CESifo WORKING PAPERS

7946
2019

November 2019

Self-Persuasion: Evidence from Field Experiments at Two International Debating Competitions

Peter Schwardmann, Egon Tripodi, Joël J. van der Weele

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7946*
Category 13: Behavioural Economics

Self-Persuasion: Evidence from Field Experiments at Two International Debating Competitions

Abstract

Does the wish to convince others lead people to persuade themselves about the moral and factual superiority of their position? We investigate this question in the context of two international debating competitions, where persuasion goals (pro or contra a motion) are randomly assigned to debaters shortly before the debate. Using incentives for truthful reporting, we find evidence of self-persuasion in the form of (i) factual beliefs that become more conveniently aligned with the debater's side of the motion, (ii) shifts in attitudes, reflected in an increased willingness to donate to goal-aligned charities, and (iii) higher confidence in the strength of one's position in the debate. Self-persuasion occurs before the debate and subsequent participation in the open exchange of arguments does not lead to convergence in beliefs and attitudes. Our results lend support to interactionist accounts of cognition and suggest that the desire to persuade is an important driver of opinion formation and political partisanship.

Peter Schwardmann
University of Munich
Department of Economics
Ludwigstraße 28/09
Germany – 80539 Munich
peter.schwardmann@econ.lmu.de

Egon Tripodi
European University Institute
Department of Economics
Via delle Fontanelle 18
Italy – 50014 San Domenico di Fiesole
egon.ripodi@eui.eu

Joël J. van der Weele
University of Amsterdam
Department of Economics & the
Tinbergen Institute / The Netherlands
vdweele@uva.nl

The authors thank Michele Belot, Davide Cantoni, Josh Dean, Marvin Deversi, Florian Englmaier, Lorenz Goette, David K. Levine, George Loewenstein, Andrej Woerner and audiences at the Advances with Field Experiments conference in Chicago, the EDP Jamboree in Louvain-la-neuve, the CESifo Behavioral Economics conference in Munich, and seminar audience at the University of Bonn for useful comments. Research funding from the CRC TRR 190 Rationality and Competition, the Research Priority Area Behavioral Economics at the University of Amsterdam, the Dutch Science Foundation (NWO) and the European University Institute is gratefully acknowledged. This project would have not been possible without the generous collaboration of the conveners of the two debating tournaments, Yimin Ge and Lena Martinovic, and the support and advice of Huyen Nguyen. We thank Hannah Rehwinkel and Irene van Rooij for excellent research assistance, and we are grateful to numerous local field assistants and members of the hosting debating societies for their support with data collection. The research hypothesis and most of the analyses were pre-registered on the AEA RCT Registry (AEARCTR-0003922). Appendix L links the pre-registration material to the final paper.

1. Introduction

How people form beliefs has been the subject of longstanding inquiry in the social sciences. Standard economic theory posits that agents interpret new evidence by using Bayes' rule, in a process of truth approximation. The behavioral approach, often associated with Kahneman and Tversky's research program, proposes that people are boundedly rational and use heuristics in their attempts to discover the truth in complex information environments, leading to systematic mistakes. In an influential set of recent papers, researchers from different backgrounds have criticized these approaches for neglecting the fundamentally social nature of human reasoning and belief formation, which originates in the need to impress and persuade others (Mercier and Sperber, 2011; Von Hippel and Trivers, 2011; Kurzban, 2012; Mercier, 2016; Simler and Hanson, 2017).

The alternative, interactionist approach maintains that our reasoning processes have developed to convince others of our position, acting like a private "press secretary" (Kurzban, 2012). In the process of persuasion, we align our beliefs and convictions with our economic and political goals, often at the expense of truth or accuracy. This view organizes a range of cognitive phenomena such as confirmation bias, overconfidence and motivated reasoning (Mercier and Sperber, 2011). It naturally explains why salespeople develop optimistic beliefs about their product, like mortgage brokers' unwarranted confidence in the U.S. housing market during the financial crisis of 2007-8, especially among those working on the sell-side (Cheng et al., 2015). It also explains why politicians, who are professional persuaders, often have opportunistic and malleable convictions and display more polarization than the general population (Fiorina and Abrams, 2008). But despite the abundance of applications, it is hard to test the interactionist account empirically, as the direction of causality between private views and the wish or need to persuade others is often unclear.

In this paper, we confront this identification challenge and deliver a direct test of the interactionist approach in a field setting. We investigate the causal effect of persuasion goals on the formation of beliefs and attitudes, a phenomenon we call "self-persuasion". Our study takes place at two international debating competitions in Munich and Rotterdam. These tournaments draw members from debating clubs from all

over Europe, who, across several rounds, debate motions on topical issues. In this context, we elicit beliefs and attitudes surrounding the debated motions in each of the qualifying rounds of the tournament, both before and after the debates. To make sure that our elicitations reflect true beliefs and attitudes, we incentivize reports with an incentive compatible scoring rule.

Several features of debating tournaments make them ideally suited for testing the interactionist approach. First, debaters are randomly assigned to pro or contra positions of a motion shortly before the start of the debate. This allows us to make causal inferences about the effect of persuasion goals. The nature of the randomization also solves two problems that may arise in the identification of self-persuasion. Because the assignment is randomized explicitly, participants know not to infer anything about the merit of the assigned debating position. Moreover, since the randomization is a natural aspect of the tournament, participants do not view it as experimental variation, ameliorating concerns of potential experimenter demand effects. Another unique aspect of our setting is that debaters' intrinsic motivation to be persuasive is high. A panel of experienced judges evaluates the quality of each debater's arguments, determining his or her success in the tournament and subsequent status in the debating community. These incentives for persuasion mimic those of professionals in politics and law. It is no coincidence that many famous politicians and lawyers honed their skills by taking part in competitive debating.[1]

We find strong evidence for self-persuasion, measured as a gap in beliefs and attitudes between debaters arguing against and those arguing in favor of the motion, shortly after the assignment of persuasion goals. First, participants are more likely to believe that a factual statement is true if the statement strengthens an argument supporting their position. Second, in a monetary allocation task between charities, debaters shift donations towards a goal-aligned charity. Third, debaters become more confident about the strength of the arguments on their side of the motion, as measured by the estimated probability that other teams on the same side of the motion will win

[1]For instance, prominent Brexiteers Boris Johnson and Michael Gove were president of the Oxford Union, a renowned debate club. Other prominent politicians who were part of debating societies include Nancy Pelosi, Jimmy Carter, Margaret Thatcher and John Major. See either the site of the National Speach and Debate Association or this site for partial lists of famous former debaters.

their debates. Beliefs elicited before the assignment of persuasion goals confirm that there are no pre-treatment difference between the two groups.

We also investigate whether the debate itself mitigates the effect of self-persuasion by exposing participants to arguments from the other side. We do not find evidence for convergence, as polarization in factual beliefs and attitudes after the conclusion of the debate is similar to that at the start. As a result, debaters leave the tournament more polarized than they started. Since debaters are never asked the same question twice, the persistence of polarization is not driven by concerns for consistency. In our setting at least, self-persuasion causes the exchange of ideas to be a catalyst of polarization rather than an antidote to it.

Our paper provides the first field evidence for the idea that persuasion goals drive non-Bayesian belief and attitude formation. This lends support to an interactionist account of human cognition (Von Hippel and Trivers, 2011; Mercier and Sperber, 2011; Simler and Hanson, 2017) and is in line with recent evidence from the laboratory.[2] Our data also allow us to comment on the mechanism underlying self-persuasion. Mercier and Sperber (2011) argue that self-persuasion is a by-product of persuasion, resulting from a cognitive failure to account for our disproportionate investment in finding the strengths in our own and the weaknesses in our interlocutor's position. Instead, Von Hippel and Trivers (2011) theorize that self-persuasion or self-deception about the moral and factual superiority of one's position is a requirement for successful persuasion. To investigate these channels, we ask debaters how many arguments they generated for each position during their preparation time. We find that these are highly skewed towards their own position, and that the imbalance can explain about half of the treatment effect. Thus, our data suggest an important role of a naive appreciation of one's own biased arguments, as well as of other mental processes likely driven by self-deception.

