Persuasion,** β_{ms}



Notes: This plot shows the estimates for demographic sub-groups of the party samples. Note that the demographics of the party groups differs by design, to approximate the demographics of each political party. The outcome is $ln(P(Outcome+1)$. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.

A3 Conceptual Framework: Additional Details

This section provides detailed derivations for results presented in Section 2.

Derivations for Equation (4). Adding 1 to both sides of Equation (3), we get

$$1 + P(\omega|s) = 1 + P(\omega) \times \Theta(s|\omega) = (1 + P(\omega))(1 + \frac{P(\omega)}{1 + P(\omega)}(\Theta(m|\omega) - 1)) \tag{14}$$

Taking logs of both sides of the equation above gives us Equation (4).

Derivations for Remark 4. Using Equation (8) we have

$$\Theta(s|\omega, m) - 1 = \frac{P(s|\omega, m) - P(\omega|m)}{P(\omega|m)} \tag{15}$$

Now, note that

$$P(\omega|m) = P(\omega, s|m) + P(\omega, \neg s|m) = P(s|m)P(\omega|s, m) + P(\neg s|m)P(\omega|\neg s, m) \tag{16}$$

Finally, plugging this into the numerator in Equation (15) and using $P(\neg s|m) = 1 - P(s|m)$ gives us the expression of interest:

$$\Theta(s|\omega, m) - 1 = \frac{(1 - P(s|m))(P(\omega|s, m) - P(\omega|\neg s, m))}{P(\omega|m)} \tag{17}$$

Derivations for Equation (10). Adding 1 to both sides of Equation (8), we get

$$1 + P(\omega|s, m) = 1 + P(\omega|m) \times \Theta(s|\omega, m) = (1 + P(\omega|m))(1 + \frac{P(\omega|m)}{1 + P(\omega|m)}(\Theta(s|\omega, m) - 1)) \tag{18}$$

Taking logs of both sides of the equation above gives us Equation (10).

A4 Selecting of Control Variables: Double Selection Lasso

Approach

To select the control variables for our baseline specification we use the post double selection lasso (PDS) methodology of Belloni et al., (2012, 2014, 2015, 2016), using the STATA package by Ahrens et al., (2020). The procedure consists of first including all potential control variables from baseline and converting categorical variables into sets of indicators. Our set of potential controls is quite rich given the large number of survey questions we ask to participants prior to the treatment. These questions can be viewed in Supplemental Materials Appendix S1.

Next we keep the controls selected from the LASSO minimization of a regression which predicts treatment assignment. Then we repeat this exercise for a regression that predicts the outcome variable and take the union of controls selected from both procedures as our baseline controls. We do this procedure separately for Democrats and Republicans, given that our randomization was stratified by party and that the demographic characteristics of party groups differ in meaningful ways.

Control Variables in the Model

From the first step, none of the control variables are selected under the lasso procedure as important predictors of treatment status. From the second step, several variables are selected which are predictive of the outcome variable.

Republican Sample

The variables selected for Republicans are whether a participant is Hispanic, is in the age group 55 to 64, voted for Hillary Clinton in 2016, voted for a third party candidate in 2016, had the view that the level of immigration in the U.S. should decrease (pre-treatment), has the view that gun control regulations should be less strict, has the view that abortion should be illegal in all cases, believes taxes are too high, believes that there should be no government intervention in the healthcare system, is an occasional twitter user, is an occasional Buzzfeed reader, is a fan of Donald Trump, is a fan of Lebron James, is a fan of Taylor Swift, is a fan of Bill Gates, is a fan of Barack Obama, has no opinion towards Barack Obama, and views the most important policy issue to their vote as healthcare.

Democrat Sample

The variables selected for Democrats are whether a participant is Black, is Hispanic, is in the age group 35 to 44, is in the age group 45 to 54, has a high school degree as their highest level of education, voted for Hillary Clinton in 2016, had the view that the level of immigration in the U.S. should decrease (pre-treatment), had the view that the level of immigration in the U.S. should stay the same (pre-treatment), has the view that abortion should be illegal in all cases, has the view that abortion should be legal in some cases, believes taxes are too high, believes that there should be no government intervention in the healthcare system, feels neutral about government intervention in the healthcare system, is a daily, weekly, or occasionally reader of the New York Times (separate indicators), is a daily or occasional viewer of TV as a news source (separate indicators), is a daily newspaper reader, is a daily, weekly, or occasional viewer of Fox News (separate indicators), is a daily or weekly Breitbart News reader (separate indicators), is an occasional Buzzfeed reader, has no opinion of Lebron James, is a Barack Obama fan, is a Donald

Trump fan, has no opinion of Donald Trump, lives in the Western U.S., region response west, and views the most important policy issue to their vote as taxes.

While these variables improve the precision of estimates, it is important to note that the models are robust to excluding all controls, as is shown in Appendix Figure A6.

