Freedom of Expression in Interpersonal Interactions

Supplemental Appendix

A. Pre-Analysis Plan

Introduction

Americans increasingly are concerned about freedom of expression in spaces such as college campuses, the workplace, and on social media (e.g. Lukianoff and Haidt 2018). One particular context that has not received as much scholarly attention is within interpersonal conversations, both online and face-to-face. Previous research has shown that individuals are often hesitant to reveal their true political opinions in conversations, instead remaining silent, self-censoring their views, or conforming to the group's majority opinion (e.g. Levitan and Visser 2016; Carlson and Settle 2016, n.d.; Noelle Neumann 1974). In a forthcoming book manuscript, we put forward a theoretical framework that can be used to generate expectations about when people are less likely to express their true opinion, such as when they face partisan disagreement or encounter highly knowledgeable discussants. Our results suggest that people prioritize preserving their esteem in others' eyes and their social relationships above the free exchange of opinion. In this study, we propose to build on our prior work by expanding our scope of inquiry to include conversational dynamics that push beyond free expression concerns in the face of partisan disagreement, to investigate new areas of *intraparty* tension that could silence some voices from conversations. In this study, we investigate these dynamics using a conjoint experiment that will allow us to evaluate the relative contribution of these new factors, as compared to other factors that have previously been identified as influential.

Previous research suggests that the most important factors affecting individuals' likelihood of expressing their true opinions in conversations were being in a political minority (defined by partisanship, policy, or general disagreement), being less knowledgeable about the topic, and the nature of the relationship between discussants. The first goal of our proposed study is to incorporate two important intraparty dynamics in interpersonal communication that could motivate people to suppress their true opinions, as these dynamics reveal rifts emerging in American politics that could affect conversational dynamics in pernicious ways. First, Hersh (2020) and Krupnikov and Ryan (forthcoming) both note that there is a small minority of the public that is extremely engaged in politics, while the vast majority does not much care for politics. This divide between those who are engaged in politics and those who are not crosses partisan lines, meaning that there can indeed be intraparty tensions that we suspect could make individuals unwilling to share their opinions. Second, the increased scholarly, journalistic, and public attention to the quality of information available in the media ecosystem has raised the possibility of yet another intraparty divide. Individuals might be uncomfortable around people within their own party who rely on news sources that they deem to be questionable. [1] In this case, individuals withhold from free exchange with their peers not on the basis of disagreement, but rather because their peers base their opinions on low quality information, thereby contributing to the misinformation problem.

[1] By questionable news sources, we are referring to "fringe," hyperpartisan news outlets, in addition to sources that have been identified as fake news. We acknowledge that perceptions of news credibility vary significantly from one person to the next, but we focus on news outlets generally deemed by fact-checking organizations to be questionable.

The results from this survey experiment will have immense implications for our understanding of free expression in everyday talk in America. Understanding when people are most likely to censor themselves under these specific contextual conditions can help us think more critically about the implication of self-censorship for representation. For example, if moderates who trust the mainstream media are continually silencing their views, the vocal minority of politically extreme people who receive information from less credible sources could come to dominate American public discourse. These opinion extremists could be more likely to be heard – and represented – by elected officials at all levels of government. Moreover, the knowledge from this study could also help academics and civil society organizations design better conversational interventions to help generate more effective dialogue. Pinpointing the factors that make people hesitant to express their views can help us think more carefully about what messaging strategies might help individuals feel more comfortable expressing their opinions. The answer to this question has important implications for the potential of everyday political talk to serve as a remedy for affective polarization, as examined by scholars such as Matthew Levenduksy, Erin Rossiter, and Magdalena Wojcieszak.

Hypotheses

Pulling together our prior results with new lines of inquiry generates the following hypotheses about the contexts of interpersonal conversation that exacerbate or mitigate true opinion expression:

- H1: Individuals will be less willing to express their true opinion in a conversation with someone who receives information from questionable news sources, compared to a conversation with someone who receives information from mainstream news sources.
- H2: Individuals will be less willing to express their true opinion in a conversation with someone who identifies with the opposite political party, compared to a conversation with someone who identifies with the same political party.
- H3: Individuals will be less willing to express their true opinion in a conversation with someone who is politically engaged, compared to a conversation with someone who is not politically engaged.
- H4: Individuals will be less willing to express their true opinion in a conversation with someone who they don't know well, compared to a conversation with someone they do
- H5: Individuals will be less willing to express their true opinion in a conversation with someone who is politically knowledgeable, compared to a conversation with someone who is not politically knowledgeable.

Building directly on our previous research, we propose to test the core hypotheses articulated above. However, part of the nature of this project is exploratory given that it is the first time anyone has examined the way in which these features affect opinion expression relative to each other. As we describe briefly in the Exploratory Analysis section of this pre-analysis plan, we anticipate examining additional relationships, such as the impact of shared identities, broadly construed to include partisanship, race, and gender. Moreover, we anticipate that individual characteristics of the respondent, such as their level of trust in the media, could moderate the treatment effects. We will pre-register these as exploratory analyses.

We do not have strong a priori expectations about which factors exert the strongest effects on opinion expression. While previous research would suggest that (dis)agreement is the most

important factor, it has generally been studied in isolation, rather than in combination with other correlated factors. We are therefore *not* pre-registering hypotheses about which attribute in our study is *most* important in determining the likelihood of free opinion expression. In an exploratory analysis, we will examine the average marginal component effect (AMCE) for each factor will be crucial in understanding which of these political divides in American politics has the strongest effect on free expression on the most personal level: in our interpersonal conversations.

Methods

Data Collection

Our experiment is conducted as part of a module on a broader survey coordinated by Yanna Krupnikov and Eitan Hersh on behalf of the Knight Foundation. The survey will be fielded to a nationally representative panel of U.S. adults drawn from Ipsos's Knowledge Panel. The study was fielded in the summer of 2021, with the exact date to be determined by Krupnikov and Hersh.

Ethical Information

This study has been approved by Institutional Review Boars at both Washington University in St. Louis and the College of William & Mary. All participants recruited from Ipsos's Knowledge Panel agree to a general consent upon becoming panelists. The survey data are collected and stored by Ipsos. The data provided by Ipsos for this research project is fully anonymized. There is therefore no risk that researchers involved in the project will be able to identify individuals who participated in the survey.

Conjoint Experiment Design

Participants will be presented with the following prompt and then asked to complete the task a total of five times. That is, participants will be presented with a profile detailing a discussion scenario with randomized attributes and asked to report how likely they would be expressing their true opinion in that conversation. They will then be presented with a new discussion scenario with randomized attributes and asked to complete the same task. This will occur a total of five times. Table 1 shows the attributes we propose to experimentally manipulate for each discussion scenario, as well as the levels of those attributes that would be randomly shown to participants. The rightmost column indicates any notes about how the levels of the attribute will be randomized. In brief, all attributes are fully randomized (i.e. shown to respondents with equal probability), with the exception of preferred news source. Here, similar to Costa (2020), we set it up so that if a partisan news source is randomly selected, the source matches the randomly selected party identification for the conversation partner. That is, if the conversation partner is a Republican and the preferred news source is mainstream partisan sources, the example would be Fox News; but if the conversation partner in this scenario were a Democrat, the example news source would be MSNBC.

Table 1: Conjoint Design

Attribute	Levels	Randomization Notes
Relationship Text to Display: Your relationship to the person	You have a close relationship with the person You have met the person before, but don't consider them to be close	Fully randomized
Conversation Context Text to Display: Context of the conversation	The conversation occurs face-to-face The conversation occurs on social media	Fully randomized
Party Identification Text to Display: The person's party identification	Strong Republican Republican Independent Democrat Strong Democrat	Fully randomized
Political Knowledge Text to Display: The person's knowledge about politics	This person knows a lot more about politics than you This person knows a lot less about politics than you	Fully randomized
Preferred News Source Text to Display: Where the person typically gets their news	This person typically gets their news from mainstream sources, such as USA Today This person typically gets their news from mainstream partisan sources, such as MSNBC This person typically gets their news from mainstream partisan sources, such as Fox News This person typically gets their news from fringe news sources that are often discredited by fact-checking organizations	Fully randomized over: mainstream nonpartisan, mainstream partisan, and questionable fringe source. However, IF mainstream partisan is randomly assigned, then the specific source listed will align with the randomly assigned party identification. If the party identification is randomly assigned to Strong Democrat or Democrat, then MSNBC would be shown; if party identification is randomly assigned to Strong Republican or Republican, then Fox News would be shown. If party identification is randomly assigned to Independent, then Fox News or MSNBC will be randomly shown with equal probability
Political Engagement Text to Display: The person's political engagement	This person is highly engaged in politics This person is not engaged in politics at all	Fully randomized
Race Text to Display: The person's race/ethnicity	White Black Latino/a Asian	Fully randomized IF: Gender = Male, enter Latino IF: Gender = Female, enter Latina
Gender Text to Display: The person's gender	Male Female	Fully randomized

Example Conjoint:

Prompt: We are interested in understanding the types of political conversations in which you would be most likely to express your true opinion. By expressing your true opinion, we mean that you would share what you really think about the political topic being discussed. We will present you with a description of a conversation, with information about the people in the conversation, as well as whether the conversation took place online or face-to-face. We will then ask you to report how likely you would be to express your true opinion in that discussion scenario. We'll ask you to do this five times.

Your relationship to the person	You have a close relationship with the person
Context of the conversation	The conversation occurs face-to-face
The person's party identification	Strong Republican
The person's knowledge about politics	This person knows a lot more about politics than you
Where the person typically gets their news	This person typically gets their news from mainstream sources, such as USA Today
The person's political engagement	This person is heavily engaged in politics
The person's race	White
The person's gender	Female

Measures

Dependent Variable:
Imagine that you were having a conversation about politics in the scenario described above. How likely would you be to express your true opinions in that conversation?

Very unlikely
Unlikely
Neither unlikely nor likely
Likely
Very likely

The dependent variable will be initially coded so that it ranges from 0 (very unlikely) to 4 (very likely) and then rescaled to range from 0 (very unlikely) to 1 (very likely).

Treatment:
The treatment in this conjoint experiment is the randomization of attribute levels shown to participants, as described in Table 1. We will code those levels as shown in Table 2.