Our findings connect with several strands of the literature. They have immediate relevance for the literature on motivated reasoning. A large literature in social psy-

[2]Several studies, using a different and narrower set of outcome variables, show that people manage their beliefs strategically in order to better convince others (Smith et al., 2017; Schwardmann and van der Weele, 2019; Solda et al., 2019).

chology and a growing literature in economics has looked at how psychological and functional goals influence belief formation processes (Kunda, 1990; Bénabou and Tirole, 2016; Gino et al., 2016). Among other things, this literature has argued that people manipulate their own beliefs in order to maintain satisfaction with past choices (Akerlof and Dickens, 1982; Goetzmann and Peles, 1997), be better bargainers (Babcock et al., 1995), raise their moral (self)image (Dana et al., 2007; Exley, 2015; Di Tella et al., 2015; Grossman and van der Weele, 2017), and motivate themselves to give their best (Compte and Postlewaite, 2004; Bénabou and Tirole, 2002). To our knowledge, we provide the first field evidence for motivated cognition, using incentivized procedures. Moreover, we demonstrate that the wish to persuade others is a powerful functional motive for belief distortion. It seems at least plausible that the persuasion motive plays a role in bargaining, self-image maintenance and self-motivation, potentially providing a unifying principle in thinking about motivated reasoning.

Our paper relates to a nascent theoretical literature that extends standard Bayesian belief updating to include the role of social interactions in belief formation. These theories formalize mechanisms through which identification with social groups (Gennaioli and Tabellini, 2019) and the production of narratives to interpret historical data (Eliaz and Spiegler, 2018) or influence the behavior of others (Bénabou et al., 2019) can lead individuals to distort their views and cause polarization. Our results broadly lend support to the common view, expressed in these models, that social interactions and persuasive communication are an important driver of belief distortion.

There is also an immediate connection of our results with the empirical literature on polarization and political opinion formation. Researchers across the social sciences have used laboratory experiments to show how confirmation bias and selective parsing of arguments can lead to attitude polarization (Lord et al., 1979; Sunstein, 2002). Several different mechanisms have been proposed to fit these data (Taber and Lodge, 2006; Kahan, 2015; Fryer et al., 2018). We add field data showing that the persuasion motive induces polarization on a range of cognitive and non-cognitive measures, suggesting that a number of different mental processes are at work. Furthermore, following the literature in experimental economics, we use incentivized procedures for truthful reporting to make sure that elicited beliefs and attitudes are sincerely held

(Schlag et al., 2015). This is crucial, as Bullock et al. (2013) show that voters display up to 80 percent less polarized attitudes when their answers are incentivized for accuracy.

Relatedly, our analysis of competitive debating contributes to a discussion about the merits of deliberative democracy. According to the ideal of deliberative democracy the exchange of opinions helps to resolve conflicts and foster social consensus (e.g. Habermas, 1984; Elster, 1998; Gutmann and Thompson, 2004). By contrast, the literature on polarization has shown that deliberation can have exactly the opposite effect (Kuhn et al., 1997), and promote radicalization in interactions between like-minded people (Sunstein, 2002). The conditions for deliberation to work best are a matter of active debate in political science (e.g. Thompson, 2008; Mercier and Landemore, 2012). We find that the *prospect* of debate increases polarization and that the subsequent debating does little to decrease it.

The remainder of the paper is structured as follows. Section 2 describes setting, sample, and procedures of the experiment. Section 3 presents results on the effects of persuasion goals on privately held views, and illustrates how debating can affect polarization. Section 4 provides evidence to inform a discussion on the psychological mechanisms of self-persuasion and the relation between self-persuasion and debater success. Section 5 concludes by discussing some implications of our results.

2. Experimental Setting

Competitive debating is a popular activity. In the US alone, the National Speech and Debate Association has enrolled about 2 million members since 1925. Many universities have debating societies that organize local or international tournaments, the most prestigious of which include the North American, European and World Championships. Contestants tend to be university students, and motions relate to topical issues in politics such as immigration, climate change and the regulation of new technology. In contrast to debates between experts or politicians, competitive debaters are randomly assigned to defend particular positions, which may or may not correspond to their private opinions.

Our study took place at two international debating competitions in March 2019: the

Munich Research Open, and the *Erasmus Rotterdam Open*. Both tournaments followed the British Parliamentary (BP) debating format, in which debates take place with two teams of two debaters arguing in favor of (Proposition) and two teams against (Opposition) a given motion. Persuasion goals (Proposition/Opposition) are randomly assigned to teams and all speakers have equal time to present their arguments. The motions are prepared by chief adjudicators before the tournament, and revealed to the debating teams fifteen minutes ahead of the debate. They are designed such that there are valid arguments for both sides. Debaters are evaluated on the quality of their arguments by a panel of three expert judges, who themselves have experience as debaters.

The competitions featured 52 (Munich) and 48 (Rotterdam) teams and took place in two phases. In the preliminary phase of the tournament (*in-rounds*), all teams debate multiple times: each round features a motion that all teams debate in parallel sessions. In each round, teams are partitioned into 13 (Munich) or 12 (Rotterdam) parallel debating sessions of four teams each using a conditional random assignment. Teams accumulate points that depend on their evaluation and determine who advances to the knock-out phase of the competition. Appendix A provides further details on the BP debating format.

2.1. Sample

Participants of international debating competitions in the BP format are predominantly undergraduate and graduate students, who are members of debating societies. They accumulate debating experience through tournament participation and regular meetings at the debating societies of their university, and sometimes also from a high-school debating career. The characteristics of BP debating attracts speakers with strong analytical skills, fast thinking and a breadth of knowledge.[3]

On average, our sample has spent more than two years in debating, has qualified for more than three semi-finals of an international tournament, is about 22 years old, and tends to hold a relatively liberal ideology. Men are somewhat over-represented and

[3]Further discussion of the characteristics of debaters that take part in this format on the website of the American Parliamentary Debate Association.

the sample is very international – less than 25 percent of participants hold nationality from the country where the tournament is hosted. The sample is similar across the two tournaments in terms of age, local representation, political views, and time spent in debating. However, there are some differences in terms of the gender balance and past achievements: the share of female debaters is 17 percentage points higher in Munich than in Rotterdam, and debaters in Rotterdam have reached semi-finals in large international competitions more than twice as many times than debaters in Munich.

More importantly for the internal validity of our findings, in Table D.2 we show balance of individual characteristics and baseline views on topics related to each motion across debaters with different persuasion goals. For some of the questions we randomized the order across subgroups. In Table D.3 we show that individual characteristics are balanced also across these subgroups.

2.2. Research Design

We only collected data during the preliminary rounds of the competitions (five in Munich and four in Rotterdam) to maintain a balanced panel of observations. Debaters answered four main surveys with the following timing:

1. **Baseline.** Administered at the very beginning of the tournament. Contains background questions as well as instructions on the quadratic scoring rule (QSR) – the procedure that we use throughout all surveys to elicit beliefs in an incentive compatible manner.

2. **Predebate.** Administered right after the preparation time of each debating session, just before the debate begins.

3. **Postdebate.** Administered right after each debate ends.

4. **Endline.** Administered after the fifth and last debate of the preliminary phase (Munich) or after the fourth round of the preliminary phase (Rotterdam).[4]

[4]This difference is due to different schedules of the tournaments. In both cases, the endline survey took place after the last round of a four-round day. In Rotterdam, the tournament started in the morning and had a full day with four rounds of debate. In Munich, the tournament started in the late afternoon with one round of debate and had four rounds of debate the day after.

Our main survey measures are the following:

- **Factual beliefs.** These were factual statements that related to the motion, and debaters had to predict whether the statements were true or false. Factual statements were constructed such that, if they were true, one side of the debate would find them "convenient" in support of their arguments. We elicit Factual beliefs related to the motions at Baseline, Predebate, and Postdebate.