A5 Out of Sample Outcomes Measured in Survey

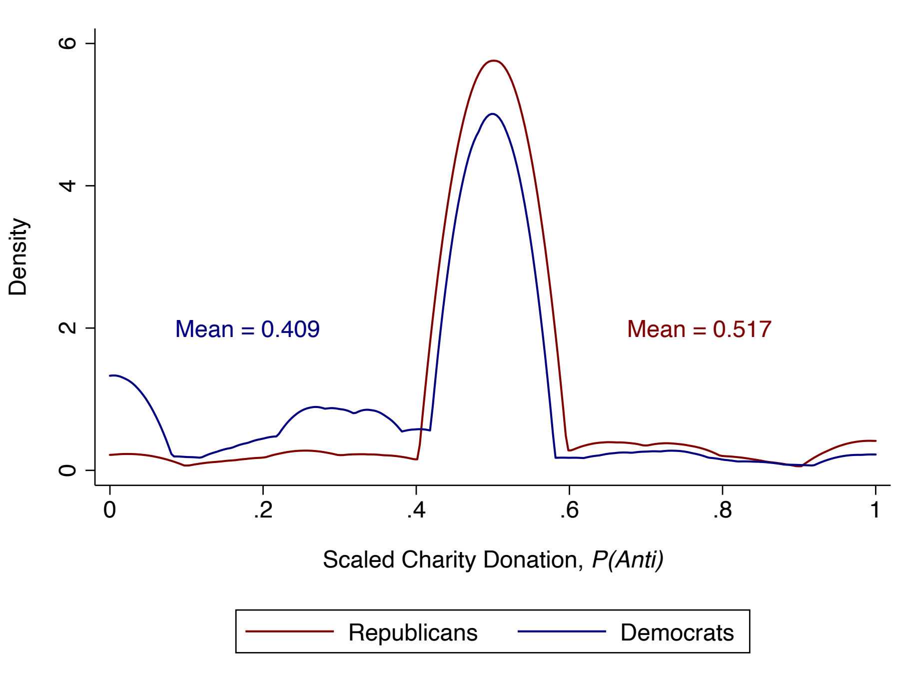
In addition to collecting information on immigrant views, we also attempted to measure two out-of-sample outcomes as a way to capture how actions might respond to a change in immigration beliefs. First, we asked participants about their intended vote in the 2020 presidential election, where we interpret a vote for Donald Trump as an anti-immigrant outcome and a vote for Joseph Biden as a pro-immigrant outcome. Second, we asked whether the participant would be interested in donating money to an anti-immigrant (Federation for American Immigration Reform (FAIR)) or a pro-immigrant charity (American Civil Liberties Union (ACLU) Immigrants' Rights Project), out of potential lottery earnings, if the participant wins a lottery that was conducted as part of the survey. The lottery offered a potential prize of $25 payable directly to the winner. If the winner chooses to donate to one of the two charities, she could choose the amount to be donated of the earnings. Any donation would be doubled before donation, such that the maximum donation amount was $50.

Figure A17 shows the results for these two outcomes as compared to the baseline immigration index outcomes. Unfortunately, the results here are noisy, imprecise and cannot reject the null hypothesis of no effect for either outcome. In cases where there is a significant change in the coefficient (or mean difference across groups), there is no corresponding significant difference in the distributions using the Kolmogorov-Smirnov test, and vice versa.

Multiple factors may contribute to the lack of effects we find for these outcomes. First, we may fail to find effects for these outcomes could be due to lack of precision and limited power. In addition, our ability to measure the voting outcome is limited by the fact that many participants may have voted prior to taking the survey, through early or on-time voting at polling places. While we intended to complete the study prior to the election on November 3, 2020, our survey recruitment period took longer than we had anticipated and continued past the presidential election to to November 10th. We need to exclude the data collected on or after the election for the voting outcome, which also limits the sample size for this test.