Table 2. Summary of Attribute Level Coding

Attribute	Code
Relationship	1=You have a close relationship to the person 0= You have met the person before, but don't consider them to be close
Conversation Context	1=The conversation occurs face-to-face 0=The conversation occurs on social media
Party Identification	0=Strong Republican 1=Republican 2=Independent 3=Democrat 4=Strong Democrat
Political Knowledge	1=This person knows a lot more about politics than you 0=This person knows a lot less about politics than you
Preferred News Source	0=This person typically gets their news from mainstream sources, such as USA Today 1=This person typically gets their news from mainstream partisan sources, such as MSNBC 1=This person typically gets their news from mainstream partisan sources, such as Fox News 2=This person typically gets their news from fringe news sources that are often discredited by fact-checking organizations
Political Engagement	1=This person is highly engaged in politics 0=This person is not engaged in politics at all
Race	0=White 1=Black 2=Latino/a 3=Asian
Gender	1=Female 0=Male

Pre-treatment Covariates:

Several pre-treatment covariates will be provided to us from Ipsos, including: age, sex, education, race/ethnicity, and party identification. Other pre-treatment covariates included in the survey are described as follows:

Table 3. Pre-Treatment Covariate Measures

Covariate	Question	Measure
Political Engagement	Have you done the following in the past year? Participated in a protest, march or rally, contacted an elected official, attended a public meeting, donated to political campaigns or causes [Yes, No to each item]	Total number of items to which respondent selected "Yes," therefore ranging from 0 to 4
Social Media Engagement	How often do you do each of the following on social media? Post links to news stories, discuss news with others on that site [often, sometimes, hardly ever, never, no opinion]	Average of the two items, where 0=never, 1=hardly ever, 2=sometimes, 3=often, and "no opinion" is dropped. Ranges from 0 to 3.
Trust in Media	How much do you trust the following? The news media [to report the news even-handedly]. Great deal, fair amount, not much, not at all, no opinion. Note that half the sample receives just "the news media" and half the sample receives "the news media to report the news even-handedly."	0=not at all, 1=not much, 2=fair amount, 3=great deal. No opinion is dropped.
Interpersonal Conversation	How often, if at all, do you use each of the following for staying up-to-date on news? "Direct communication with people in your local area (outside of your household), including in person, on the phone, or online." Daily, weekly, monthly, less than monthly, never.	0=never, 1=less than monthly, 2=monthly, 3=weekly, 4=daily

Outcome Neutral Quality Checks

In order to verify that randomization was successful, we will regress covariates of interest on the treatment indicators for each attribute level. An indication for successful randomization would show that none of the treatment indicators has a statistically significant association with the covariates of interest. Any imbalance detected will be noted in the manuscript.

Sampling Plan

Recruitment

Our respondents are recruited from Ipsos's Knowledge Panel. Information on how Ipsos recruits their panelists can be obtained directly from Ipsos. In their words, "Panel members are recruited using probability selection algorithms for both random-digit dial (RDD) telephone and address-based sampling (ABS) methodologies. Unlike other Internet research panels that sample only individuals with Internet access and who volunteer for research (i.e. opt-in non-probability panels), KnowledgePanel does not accept self-selected volunteers are part of the KnowledgePanel. Instead, KnowledgePanel is based on a household sampling frame which recruits households: with unlisted telephone numbers, without landline telephones, that are cell phone only, without current Internet access, without devices to access the Internet…Currently, approximately 30% of panel members were recruited through RDD methodology, while 70% were recruited using an ABS methodology. For both ABS and RDD recruitment, households without an Internet connection were provided with a web-enabled device and free Internet service. After initially accepting the invitation to join the panel, participants are asked to complete a short demographic survey (the initial profile survey); answers to these questions allow efficient panel sampling and weighting for surveys. Completion of the profile survey allows participants to become panel members."

Inclusion and Exclusion Criteria

We do not have any inclusion or exclusion criteria.

Analysis Plan

Pre-Processing Steps

The primary dependent variable in this study is the likelihood that the respondent would express his or her true opinion in the conversation described in the profile, which is measured on a scale from 0 (very unlikely) to 1 (very likely). Table 2 describes how the attribute treatments will be coded for analysis. Table 3 describes how pre-treatment covariates will be coded for analysis.

However, we need to go a step further to test some of our hypotheses. For example, we are not particularly interested in whether respondents are more or less likely to express their true opinions to Republicans, compared to Democrats. Rather, we are interested in whether respondents are more or less likely to express their true opinion to copartisans, compared to out-partisans. As such, we will need to construct a copartisanship variable. Table 4 provides details on how we will construct variables that are conditional on respondent characteristics and attribute levels.

Table 4. Respondent-Attribute Variable Coding

Variable	Code
Copartisanship	1 = Attribute is {Democrat or Strong Democrat} AND Respondent PID is {Strong Democrat, Democrat, or Independent who Leans Democrat} 1 = Attribute is {Republican or Strong Republican} AND Respondent PID is {Strong Republican, Republican, or Independent who Leans Republican} 1 = Attribute is {Independent} AND Respondent PID is {Independent} 0 = Attribute is {Democrat or Strong Democrat} AND Respondent PID is {Strong Republican, Republican, Independent who Leans Republican, or Independent} 0 = Attribute is {Republican or Strong Republican} AND Respondent PID is {Strong Democrat, Democrat, Independent who Leans Democrat, or Independent} 0 = Attribute is {Independent} AND Respondent PID is {NOT Independent}
Coethnicity	1 = Attribute is {White} and respondent race is {White} 1 = Attribute is {Black} and respondent race is {Black} 1 = Attribute is {Latino/a} and respondent race is {Latino/a} 1 = Attribute is {Asian} and respondent race is {Asian} 0 = Otherwise (i.e. if Attribute race is not the same as respondent race)
Gender	1 = Attribute is {Male} and respondent gender is {Male} 1 = Attribute is {Female} and respondent gender is {Female} 0 = Attribute is {Male} and respondent gender is {Female} 0 = Attribute is {Female} and respondent gender is {Male}

Analysis

Table 5 outlines our empirical strategy for each hypothesis outlined in the Hypothesis section of this pre-analysis plan.

Table 5. Hypotheses and Estimation Strategy

Hypothesis	**Analysis Plan**	**Interpretation**
H1: Individuals will be less willing to express their true opinion in a conversation with someone who receives information from questionable news sources, compared to a conversation with someone who receives information from mainstream news sources.	We will estimate the AMCE of *questionable news source* by estimating a simple linear regression where the dependent variable is the likelihood of expressing one's true opinion (5 point scale, ranging from 0-1), and right-hand side variables include indicators for each attribute level. The unit of analysis is the profile, rather than the respondent. We use cluster-robust standard errors at the respondent level to correct for correlation between responses within a subject.	If individuals are less willing to express their true opinion in a conversation with someone who receives information from questionable news sources, compared to a conversation with someone who receives information from mainstream news sources, the coefficient on *questionable news source* should be negative and statistically significant.
H2: Individuals will be less willing to express their true opinion in a conversation with someone who identifies with the opposite political party, compared to a conversation with someone who identifies with the same political party.	We will estimate the AMCE of *copartisanship* by estimating a simple linear regression where the dependent variable is the likelihood of expressing one's true opinion (5 point scale, ranging from 0-1), and right-hand side variables include indicators for each attribute level, including an indicator for copartisanship – as described in Table 3. The unit of analysis is the profile, rather than the respondent. We use cluster-robust standard errors at the respondent level to correct for correlation between responses within a subject.	If individuals are less willing to express their true opinion in a conversation with someone who is not a copartisan, compared to a conversation with someone who is a copartisan, then the coefficient on *copartisanship* should be positive and statistically significant.
H3: Individuals will be less willing to express their true opinion in a conversation with someone who is politically engaged, compared to a conversation with someone who is not politically engaged.	We will estimate the AMCE of *politically engaged* by estimating a simple linear regression where the dependent variable is the likelihood of expressing one's true opinion (5 point scale, ranging from 0-1), and right-hand side variables include indicators for each attribute level. The unit of analysis is the profile, rather than the respondent. We use cluster-robust standard errors at the respondent level to correct for correlation between responses within a subject.	If individuals are less willing to express their true opinion in a conversation with someone who is politically engaged, compared to a conversation with someone who is not politically engaged, the coefficient on *politically engaged* should be negative and statistically significant.
H4: Individuals will be less willing to express their true opinion in a conversation with someone who they don't know well, compared to a conversation with someone they do know well	We will estimate the AMCE of *close relationship* by estimating a simple linear regression where the dependent variable is the likelihood of expressing one's true opinion (5 point scale, ranging from 0-1), and right-hand side variables include indicators for each attribute level. The unit of analysis is the profile, rather than the respondent. We use cluster-robust standard errors at the respondent level to correct for correlation between responses within a subject.	If individuals are less willing to express their true opinion in a conversation with someone with whom they do not have a close relationship, compared to a conversation with someone to whom they are close, the coefficient on *close relationship* should be positive and statistically significant.
H5: Individuals will be less willing to express their true opinion in a conversation with someone who is politically knowledgeable, compared to a conversation with someone who is not politically knowledgeable.	We will estimate the AMCE of *politically knowledgeable* by estimating a simple linear regression where the dependent variable is the likelihood of expressing one's true opinion (5 point scale, ranging from 0-1), and right-hand side variables include indicators for each attribute level. The unit of analysis is the profile, rather than the respondent. We use cluster-robust standard errors at the respondent level to correct for correlation between responses within a subject.	If individuals are less willing to express their true opinion in a conversation with someone who is politically knowledgeable, compared to a conversation with someone who is not politically knowledgeable, the coefficient on *politically knowledgeable* should be negative and statistically significant.

Statistical Significance
Throughout this study, we use the *p*-value of 0.05 as the value for statistical significance of our two-tailed tests. We consider results where $p<.10$ to be suggestive.

Missing Data
We do not expect missing data on most demographic pre-treatment covariates because they are maintained and supplied by Ipsos. Should we observe missing values on other pre-treatment covariates of interest, such as those in Table 3, these respondents will be dropped from the corresponding analyses that rely on that covariate.

Robustness Checks
We plan to conduct the following robustness checks. We anticipate presenting these results in an online appendix to accompany the manuscript. The results may be discussed in the main text and/or in footnotes of the manuscript.

1. Exclude Independent respondents from the analysis.
2. Recode the copartisanship variable to consider Independent-leaners to be Independents rather than partisans.
3. Recode the coethnicity variable to be White vs. [Latino, Black, or Asian]
4. Present results estimated without cluster-robust standard errors
5. Present results estimated with ordered logit instead of ordinary least squares
6. Recode the dependent variable to a dichotomous measure where 1={Likely or Very Likely} and 0={Neither Likely nor Unlikely; Unlikely; or Very Unlikely}. Replicate main hypothesis tests with this dependent variable and a logit model.

Exploratory Analyses
While our central hypotheses are articulated in H1-H5, there are several additional analyses that we plan to conduct. These analyses may be discussed in the manuscript, but will be clearly labeled as exploratory rather than hypothesis testing.