- **Attitudes:** We asked debaters to allocate money between a "neutral" charity and a charity that was aligned with one side of the motion. Each charity was described to respondents in a short paragraph on the same survey sheet. We elicit Attitude related to the motions at Predebate, and Postdebate.

- **Confidence in proposition:** We elicited the subjective probability that a majority of parallel debates (excluding the debater's own debate) in the round will be won by the proposition side of the debate. This is a measure of the perceived advantage of a persuasion goal, independent of a speaker's confidence in her own ability. We elicit Confidence in proposition only at Predebate.

Next we provide an example of a motion and an associated factual statement, charity and confidence question from the surveys. Appendix B provides detailed examples of factual belief elicitations from motions in our debates.

Example of motion: When tech companies own platform utilities and platform products, this House would break them up.

Factual statement: According to a 2018 survey from the Pew Research Center, over 60 percent of Americans believe that major tech companies should be more regulated than they currently are.

Charity: The Open Markets Institute (OMI). OMI uses journalism to promote greater awareness of the political and economic dangers of monopolization, identifies the changes in policy and law that cleared the way for such consolidation and foster discussions with policymakers and citizens as to how to update America's traditional political economic principles for our 21st century digital society.

Confidence statement: Excluding the debate happening in this room, in at least half of the parallel debates of this round, one of the two teams on the Government side of this motion will rank 1st.

We incentivized our main outcome variables as follows. For the Factual beliefs and the Confidence elicitation, subjects were incentivized with a binarized quadratic scoring rule that paid in lottery tickets. By providing a report $r \in [0, 100]$, given the objective binary answer $R \in \{0, 1\}$, a subject receives a lottery ticket that paid off a monetary prize of 30 euros with the following winning probability

$$w = 1 - \left(R - \frac{r}{100}\right)^2.$$

Of all elicitations of this kind, only one was randomly selected to be paid at the end of the study. Our general instructions used both the mathematical equation, a simple quantitative illustration, and an intuitive explanation that incentives were designed so that the truthful reporting optimizes the likelihood of winning the prize of 30 euro (see Appendix I). [5]

For the Attitude variable, subjects allocated up to 10 euro between two different charities, where the budget constraint was concave in order to discourage extreme choices. One of the choices was randomly selected and the experimenters made the charitable payments on the subjects' behalf.

In addition to these incentivized measures, we elicited some background variables, including gender, debating experience and performance, as well as some basic socio-demographics.[6] In our Endline survey, we also asked several questions on "impressions", for example, about factual statements and the goal of the research. These variables served to check the robustness of our main results. Table 1 summarizes how survey elements were distributed across the different surveys.

[5] In theory, this procedure makes the quadratic scoring rule incentive compatible for all risk preferences (Hossain and Okui, 2013; Schlag and Van der Weele, 2013). Whether this is actually the case in practice is a matter of ongoing debate.

[6] The Baseline survey also included some incentivized factual knowledge "decoy" questions about topics not related to the motions. These questions served to obfuscate the elicitation of Factual Beliefs related to the motions and not give away the topics of the motions that were still secret at that point.

Table 1: Timing and Content of Debater Surveys

			Incentivized Outcome Variables			
Survey	Timing	Background Info	*Factual beliefs*	*Attitudes (charities)*	*Confidence in proposition*	Impressions
Baseline	Beginning of tournament	X	X			
Predebate	Right before each debate		X	X	X	
Postdebate	After each debate		X	X		
Endline	After last debate					X

Debates were moderated by a panel composed of three (sometimes two) judges. These were experienced debaters themselves trained to evaluate debaters' speeches according to standardized international criteria. After the debate, judges deliberated in private to produce the "ballot", an official score sheet that consists of the technical score on the quality of arguments made by each debater in each debate and determines the ranking of teams in each debate. In addition, we asked judges to independently fill out a "judge survey" where they assign a broad persuasiveness score to each debater. We told judges that this score should consider quality of arguments as well as body language, tone, and other markers that make a speech persuasive to a general population.

The four debater surveys as well as the judge survey were administered by an enumerator, who also attended the debate and filled out a separate "enumerator survey" that was designed to capture both objective and subjective measures of how heated debates were, and whether facts and charities included in the survey questions were mentioned by debaters to make their case. Enumerators were asked to take note of any anomaly that might have occurred during the debate.

The full content of all surveys is described in detail in Appendix I. Appendix J provides all motions, survey questions and charities used for the attitude elicitation.

2.3. Survey Versions and Administration Procedures

Before each tournament, we interacted with the chief adjudicators to converge on a final set of motions for the debate. For each motion, we developed four factual ques-

tions (A, B, C, D) and found two motion-related charities (E, F). We varied the order in which factual questions and charities were presented between two different subgroups, as illustrated in Table 2. We created these subgroups in advance using lists of registered participants and identified a debater's subgroup by adding an ID number to their name tag.

The use of multiple questions in different orders assures that no debater answers the same question twice and that no result depends on the answer to a single question or the order in which questions were asked. It also eliminates the desire to provide consistent answers to repeated questions and reduces potential experimenter demand effects. Moreover, since baseline and predebate questions were different both within and across subgroups, participants could not be influenced through discussion of the answers with others.

The baseline survey was administered in a large common room after some introductory remarks by the organizers and one of the researchers. In this room, debaters were given 10 minutes to read carefully a set of general instructions for the surveys, and subsequently had 25 minutes to answer the baseline survey. The survey is similar for all participants except for the factual questions that directly relate to the in-rounds motions, which differed between subgroups as displayed in Table 2.

Table 2: Distribution of Factual Questions and Charities Over Surveys

	Motion factual questions			Motion charities	
	Baseline	Predebate	Postdebate	Predebate	Postdebate
Subgroup 1	A	D	B, C	E	F
Subgroup 2	B	C	A, D	F	E

Note: Distribution of four factual questions per motion and two motion-related charities over surveys. Each letter corresponds to one factual question/charity.

In each debating round, the motions were announced in the central meeting room, and debaters made their way to the assigned debating room after announcements. Enumerators distributed the predebate survey in the separate debating rooms. While seated at their desks, debaters were given up to five minutes to answer and enumera-

tors ensured that they did not use this time to prepare for the debate. At the beginning of the debate enumerators also distributed the judge survey, in which judges indicated their evaluations of persuasiveness. Judges had the entire debate session plus their regular judge deliberation time to fill out this survey.

After the predebate survey, the judges opened the debate. During the debate itself, which lasts about an hour, enumerators filled in their own surveys, noting down participant IDs and debate impressions. After the judges declared the end of the debate, enumerators distributed the postdebate survey, which debaters had five minutes to answer.

The endline survey was administered just outside of each debate room right after the end of the last round of debates covered by our intervention. Debaters had twenty minutes to answer this survey, which they did in the corridors outside the debating room. Enumerators insisted with subjects to not interact with others or mobile devices during this time.

3. Results

Our main focus lies on the question of how persuasion goals affect self-persuasion, as measured by our predebate elicitations on Factual Beliefs, Attitudes and Confidence. A secondary question relates to the impact of the debate itself on polarization among debaters.