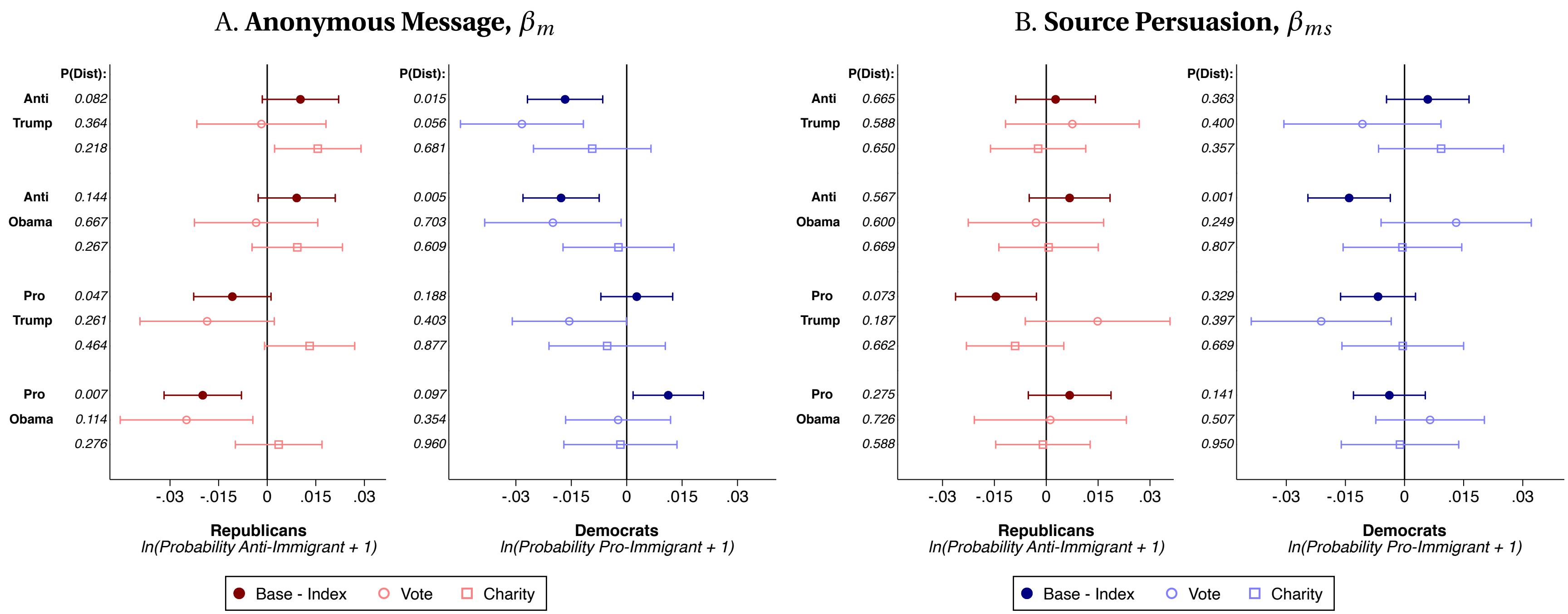
For the charity donation outcome, our total variation in the outcome is quite limited. Figure A16 plots the charity donation outcome in the control group of the study, where this outcome is normalized to be between 0 and 1. Here, 0 corresponds to a choice to donate the entire potential prize to the ACLU (pro-immigrant charity), 1 corresponds to the choice to donate the entire potential prize to FAIR (anti-immigrant charity), and 0.5 corresponds to the choice not to donate to either charity. Nearly all survey participants chose to award themselves the money (if they were to win). One reason that we may not find frequent responses to donate to one of the charities is that the prize may not be big enough, a constraint that we faced given our research budget. Effectively, the lack of variation we see in this outcome contributes to our limited ability to observe meaningful or significant changes in the charity donation choice after different treatments. It is not surprising that we find noisy and imprecise effects for this outcome given that nearly all participants chose to give themselves the prize, regardless of treatment status.

Figure A16: Charity Donation Outcome in No Audio Control Group



Notes: This plot shows the baseline distribution of the charity donation outcome in the no audio control group. The charity outcome is measured from a question where participants are told that they have been automatically entered into a lottery for $25, of which they can donate a portion to either an anti-immigrant charity, the Federation for American Immigration Reform (FAIR), or a pro-immigrant charity, the American Civil Liberties Union (ACLU) Immigrants' Rights Project, at a rate that will double any dollar that is donated. This outcome re-scales the donation choice based on these dollar donations to have a scale of $[0, 1]$. A value of 0.5 corresponds to the participant electing not to make any donation.

Figure A17: **Decomposition II:** Out of Sample Outcomes



Notes: This plot shows the estimates for the baseline outcomes as compared to two out of sample outcomes. The outcome is $ln(P(Outcome+1)$. Vote is whether the participant stated that they intended to vote for Joseph Biden or Donald Trump in 2020, where a Biden vote is coded as pro-immigrant and a Trump vote is coded as anti-immigrant. This outcome is measured as $ln(1[Vote]+1)$. The charity outcome is measured from a question where participants are told that they have been automatically entered into a lottery for $25, of which they can donate a portion to either an anti-immigrant charity, the Federation for American Immigration Reform (FAIR), or a pro-immigrant charity, the American Civil Liberties Union (ACLU) Immigrants' Rights Project, at a rate that will double any dollar that is donated. This outcome re-scales the donation choice based on these dollar donations to have a scale of $[0,1]$. For example, a fully anti-immigrant choice would be donating all of the lottery winnings to the anti-immigrant charity and a fully pro-immigrant choice would be donating all lottery winnings to the pro-immigrant charity. Again, this outcome is entered in the regression as $ln(Charity+1)$. All regressions are adjusted for covariates in Appendix A4. "P(Dist)" is the p-value from a Kolmogorov-Smirnov test of equality of distributions between the treatment and comparison groups for each estimate.