Moderators
The first group of exploratory analyses involve individual-level characteristics that might moderate the effect of the attributes on opinion expression. We plan to explore these relationships by interacting the individual-characteristic with the attribute in the model. A priori, we are interested in the following relationships, but may explore additional moderators as well:

1. Strength of Partisanship
    a. Weaker partisans, relative to stronger partisans, will be less willing to express their opinion when the discussant is an out-partisan
    b. Weaker partisans, relative to stronger partisans, will be less willing to express their opinion when the discussant is politically knowledgeable
    c. Weaker partisans, relative to stronger partisans, will be less willing to express their opinion when the discussant typically gets their news from questionable fringe sources
    d. Weaker partisans, relative to stronger partisans, will be less willing to express their opinion when the discussant typically gets their news from mainstream partisan sources
2. Trust in Mainstream Media

    a. Respondents who are more trusting of the mainstream media, relative to respondents who are less trusting of the mainstream media, will be more willing to express their opinion when the discussant typically gets their news from mainstream sources
3. Political Engagement
    a. Respondents who are less politically engaged, relative to respondents who are more politically engaged, will be more willing to express their true opinion when the discussant is politically engaged
4. Gender
    a. Female respondents, relative to male respondents, will be less willing to express their true opinion when the discussant is politically knowledgeable
    b. Female respondents, relative to male respondents, will be less willing to express their true opinion when the discussant is politically engaged
    c. Female respondents, relative to male respondents, will be less willing to express their true opinion when the discussant is an out-partisan

Shared Identity

1. Gender
    a. Respondents are more likely to express their true opinion to a discussant who shares their gender
    b. Female respondents, relative to male respondents, will be less willing to express their true opinion when the discussant is male
2. Co-ethnicity
    a. Respondents will be more willing to express their true opinion to a co-ethnic

Post Analysis

Data Archiving
Our data will be made available to other researchers. We will post the study materials and all data on the OSF website upon publication of the research based on the data.

Submission and Modification
We will pre-register this pre-analysis plan on the Open Science Framework website (https://osf.io/prereg). Any changes to the pre-analysis plan we initially file will be made transparently: we will report the change and the justification for the change. Any unregistered analyses will be transparently reported as "unregistered," "exploratory," or "preliminary" findings.

Publication
We will report the results of all preregistered analyses, regardless of outcome.

B. Supporting Tables

Table B.1. Effect of Attributes on Likelihood of True Opinion Expression (supports Figure 1)

	Likelihood of True Opinion Expression
Close Relationship	0.053***
	(0.005)
Face-to-Face Context	0.044***
	(0.006)
Copartisan	0.072***
	(0.007)
Discussant More Knowledgeable	0.002
	(0.005)
Discussant Prefers Fringe Media	-0.036***
	(0.010)
Discussant Prefers Partisan Media	-0.015^
	(0.009)
Discussant Highly Engaged	0.002
	(0.005)
Discussant is Same Race/Ethnicity	0.023***
	(0.007)
Discussant is Same Gender	-0.032***
	(0.008)
Intercept	0.555***
	(0.010)
Observations	13,803
Multiple R-Squared	0.027
Adjusted R-Squared	0.02637
F-statistic	35.35*** (DF = 9; 2,776)

Note: p <.10^, p < .05*, p < .01**, p < .001***

C. Pre-registered Robustness Checks

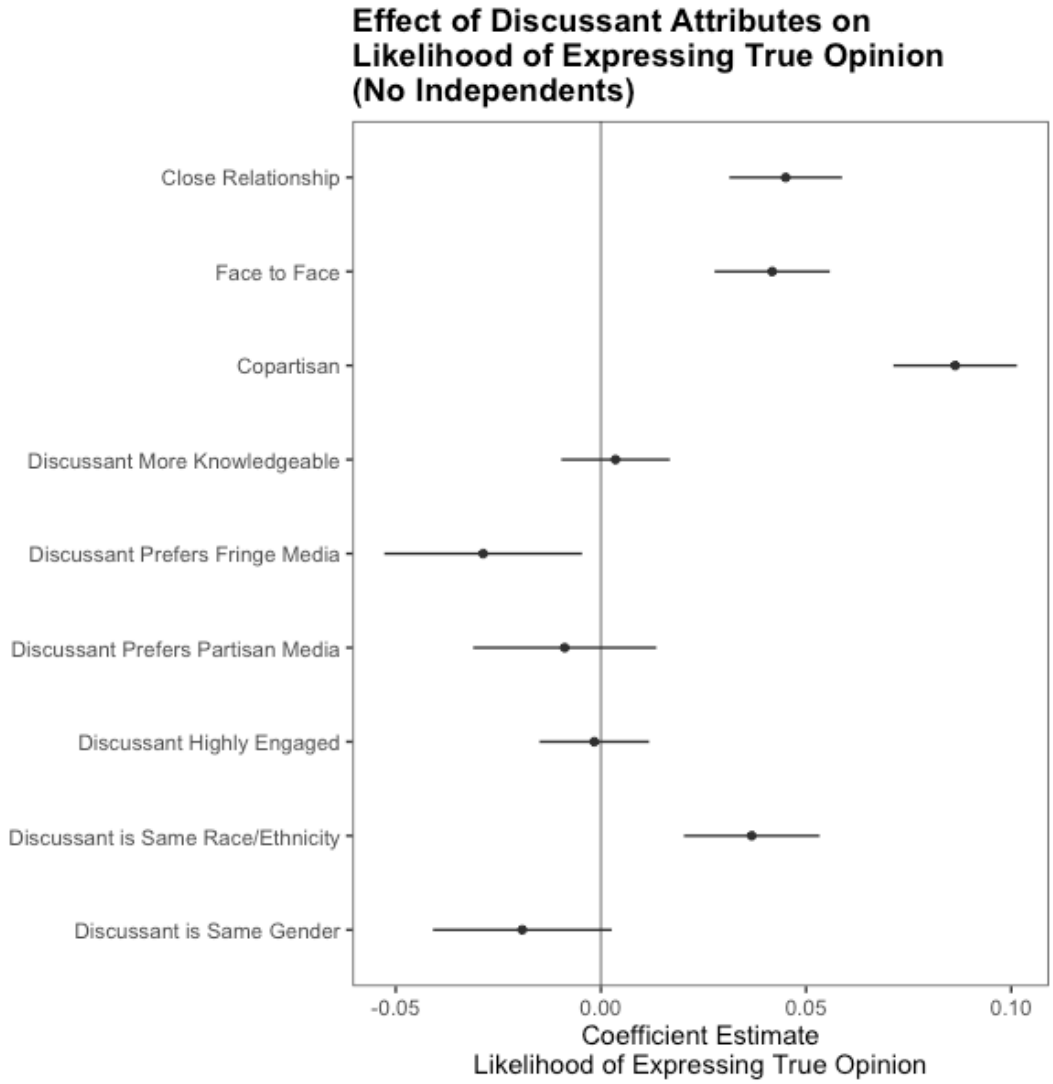

Figure C.1. Effect of discussant attributes on likelihood of expressing true opinion, with Independent respondents excluded. Coefficients estimated using a linear model with robust standard errors clustered at the respondent level. Bars reflect 95 percent confidence intervals.

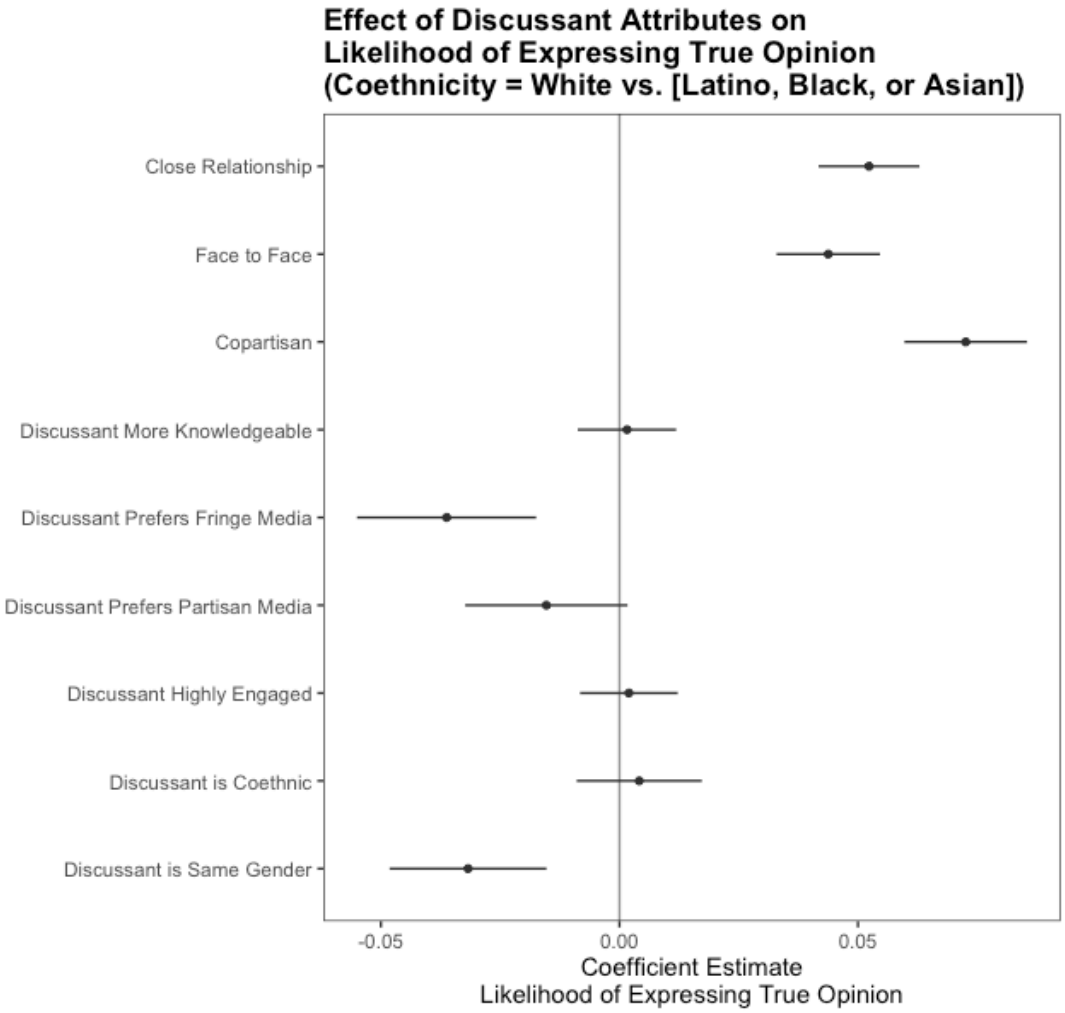

Figure C.2. Effect of discussant attributes on likelihood of expressing true opinion, coding coethnicity as white vs. nonwhite (e.g. white-white pairs are coethnic; nonwhite-nonwhite pairs are coethnic). Coefficients estimated using a linear model with robust standard errors clustered at the respondent level. Bars reflect 95 percent confidence intervals.