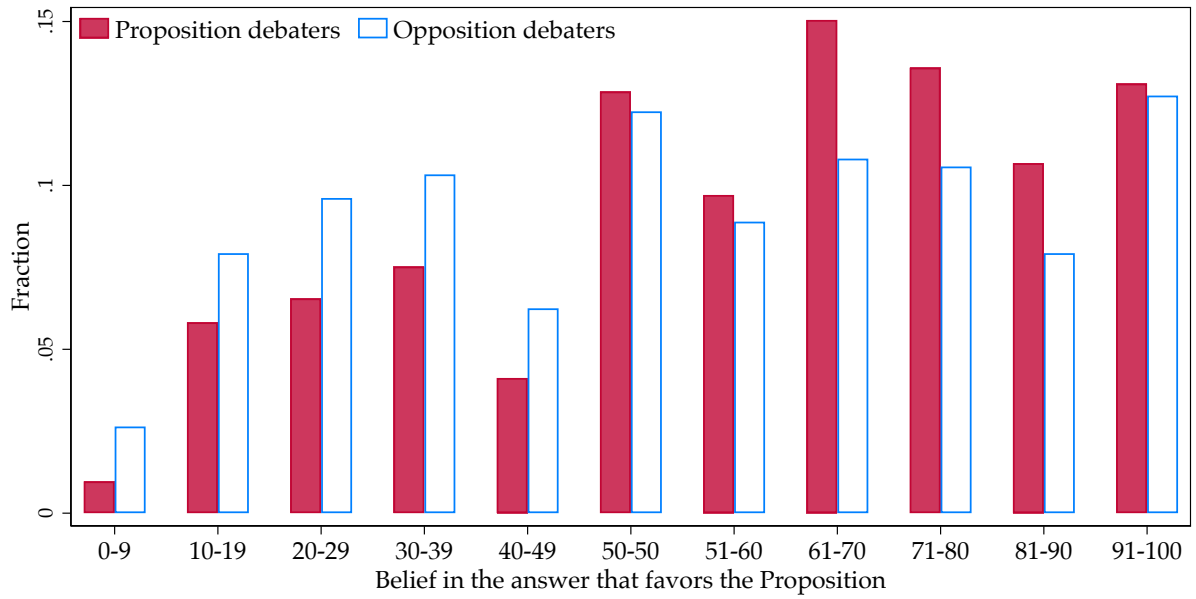
3.1. Self-Persuasion

We compare differences in our main outcome variables, i.e. Factual Beliefs, Attitudes and Confidence, between debaters on the proposition and on the opposition side of the motion. We look at the predebate elicitations, which reflect only the cognitive processes taking place in the 15 minute preparation period after persuasion goals are assigned, and are not affected by the actual debating itself. In Appendix G, we assess the extent to which debaters were able to infer our research hypotheses, and use this as input to a robustness analysis of our self-persuasion results to experimenter demand effects.

3.1.1. Do Persuasion Goals Affect Factual Beliefs?

For every factual belief question, one state (e.g. the statement is true) is more favorable to the proposition of the motion than the other state (e.g. the statement is not true). In order to compare questions, we transform each belief into the subjective probability that the state that favors the proposition is true. When a factual statement is favorable to the proposition (opposition), this corresponds to the reported subjective probability that the statement is true for speakers on the proposition (opposition) side of the debate, and to the complementary probability for speakers on the opposition (proposition) side. More background information on which states are considered favorable to the proposition is provided in Appendix B.

Figure 1: Factual Beliefs, by Persuasion Goal



Note: Predebate beliefs elicited from debaters over multiple rounds are pooled and each report $r \in [0, 100]$ is transformed as the complement to 100 if the report is not aligned with the proposition. The pooled and transformed beliefs are then grouped in equally spaced probability brackets – except for the intermediate 50-50 category.

Figure 1 reports beliefs that are grouped in equally spaced probability brackets, except for the intermediate 50-50 category. These data show that debaters are more likely to believe in the answer that favors the proposition, if they themselves are in the propo-

sition.[7]

To assess the statistical significance and the magnitude of this effect, and gain greater comparability of subjective probabilities on the truthfulness of different factual statements, we conduct both a normal standardization of the reported belief (separately for each question) and adjust the sign of the standardized belief. In turn, a positive (negative) sign of such standardized outcome captures alignment with the state that favors the proposition (opposition). After adjusting the sign, the standardized belief remains normally distributed with zero mean and unit standard deviation. This transformation yields an individual level outcome variable $b_{i,m}$ that admits a straightforward interpretation in terms of debater i's belief alignment with the proposition of motion m.

We estimate the gap in belief alignment with the proposition in a regression model

$$b_{i,m} = \alpha_i + \beta Proposition_{i,m} + \delta_m + \varepsilon_{i,m} \tag{3.1}$$

in which we include motion fixed effects δ_m and debater fixed effects α_i and allow for the error term to be correlated within each team of debaters.

Table 3 shows the results of the estimation. We confirm the finding that proposition debaters report beliefs that are markedly different from the beliefs reported by opposition debaters. Because of the randomized allocation of persuasion goals, this pattern cannot be explained by pre-existing differences between debaters on the two sides of the debate and has a causal interpretation. Factual Beliefs of proposition debaters are 21.5 percent of a standard deviation (column 1, $p < 0.001$) closer to the proposition alignment. This effect is robust to the omission of fixed effects (column 2) and the inclusion of controls (column 3).

[7] Note that on both sides of the debate, debaters are more likely to believe that the answer favors the proposition. This is partly driven by the correct answer being aligned with the proposition relatively more frequently.

Table 3: Panel Regressions for Effects of Persuasion Goals on Factual Beliefs

	Beliefs align with proposition		
	(1)	(2)	(3)
Debater in proposition	0.215***	0.217***	0.203***
	(0.062)	(0.061)	(0.062)
Socio-demographic and experience controls			✓
Debater fixed effects	✓		
Round FEs	✓	✓	✓
Observations	884	884	851

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

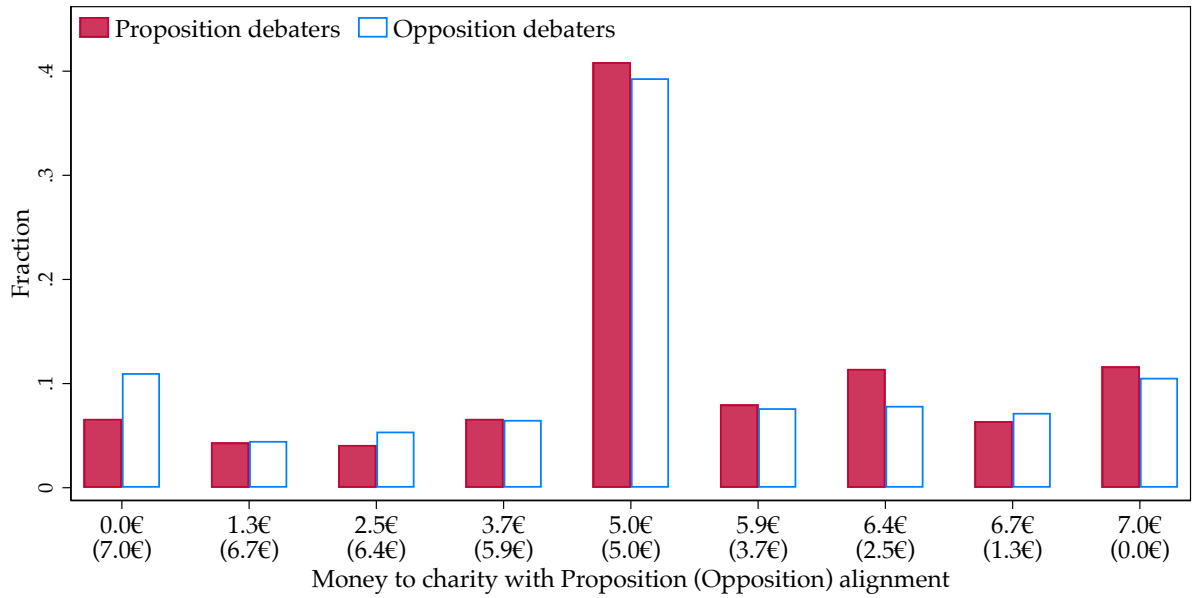
Notes: Standard errors in parentheses are clustered at the team level. Socio-demographic controls include age, gender, and an indicator for whether the debater's nationality is from the country that hosts the competition. Experience controls include the reported number of international tournaments in which the debater has made it to semi-finals, and a categorical variable capturing the number of years the debater has been actively debating. Some observations are lost in column (3) due to missing control variables.

Result 1 (Factual Self-Persuasion)**.** *Persuasion goals make individuals distort their perception of facts in the direction that strengthens the positions they need to defend.*

3.1.2. Do Persuasion Goals Affect Attitudes?

We measure attitudes towards the persuasion goal by how much money the debater allocates to a charitable cause that is aligned to her persuasion goal relative to a neutral charity. Remember that allocations lie on a concave budget constraint to encourage choices in the interior of the donation space.