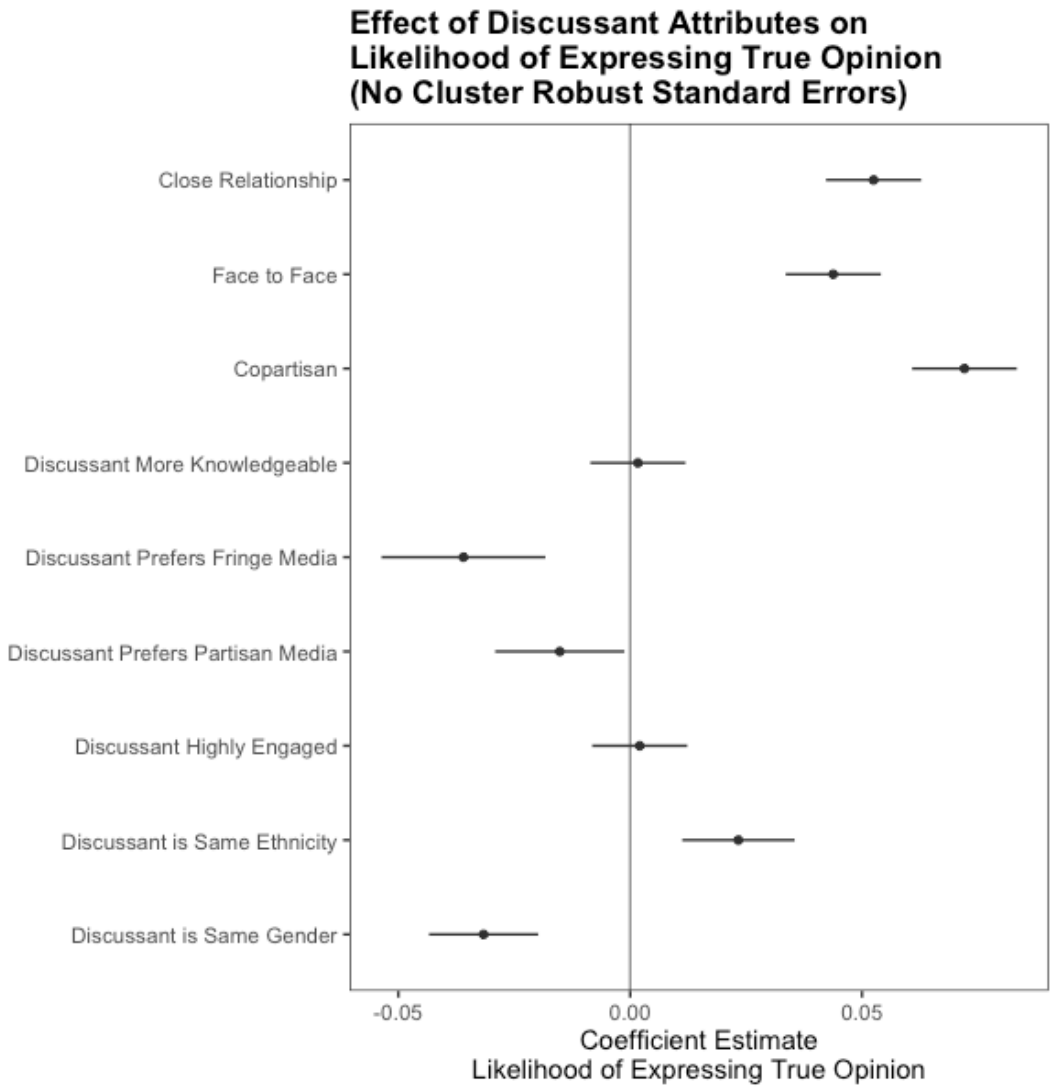

Figure C.3. Effect of discussant attributes on likelihood of expressing true opinion, as displayed in Figure 1 and Table B.1., but without cluster robust standard errors. Coefficients estimated using a linear model. Bars reflect 95 percent confidence intervals.

Table C.1. Effect of Discussant Attributes on True Opinion Expression (Ordered Logit Model)

	Likelihood of True Opinion Expression
Close Relationship	0.312
	(0.031)
Face-to-Face Context	0.250
	(0.030)
Copartisan	0.404
	(0.033)
Discussant More Knowledgeable	0.007
	(0.030)
Discussant Prefers Fringe Media	-0.203
	(0.053)
Discussant Prefers Partisan Media	-0.090
	(0.041)
Discussant Highly Engaged	0.018
	(0.030)
Discussant is Same Race/Ethnicity	0.147
	(0.036)
Discussant is Same Gender	-0.202
	(0.035)
Intercepts	
Very Unlikely | Unlikely	-1.989
	(0.060)
Unlikely | Neither Likely Nor Unlikely	-0.995
	(0.057)
Neither Likely Nor Unlikely | Likely	0.261
	(0.057)
Likely | Very Likely	1.502
	(0.058)
Observations	13,083
Residual Deviance	42232.43
AIC	42258.43
Adjusted R-Squared	0.02637

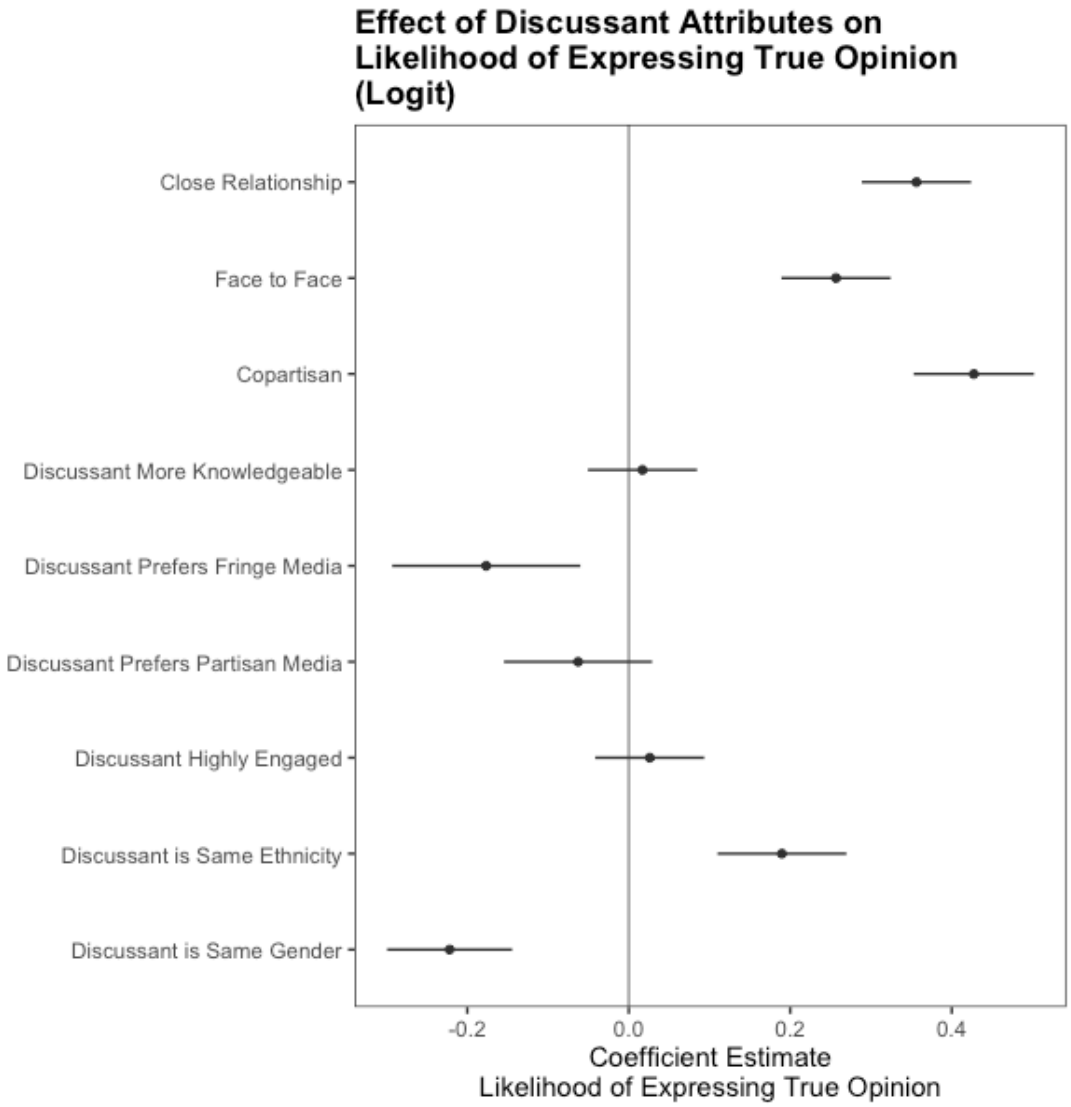

Figure C.4. Effect of discussant attributes on likelihood of expressing true opinion, using a logit model. The dependent variable is coded such that 1=Likely or Very Likely; and 0=Very Unlikely, Unlikely, or Neither Likely nor Unlikely. Bars reflect 95 percent confidence intervals.