Figure 2: Chosen Donation Bundles by Persuasion Goal



Note: Predebate allocations of charitable donations over multiple rounds are pooled and each allocation $a \in \{0, .., 8\}$ is transformed as the complement to 8 if the allocation does not favor the charity with relative proposition alignment.

Figure 2 depicts donation choices across all motions. Allocations on the right side favor the charity aligned with the proposition and choices on the left side favor the charity aligned with the opposition. About 40 percent of allocation choices feature an equal split. Among the remaining observations we see a tendency for debaters to favor charities that are aligned with their persuasion goal.

To estimate the size and statistical significance of the effect, we use a fixed effects regression framework similar to model 3.1, in which the ordinal outcome capturing how favorable the debater's allocation is to the proposition charity is treated as a continuous variable.[8] We complement this analysis with regressions that use as continuous outcomes directly the monetary amounts donated to proposition and opposition charities implied by the bundle chosen by the debater.

[8]The more appropriate regression model would take into account the discrete ordinal nature of the outcome variable. However, ordered log-odds estimated from ordered Logit models are very hard to interpret. We provide panel estimates of the ordered Logit model in Table D.4. These are qualitatively very similar and support the main analysis presented here.

Table 4: Panel Regressions for Effect of Persuasion Goals on Attitudes

	Donation bundle favorable to Proposition charity			Money to charity in Proposition	Money to charity in Opposition
	(1)	(2)	(3)	(4)	(5)
Debater in proposition	0.306**	0.297**	0.300**	0.316***	-0.239*
	(0.132)	(0.136)	(0.145)	(0.122)	(0.124)
Socio-demographic and experience controls			✓		
Debater fixed effects	✓			✓	✓
Round FEs	✓	✓	✓	✓	✓
Observations	883	883	850	883	883

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Standard errors in parentheses are clustered at the team level. Socio-demographic controls include age, gender, and an indicator for whether the debater's nationality is from the country that hosts the competition. Experience controls include the reported number of international tournaments in which the debater has made it to semi-finals, and a categorical variable capturing the number of years the debater has been actively debating. Some observations are lost in column (3) due to missing control variables.

Table 4 presents the results of the estimation. We confirm the impressions from visual inspection of the pooled outcomes: persuasion goals lead proposition debaters to choose an allocation of charitable donations that is 0.306 positions more favorable to the charity with proposition alignment (column 1, $p = 0.023$). Columns (4) and (5) aid the interpretation of this point estimate: From a total concave budget to allocate between two charities that can range from 7 to 10 euro, proposition debaters tend to sacrifice 0.239 euro that could go to the charity with opposition alignment to give 0.316 euro more to the charity with proposition alignment. The asymmetry of the transfer is largely due to the frequency of extreme aligned allocations among opposition debaters.

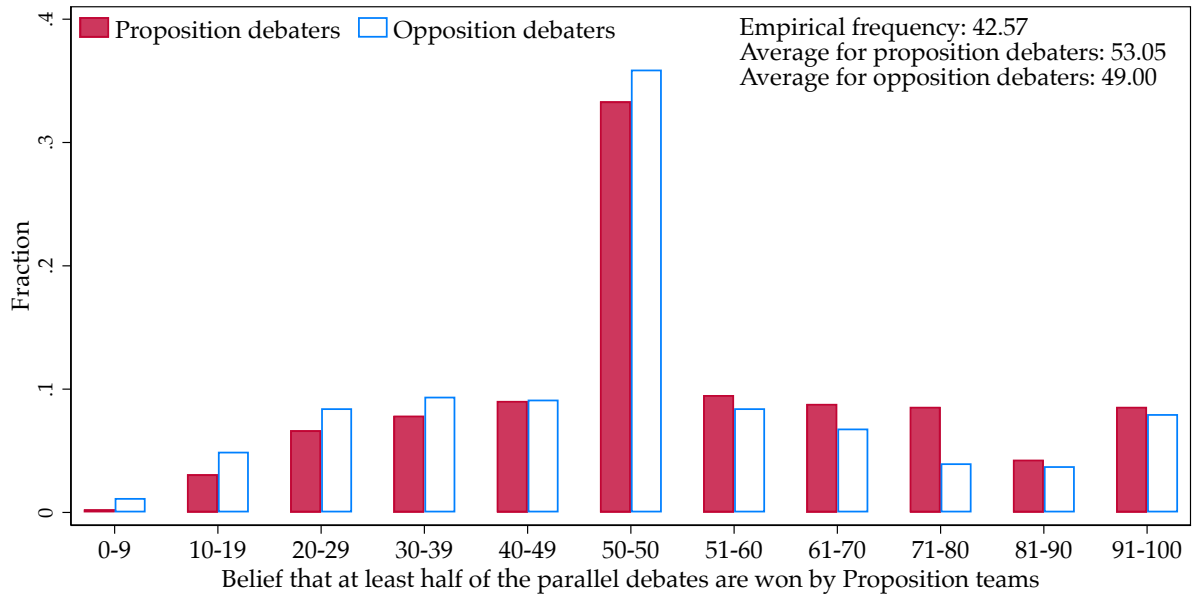
Result 2 (Moral Self-Persuasion)**.** *Individuals favor social causes aligned with their persuasion goals.*

3.1.3. Do Persuasion Goals Affect Confidence in One's Side of the Debate?

Our third outcome measure is debaters' Confidence in the strength of the proposition side of the debate. This is reported by debaters as the probabilistic prediction that at least half of the parallel debates will be won by proposition teams. Importantly, since debaters are not betting on the outcome of the parallel debates and not on their own performance, this belief reflects the perceived strength of the debating position abstracting from beliefs in their own ability.

Figure 3 depicts probabilistic beliefs that the proposition will win in more than half of the parallel sessions, grouped by equally spaced probability brackets – except for the intermediate 50-50 category. Beliefs are polarized across the two sides of the debate: 38 percent of the beliefs reported by proposition debaters lie above 50 percent, while only 30 percent of opposition debaters state beliefs higher than 50 percent.

Figure 3: Perceived Advantage of the Proposition, by Persuasion Goal



Note: Predebate Confidence in the proposition, measured as the probability that at least half of the parallel debates are won by proposition teams, reported from debaters over multiple rounds are pooled. The pooled confidence reports are then grouped in equally spaced probability brackets – except for the intermediate 50-50 category.

When it comes to the empirical distribution, the proposition team wins the majority of parallel debates in each round only 43 percent of the time. Debaters' average

probabilistic beliefs in this event are 49 percent in the opposition and 53 percent in the proposition. Hence, all debaters tend to overestimate the chances of proposition teams in these debates, but debaters in the proposition exhibit a greater bias.

To estimate the effects of persuasion goals on the perceived strength of the proposition, we can directly use the raw belief data on Confidence in the proposition as outcome in a fixed effects regression framework similar to equation (3.1). The results of this analysis are reported in Table 5. Debaters in proposition teams are significantly more likely to believe that proposition teams will win the majority of debates. The reported probability assigned to the event that the majority of parallel debates will be won by proposition teams is higher by about 4.5 percentage points (column 1, $p < 0.005$) for debaters who propose the motion relative to those who oppose it. This estimated effect is also about 20 percent of a standard deviation in the outcome – a similar magnitude to the self-persuasion effects on factual beliefs reported in the previous section, and also remarkably similar to estimates in Schwardmann and van der Weele (2019).

Table 5: Panel Regressions for Effects of Persuasion Goals on Confidence

	Confidence in proposition teams		
	(1)	(2)	(3)
Debater in proposition	4.531***	4.389***	4.319***
	(1.498)	(1.492)	(1.554)
Socio-demographic and experience controls			✓
Debater fixed effects	✓		
Round FEs	✓	✓	✓
Observations	883	883	850

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Standard errors in parentheses are clustered at the team level. Socio-demographic controls include age, gender, and an indicator for whether the debater's nationality is from the country that hosts the competition. Experience controls include the reported number of international tournaments in which the debater has made it to semi-finals, and a categorical variable capturing the number of years the debater has been actively debating.