D. Pre-registered Exploratory Analyses

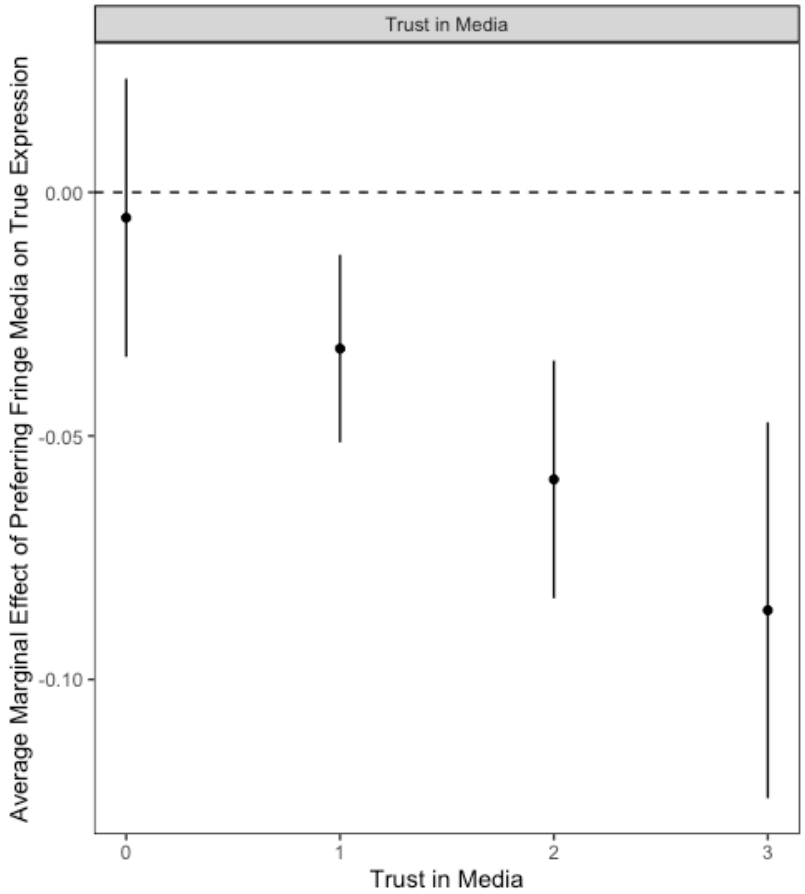

Figure D.1. Marginal effect of the discussant preferring fringe media sources (relative to mainstream media sources) on opinion expression for people at different levels of trust in media. 0 represents the least trust in media; 3 represents the highest trust in media. Vertical lines represent 95 percent confidence intervals.

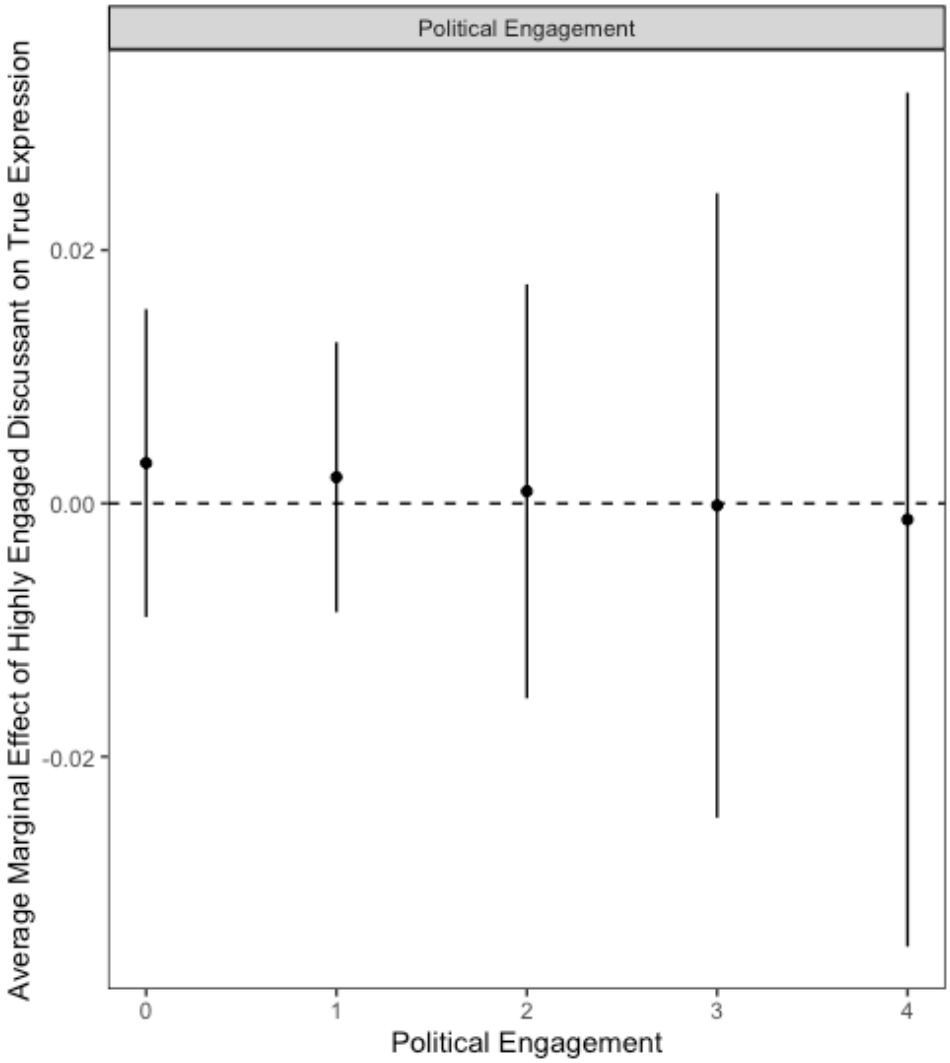

Figure D.2. Marginal effect of the discussant being highly politically engaged (relative to not being politically engaged) on opinion expression for different levels of respondent political engagement. 0 reflects 0 political engagement activities (least engaged); 4 reflects 4 political engagement activities (most engaged).

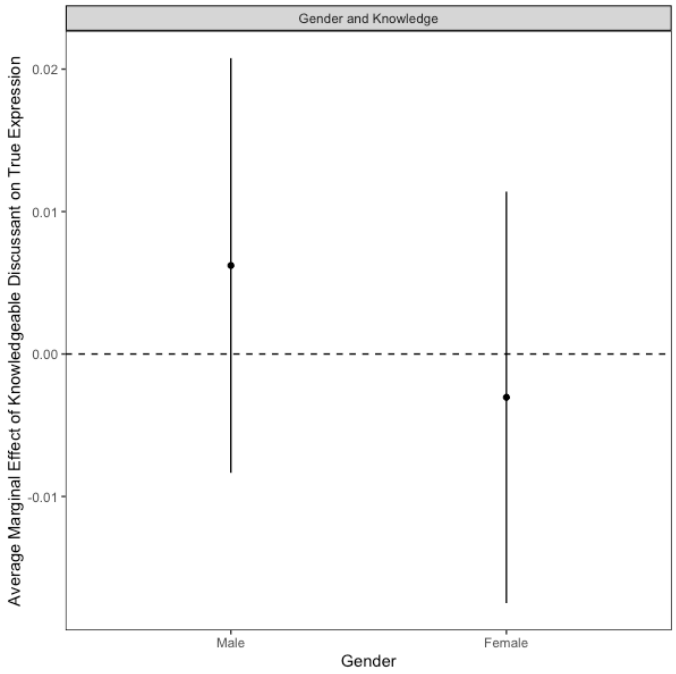

Figure D.3. Marginal effect of a more knowledgeable discussant on opinion expression by respondent gender. Bars reflect 95 percent confidence Intervals.

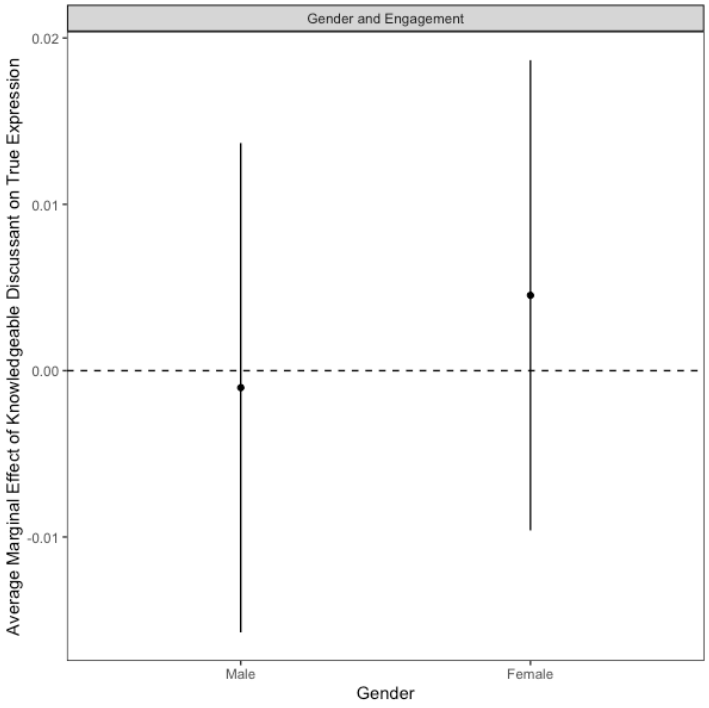

Figure D.4. Marginal effect of a more engaged discussant on opinion expression by respondent gender. Bars reflect 95 percent confidence intervals.

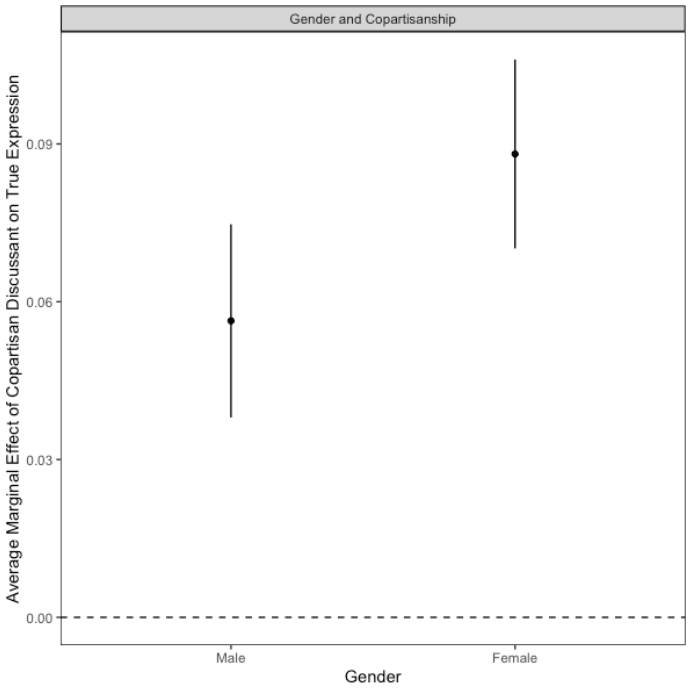

Figure D.5. Marginal effect of a copartisan discussant on opinion expression by respondent gender. Bars represent 95 percent confidence intervals.

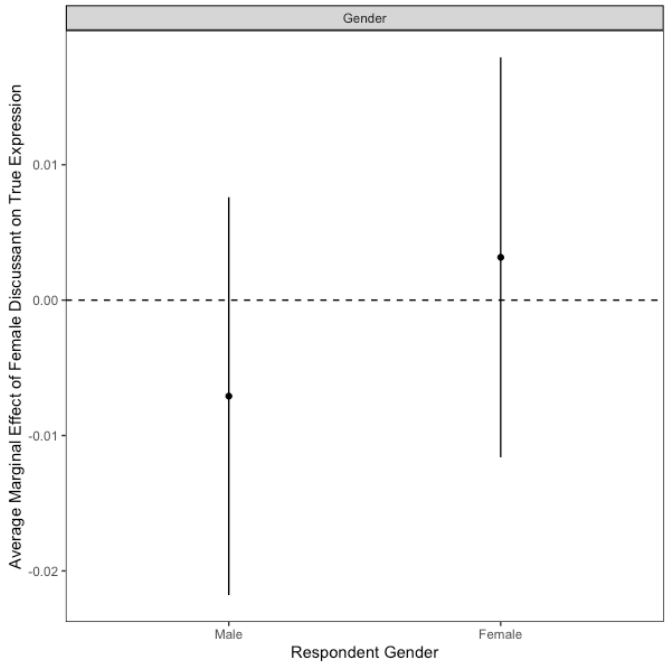

Figure D.6. Marginal effect of a female discussant on opinion expression by respondent gender. Bars represent 95 percent confidence intervals.

E. Marginal Means for Main Result

Although we did not pre-register that we would use marginal means to analyze the results from our conjoint experiment, we present that analysis here. Looking across all ratings, 10% of respondent choices were "very unlikely" (0), 13% were "very unlikely" (.25), 28% were "neither likely nor unlikely" (0.5), 27% were "likely" (0.75), and 22% were "very likely" (1). The overall average was .59. The interpretation of marginal means differs from the interpretation of the AMCE reported in the main text, but the direction and statistical significance is the same for each attribute across both presentations. Figure E.1 was generated using the cregg (Leeper and Barnfield 2020) and ggplot2 (Wickham 2016) packages in R and code adapted from the replication materials provided for Costa (2020).

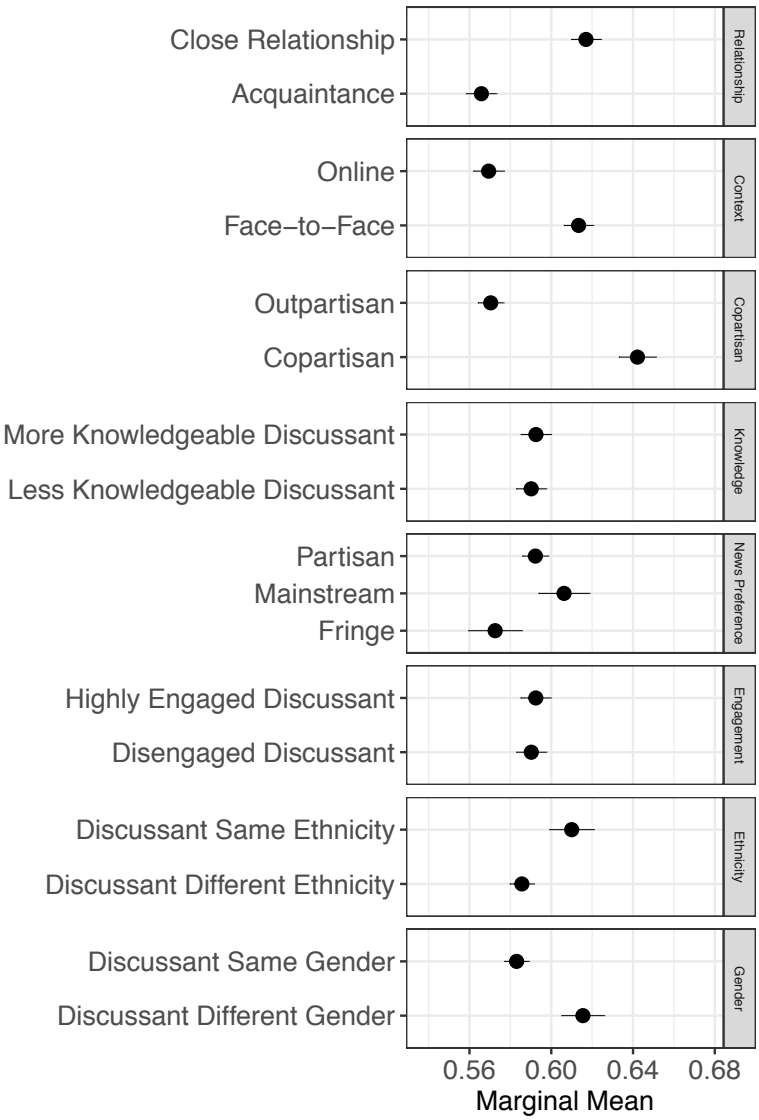

Figure E.1. Marginal means for each attribute using cluster robust standard errors by subject. Bars represent 95 percent confidence intervals.

F. Marginal Means for Subgroup Analyses

We pre-registered that we would examine whether several individual characteristics moderated the effect of each attribute level on respondents' likelihood of reporting their true opinions in the described conversation. We presented these results in Appendix D, plotting the marginal effect of the attribute for various subgroups, based on the AMCE. Leeper, Hobolt, and Tilley (2020) argue that marginal means are a better way to evaluate subgroup analyses, so we present plots of these analyses here. The figures that follow were generated using the cregg package in R (Leeper and Barnfield 2020) and code adapted from replication materials provided for Leeper, Hobolt, and Tilley (2020). We note that these analyses were not pre-registered.

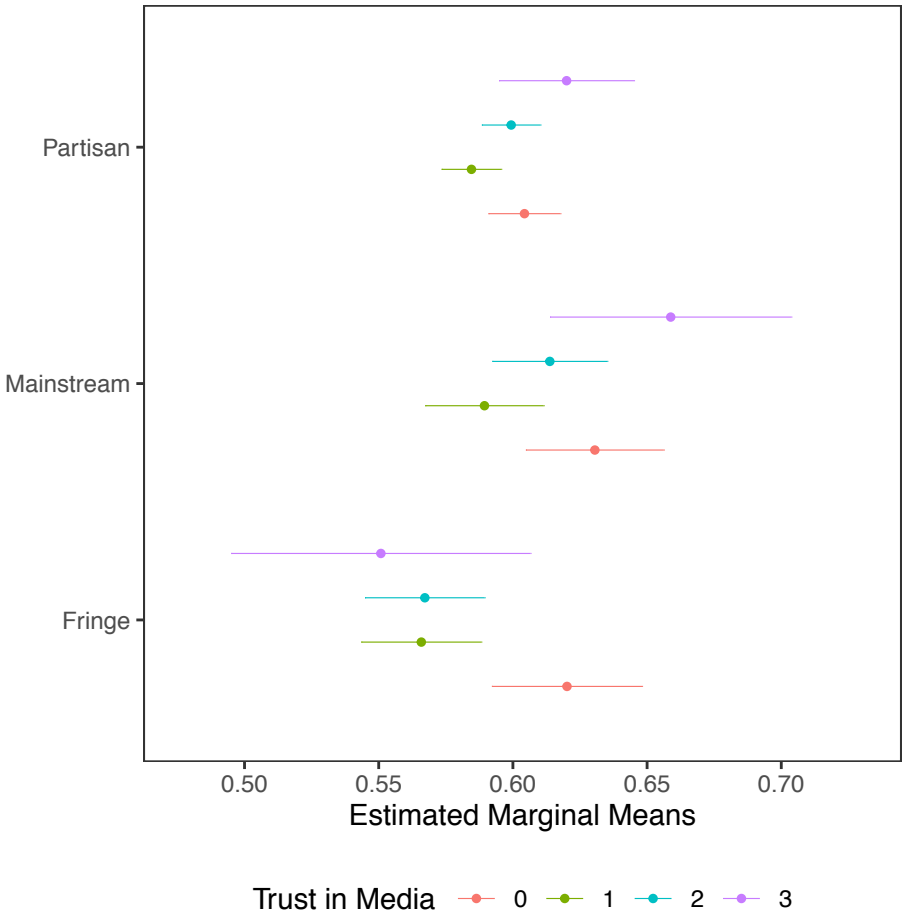

Figure F.1. Estimated marginal means for discussant's preferred news source (partisan, mainstream, and fringe) at different levels of respondent trust in media (0=lowest trust; 3=highest trust). Horizontal lines represent 95 percent confidence intervals. Consistent with Figure D.1, those who trust the media the least (red circle) reported that they would be more likely to express their opinion to someone who relied on fringe media sources than people who had more trust in media (blue and green circles).