Result 3 (Confidence)**.** *Persuasion goals make individuals relatively more confident about the strength of the positions they defend.*

3.2. Debates and the Dynamics of Polarization

The power of debate to moderate differences of opinion is at the core of the ideal of deliberative democracy. In this section we assess whether the debate reduces the polarization caused by self-persuasion. To investigate this, we compare the beliefs and attitudes at the start of the debate, as measured in the predebate survey with those at the end, as expressed in the postdebate survey. As a measure of dispersion we use the sample variance σ^2 in beliefs and attitudes. To track disagreement both within and between the proposition and opposition sides, we decompose this variance in *between* group and *within* group variation. In particular, σ^2 can be written as the weighted average of Mean Squares Between groups (MSB) and Mean Squares Within groups (MSW) as follows[9]

$$\sigma^2 = \frac{k-1}{n} MSB + \frac{n-k}{n} MSW,$$

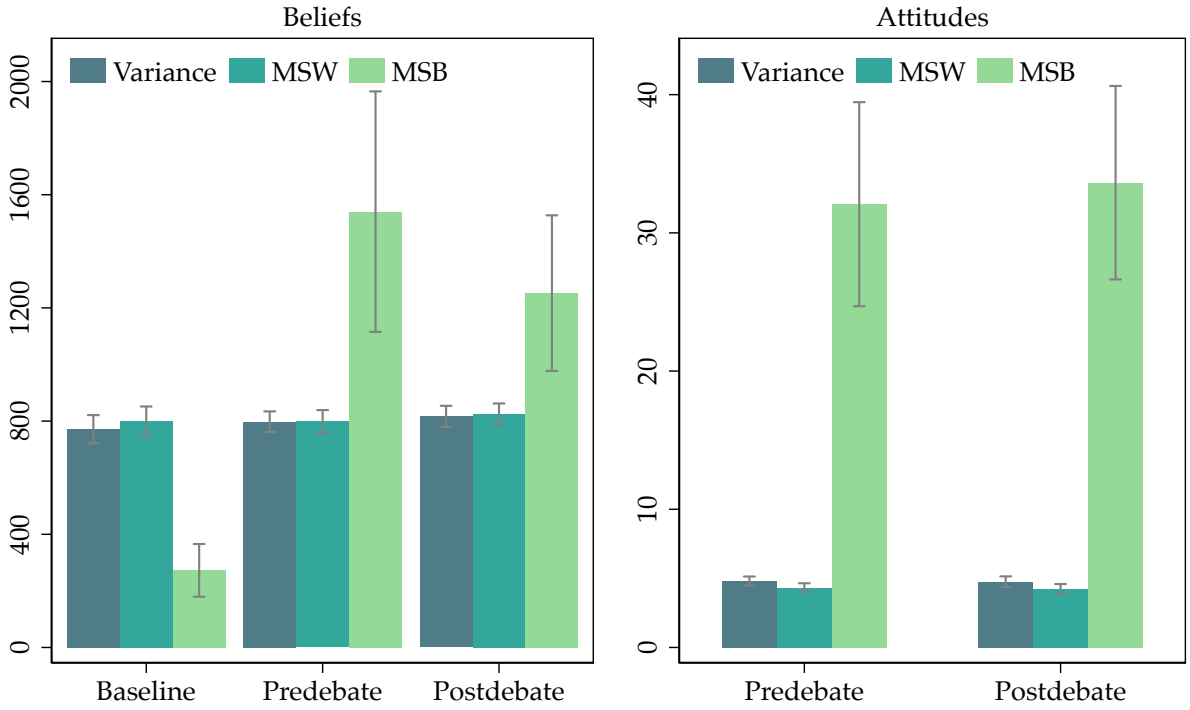
where n is the sample size and k is the number of groups. For each Factual Belief and Attitude elicitation have two subgroups ($k = 2$) and a sample of about $n = 50$ observations (this represents half of the participants in each tournament, as we randomized the order of elicitations between two subgroups). We have two questions and two charities for each of the nine different motions, leading to 18 observations of within and between group polarization for each variable. This allows us to statistically compare the distributions of Total variance (σ^2), MSB and MSW across different stages of the debate.

Figure 4 shows the resulting statistics. The comparison of pre- and postdebate, shows that the MSB for Beliefs decreases slightly (by 0.12 of a standard deviation), but not significantly so (*Mann-Whitney* test $H_0 : MSB_{Pre} = MSB_{Post}$, $p = 1.000$). When it comes to Attitudes, polarization actually increases slightly (by 0.05 of a standard

[9]Using the well known decomposition of the Total Sum of Squares in the sum of Between Sum of Squares (BSS) and Within Sum of Squares (WSS), and the definition of mean squares as the sum of squares statistics over their degrees of freedom ($MSB := BSS/(k-1)$, and $MSW := WSS/(n-k)$).

deviation), but again without statistical significance. To check whether our measure are capable of picking up changes in polarization documented in the previous subsection, we also include the polarization in Factual Belief at baseline. This comparison shows that the MSB for factual beliefs increases significantly from baseline to predebate (*Mann-Whitney* test $H_0 : MSB_{Base} = MSB_{Pre}$, $p = 0.023$). This shows that the MSB measure captures the polarizing effects of self-persuasion. Moreover, it also increases between baseline and postdebate by 0.57 of a standard deviation (*Mann-Whitney* test $H_0 : MSB_{Base} = MSB_{Post}$, $p = 0.031$), showing that the overall debating experience leads to an increase in polarization.[10]

Figure 4: Variance Decomposition of Beliefs and Attitudes



Note: For each elicitation of factual beliefs and attitudes from an identical question that debaters answer in the same survey we have a sample of about 50 responses from both proposition and opposition debaters. Over both tournaments we have 18 belief questions elicited at baseline and postdebate, 18 belief questions elicited at predebate and postdebate, and 18 allocations of donations between different charities elicited at predebate and postdebate. Ranges indicate standard errors.

To assess the robustness of these findings, we consider two other prominent ap-

[10]Figure D.1 and Figure D.2 dissect the evolution of disagreement between debaters question by question, and demonstrate that polarization occurs on a broad range of issues.

proaches in the literature in Appendix C. Mimicking the variance decomposition, (Desmet et al., 2017)'s measure of cultural distance increases significantly from baseline to postdebate, and is reduced slightly from predebate to postdebate–although not significantly so. The polarization index by Duclos et al. (2004) shows that the polarization of factual beliefs appears stable through the three elicitations. This index however does not perform too well with survey responses that have a high mass of reports at focal points (e.g. for factual beliefs these are 0, 50, and 100). Distributions with (more than one) artificially strong modes are spuriously identified as substantially polarized, making relatively small changes in actual polarization hard to detect.

Finally, we investigate whether the dynamics of polarization are related to emotions during the debate. There is some research in political science showing that incivility during debates may lead people to take opposing views less seriously (Mutz, 2007). To get a sense of how emotional the debaters were during the debate, the enumerators recorded both subjective measures of the "heatedness" of a debate, and the number of interruptions during the debate. The analysis in Appendix F shows that debaters whose baseline beliefs are aligned with their persuasion goals also give more heated speeches, but greater heat in a debate does not moderate the convergence of views (see Appendix C).

Result 4. *We find no evidence that debates lead to convergence of attitudes and beliefs among debaters.*

4. Mechanisms and Consequences of Self-Persuasion

We now discuss several secondary research questions. First, we delve deeper into the psychological mechanisms behind self-persuasion. We then discuss the relation between self-persuasion and debating success.

4.1. Psychological Mechanisms of Self-Persuasion

What psychological mechanisms underlie the self-persuasion documented in the previous section? The randomization of persuasion goals across debaters allows us to

rule out some explanations suggested in the previous literature, such as the priming of political affiliations (e.g. Petersen et al., 2013) or confirmation bias (e.g. Fryer et al., 2018). Subjects also had very little opportunity to acquire new information, and thus engage in selective search (Taber and Lodge, 2006). Furthermore, debaters are unlikely to actively think about research hypotheses and bias their responses accordingly, since the randomization is such a natural part of the tournament. In Appendix G we do more analyses to rule out experimenter demand effects.

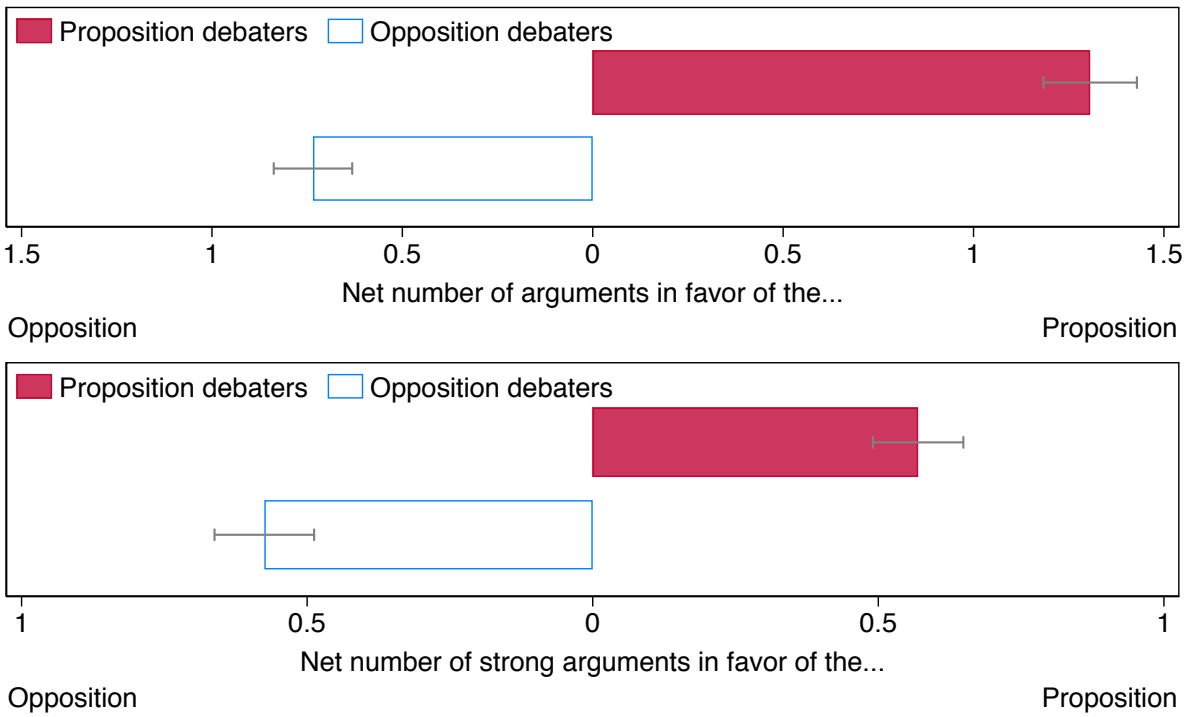
Two remaining and plausible mechanisms by which social interactions cause bias in beliefs and attitudes are self-deception and bounded rationality. Self-deception refers to a process of motivated reasoning in which subjects "choose" their beliefs. In this account, put forward in Von Hippel and Trivers (2011), self-persuasion is a subconscious strategy aimed at increasing persuasiveness. It does so by reducing nervousness, give-away tells or other manifestations of doubt or cognitive dissonance arising from a discrepancy between one's persuasion goals and true beliefs. This theory has received support in recent laboratory studies (Smith et al., 2017; Schwardmann and van der Weele, 2019; Solda et al., 2019).

Alternatively, bounded rationality or cognitive heuristics may drive the impact of persuasion goals. According to Mercier and Sperber (2011), persuasion objectives lead us to generate arguments disproportionately in favor of our own view. Thus, in the process of preparing for the debate, debaters may naturally gather more arguments for their position than against it. They may take this asymmetry as evidence for the strength of their position because they fail to take account of their biased selection of arguments. Such "selection neglect" has been documented in multiple studies (Juslin et al., 2007; Barron et al., 2019). Related ideas underpin the notions of "availability bias" (Tversky and Kahneman, 1973) or "persuasive argument theory" (Vinokur and Burstein, 1974), which maintain that the number, novelty or salience of arguments drive belief formation.

Selection neglect implies that if debaters generate more arguments on their own side of the debate, this asymmetry mediates self-persuasion. To test this, we asked debaters in the predebate survey for the number of arguments they came up with during their preparation time, both for and against the motion. We also asked them how many of

these arguments they considered to be "very strong". Figure 5 shows the average net number of arguments debaters came up with on both sides by treatment. As is clear from the graph, debaters engage in asymmetric selection of arguments. On average, they come up with one additional argument and one half of a "strong" argument in favor of their own side.

Figure 5: Differences in the Number of Arguments



Note: Ranges indicate standard errors.

To address the impact of this asymmetry, we conduct a parametric causal mediation analysis (Imai et al., 2010b) - see Appendix H for details. This analysis shows the extent to which self-persuasion is mediated through s_i, the number of aligned arguments as a fraction of total arguments considered during preparation time. In Table 6 we show the results of this analysis for our three main outcome variables. The analysis reveals that the share of aligned arguments drives between 29 percent and 57 percent of the self-persuasion effect. The fraction is largest for Confidence and smallest for Factual Beliefs.

These results suggest that selection neglect plays an important role in self-persuasion, but that mechanisms of self-deception are about equally, if not more, important. Note

that these statements are surrounded by substantial uncertainty: on the one hand, we cannot rule out that selection neglect is itself (partially) driven by self-serving motives (Exley and Kessler, 2018), leading to a possible overestimation of the importance of the heuristic explanation. On the other hand, our measures of the number of arguments are likely to be affected by measurement error, leading to a potential underestimation.[11] The constraints of our field experiment do not allow for a deeper investigation, which we therefore leave for further research.

Table 6: Decomposition of Treatment Effect in Mediated and Direct Effect

	Beliefs	Attitudes	Confidence
Average causal effect mediated by s_i (ACME)	0.058	0.158	2.340
	(0.045)	(0.075)	(1.131)
Average direct effect (ADE)	0.143	0.129	1.714
	(0.075)	(0.156)	(1.854)
Average treatment effect (ATE)	0.201	0.287	4.110
	(0.066)	(0.137)	(1.558)
ACME/ATE	0.289	0.551	0.569

Note: Estimates obtained following the procedure outlined in Appendix D of Imai et al. (2010a): we estimate the Linear Structural Equation Model using random effects regressions with the full set of controls as in Section 3.1, and we use the estimated sampling distributions to draw 100 simulations of potential mediators and potential outcomes. We average the differences of potential outcomes across the 100 simulations to obtain an estimate of the mediated effect. We repeat the procedure 1000 times from bootstrap samples to obtain standard errors of the estimates.

4.2. Self-Persuasion and Debating Success

We now turn to the relation between self-persuasion and success in the debating competition. This relation is of interest for two reasons. First, it can inform our view of the the psychological mechanisms underlying self-persuasion that we discussed above.

[11] See also Appendix H, where we discuss (i) the *sequential ignorability* assumption needed to identify causal mediation effects, and (ii) measurement error potentially attenuating the estimates of these effects (le Cessie et al., 2012).

A negative relation with debating success is consistent with an explanation of self-persuasion in terms of cognitive errors. By contrast, a positive relation is in line with strategic self-deception, where cognition is optimized for persuasiveness. Second, the success of self-persuasion in the context of a debating competition may tell us something about its prevalence in broader contexts. If self-persuasion is detrimental to persuasiveness, it would be less likely to constitute a widely observed phenomenon. However, if self-persuasion is not detrimental to persuasiveness, we might expect it to be common, even for people, such as politicians, whose professional success relies on persuasion.