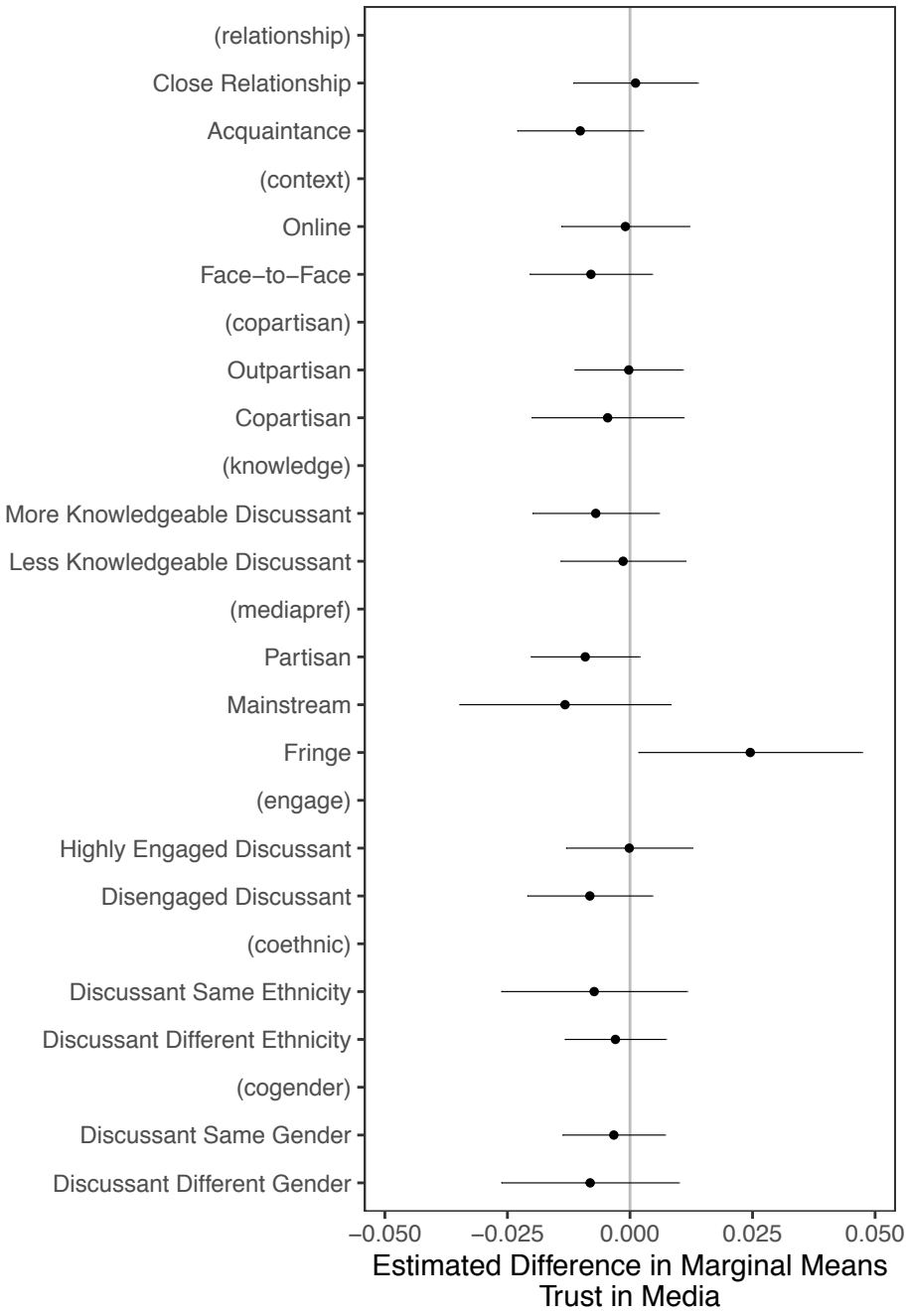

Figure F.2. Estimated difference in marginal means between those with higher (2-3) and lower (0-1) trust in media, across all attributes. Horizontal lines represent 95 percent confidence intervals. There are no statistically significant differences in marginal means of likelihood of expressing one's true opinion across any of the attributes between more and less trusting respondents, except for fringe media. Respondents with lower trust in media were more likely to report that they would express their true opinion to someone who prefers fringe media than people with higher trust in media.

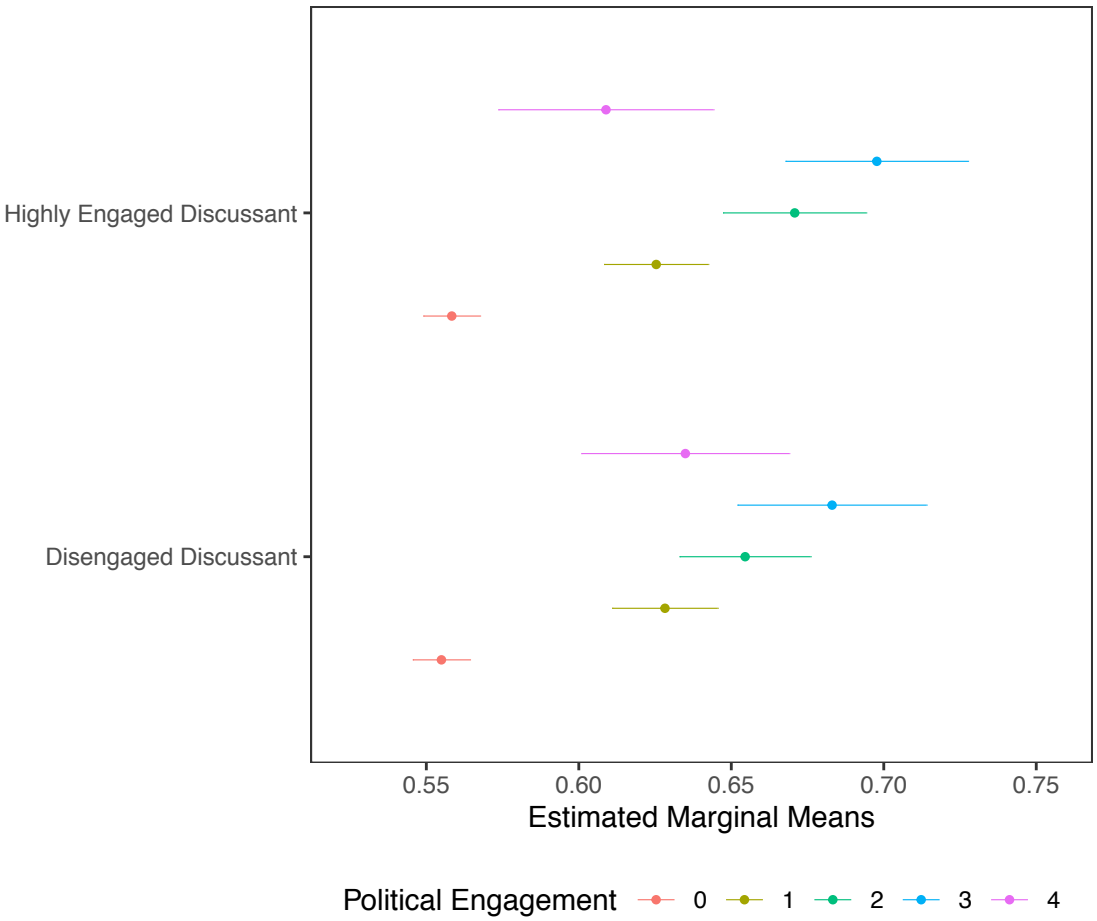

Figure F.3. Estimated marginal means for discussant's level of political engagement (highly engaged vs. not engaged) for respondents who reported participating in different numbers of political engagement activities (0-4). Horizontal lines represent 95 percent confidence intervals. Consistent with Figure D.2, the discussant's level of engagement affected the likelihood one would express their true opinion uniformly for respondents with different levels of political engagement. For both highly engaged and disengaged discussants, the least engaged respondents (red circles) were less likely to express their true opinions than their more engaged peers.

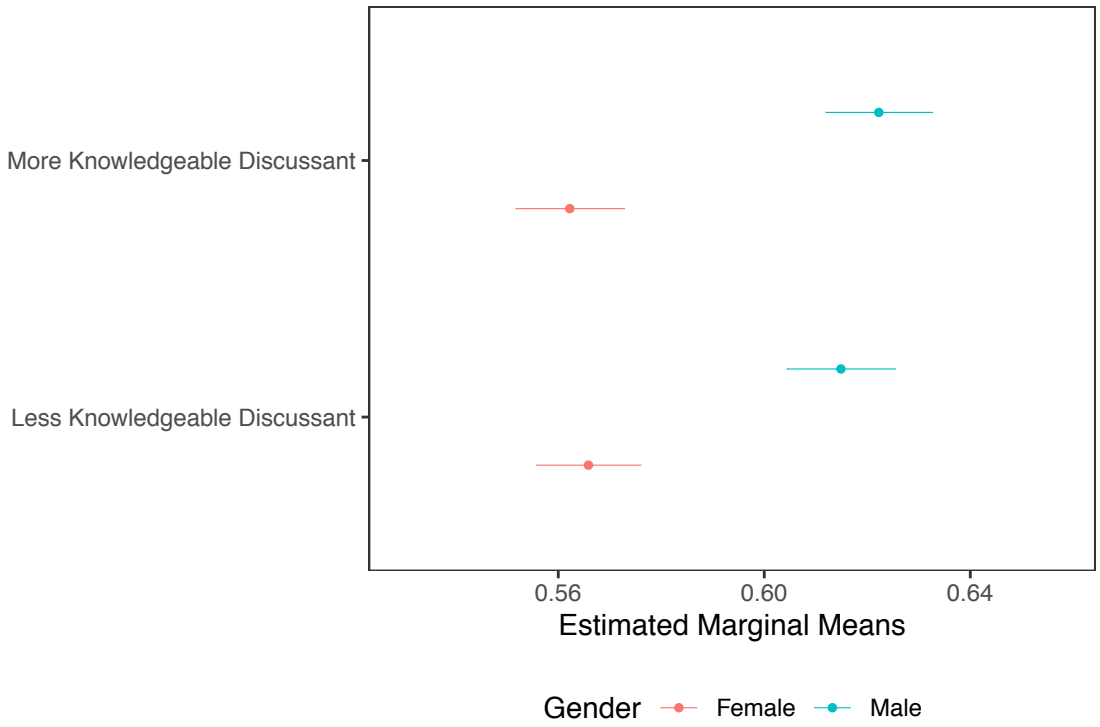

Figure F.4. Estimated marginal means for discussant knowledge level (more knowledgeable, less knowledgeable) at different levels of respondent gender (male, female). Horizontal lines represent 95 percent confidence intervals. Consistent with Figure D.3, women reported being less likely to report their true opinions than men for both more and less knowledgeable discussants.

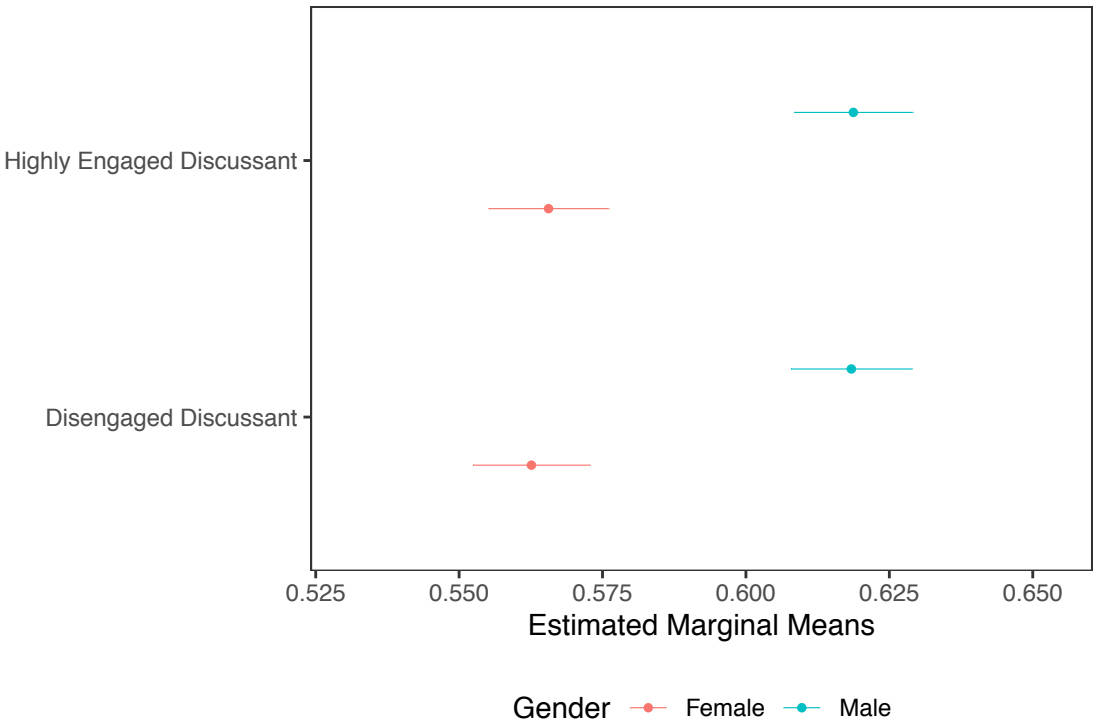

Figure F.5. Estimated marginal means for discussant political engagement (highly engaged, not engaged) at different levels of respondent gender (male, female). Horizontal lines represent 95 percent confidence intervals. Consistent with Figure D.4, women reported being less likely to report their true opinions than men for both more and less engaged discussants.

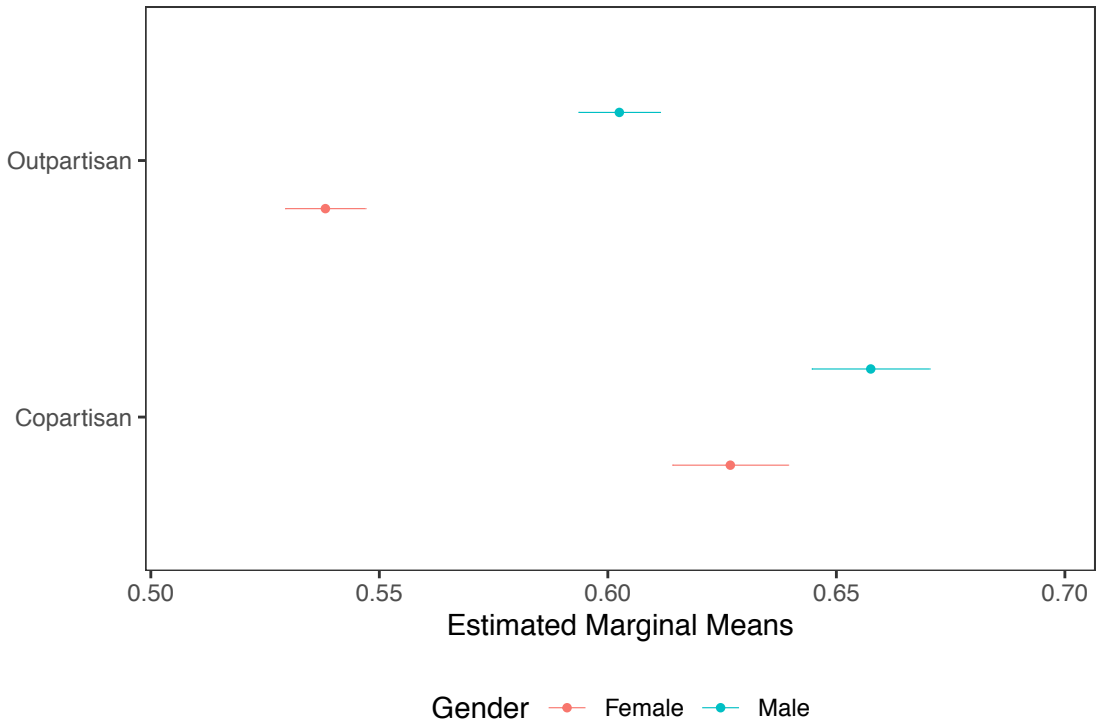

Figure F.6. Estimated marginal means for copartisan and outpartisan discussants at different levels of respondent gender (male, female). Horizontal lines represent 95 percent confidence intervals. Consistent with Figure D.5, both men and women reported being more likely to express their true opinions to copartisans than to outpartisans, but this gap was much larger for women than it was for men.

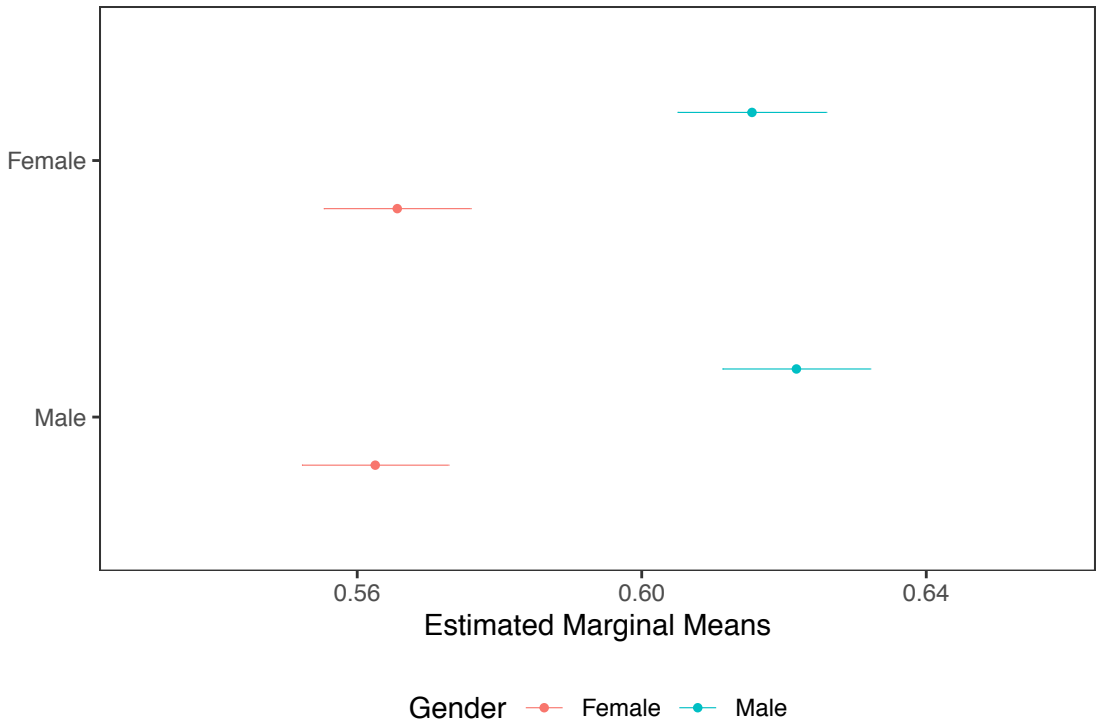

Figure F.7. Estimated marginal means for discussant gender (male, female) at different levels of respondent gender (male, female). Horizontal lines represent 95 percent confidence intervals. Consistent with Figure D.6, women reported being less likely to report their true opinions than men for both male and female discussants. Women were equally likely to express their true opinions to female as male discussants; men were equally likely to express their true opinions to female as male discussants.

G. Ethics Statement

This study was approved by the [INSTITUTION REDACTED FOR ANONYMITY] Institutional Review Board under exempt status (ID #202105136) on June 3, 2021.

H. Exploratory Analysis

Through the peer-review process, an insightful reviewer requested to see our main results broken down by the partisanship of the respondent. These results are presented in Figure H.1 below.

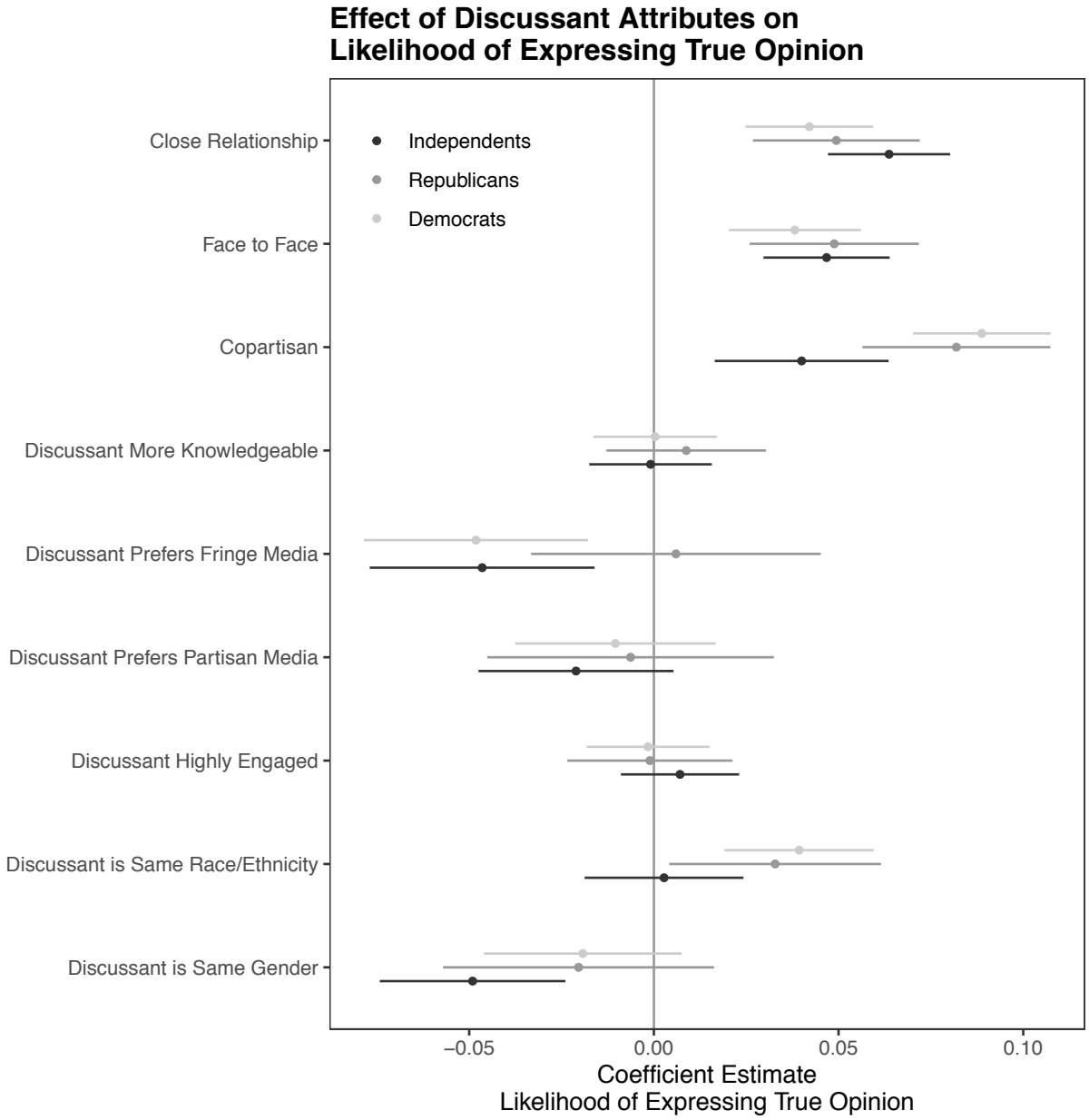

Figure H.1. Effect of discussant attributes on likelihood of expressing true opinion, by respondent Party ID.

Note: Coefficients estimated using a linear model with robust standard errors clustered at the respondent level. Bars reflect 95 percent confidence intervals