Unfortunately, our dataset is not ideally suited to look at the causal effect of self-persuasion. The ideal experiment would create exogenous variation in self-persuasion. However, this would require changing debating objectives and procedures, which was not possible at such high profile competitions. Nevertheless, correlations may give us a valuable input for future research. Moreover, we can exploit the alignment of factual beliefs at baseline, which is random, to look at the effect of belief alignment on persuasiveness.

Is self-persuasion more prevalent among successful debaters? If successful debaters are more likely to engage in self-persuasion, we should expect a positive interaction effect between debater success and self-persuasion. To look at this, we add an interaction term to the regression model 3.1, used to study self-persuasion on all our three outcomes. Debater success is measured by "achievements" – the number of semi-finals reached by debaters in international tournaments–elicited in the baseline survey before treatment. Table 7 presents the results of such estimation. In each regression, we control for debating experience by including the number of years a debater has been active.

Table 7: Panel Regressions for Heterogeneous Effects of Persuasion Goals

	Factual Beliefs		Attitudes		Confidence	
	(1)	(2)	(3)	(4)	(5)	(6)
Debater in proposition	0.203***	0.229***	0.300**	0.211	4.319***	2.784*
	(0.062)	(0.070)	(0.145)	(0.167)	(1.554)	(1.640)
Debater in proposition × Achievements		-0.007		0.024		0.419*
		(0.011)		(0.033)		(0.255)
Socio-demographic and experience controls	✓	✓	✓	✓	✓	✓
Round FEs	✓	✓	✓	✓	✓	✓
Observations	851	851	850	850	850	850

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$

Notes: Standard errors in parentheses are clustered at the team level. Socio-demographic controls include age, gender, and an indicator for whether the debater's nationality is from the country that hosts the competition. Experience controls include the number of years the debater has been actively debating.

The results in column 1 indicate that self-persuasion on factual beliefs is not related to success in past tournaments: more and less successful debaters engage in self-persuasion to a similar extent. Though not (highly) significant, we find higher estimates for the interaction term for attitudes (column 4, $p = 0.471$) and confidence (column 6, $p = 0.100$). For debaters who have never made it to the semi-finals of an international tournament we estimate that for these variables the self-persuasion effect is 30 and 35 percent smaller, respectively.

Does belief and attitude alignment help persuasiveness? We analyze whether judges' evaluations of debaters' persuasiveness correlate with the alignment of debaters with their persuasion goal. We have four measures of a debater's alignment with the persuasion goal: Factual Belief alignment at baseline, Factual Belief alignment at predebate, Attitude alignment at predebate, and Confidence in Proposition at predebate. Note that only the first of these measures counts as exogenous variation, as it was measured before the treatment was administered. As measures of persuasiveness in the tournament we have both a broad persuasiveness score provided by each judge independently, as well as a technical score of the quality of debater's arguments

that is given by judges in agreement after the debate is over.

Table 8 presents correlations between our measures of alignment and persuasiveness across all rounds of debate. None of our alignment measures is a significant predictor of persuasiveness. One explanation for this null result is that measurement error attenuates the relations between the variables. In fact, while alignment with the persuasion goal may be partially or wholly captured using Factual Beliefs, Attitudes, and Confidence, actual debater's alignment remains a latent variable. In addition, the low inter-rater agreement between judges (*Cohen's Kappa* $= 0.083$) on the broad persuasiveness of each debater also raises concerns regarding the overall quality of judges' unincentivized responses.[12]

Table 8: Pearson's Correlation Between Persuasion Outcomes and Alignment Variables

	Broad persuasiveness	Quality of arguments
	(1)	(2)
Baseline belief alignment	-0.006	0.035
	(0.859)	(0.302)
Predebate belief alignment	-0.019	0.025
	(0.572)	(0.451)
Predebate attitude alignment	0.181	0.041
	(0.590)	(0.228)
Predebate confidence in own position	0.006	0.019
	(0.851)	(0.571)
Observations	883	883

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: P-value for statistical significance in parentheses. Broad persuasiveness is evaluated by each judge on a panel independently; but we average the individual scores at the debater-round level. Alignment variables transform our main raw outcomes as in Section 3.1, and change the sign of these outcomes for opposition debaters to obtain variables that become larger (smaller) as the debater exhibits greater (less) alignment with their persuasion goal.

[12]The *Cohen's Kappa* coefficient ranges between 0 (expected level of agreement that can be obtained by chance) and 1 (perfect agreement).

In summary, although we find slightly more positive than negative point estimates, there are only weak correlations between debater success and the alignment of their attitudes and beliefs with their persuasion goal. The available variation in our dataset does not allow definite conclusions about the relation between self-persuasion and debater success. This remains an important area for future research.

5. Conclusion

Our data show that people will distort their factual beliefs, attitudes and confidence when they have to argue for their position. Debaters whose persuasion goals are randomly varied will believe in "alternative facts", despite incentives for accuracy and exposure to opposing views. We call this effect *self-persuasion*, as debaters convince themselves of their position before they even begin to persuade others. Our findings lend support to theories that reserve a fundamental role for social influence and persuasion in the development and operation of our cognitive capacities (Von Hippel and Trivers, 2011; Mercier and Sperber, 2011). These results obtain in a field setting, in a sample that is a regular supplier of future elites and politicians. We find no evidence that self-persuasion is detrimental to success, and hence no reason to suspect that it disappears with experience.

While our dataset does not allow investigation of long-term dynamics, we conjecture that persuasion goals help explain instances of polarization in a range of contexts where debate and persuasion play a role. For instance, self-persuasion offers a reason why polarization is more severe in the US congress than it is in the American public (Fiorina and Abrams, 2008), why it is so strong on social media platforms, especially if people are exposed to opposing views (Bail et al., 2018), and why people who joined the Republican party exclusively for their view on abortion then saw their other beliefs fall in line with the party (Gould and Klor, 2019). It also suggests alternative motives for political behavior than are commonly assumed. For instance, canvassing and proselytizing activity may be important not just to grow the base, but also for deepening the convictions of existing followers. Similarly, opportunistic political U-turns or flip-flops may be the cause of genuine conversion in the process of defending the new

position.

In the field of behavioral economics and social psychology, self-persuasion has the potential to unify phenomena that cannot be explained by Bayesian updating and are currently being studied separately. It helps explain why people engage in various self-enhancement strategies and become overconfident about their abilities (Trivers, 2011; Schwardmann and van der Weele, 2019), why they are more eager to confirm than disconfirm their views (Nickerson, 1998; Benjamin, 2019), why they look for exculpatory narratives and exploit wiggle room in moral dilemmas (Dana et al., 2007; Exley, 2015; Di Tella et al., 2015), and why they appear conveniently unaware of their darker motives (Kurzban, 2012; Simler and Hanson, 2017).

Further research is necessary to test the explanatory power of self-persuasion and the interactionist approach across domains. Our findings raise expectations that such a research program will lead to substantial revisions in the standard view of human cognition, a view eloquently expressed by John Maynard Keynes when accused of inconsistency: "When the facts change, I change my mind. What do you do Sir?". For many people the answer appears to be "the reverse".

References

Akerlof, George A. and William T. Dickens, "The Economic Consequences of Cognitive Dissonance," *The American Economic Review*, 1982, *72* (3), 307–319.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin F. Camerer, "Biased judgments of fairness in bargaining," *The American Economic Review*, 1995, *85* (5), 1337–1343.

Baetschmann, Gregori, Kevin E Staub, and Rainer Winkelmann, "Consistent estimation of the fixed effects ordered logit model," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2015, *178* (3), 685–703.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences*, 2018, *115* (37), 9216–9221.

Barron, Kai, Steffen Huck, and Philippe Jehiel, "Everyday econometricians: Selection neglect and overoptimism when learning from others," Technical Report, WZB Discussion Paper 2019.

Bénabou, Roland and Jean Tirole, "Self-Confidence and Personal Motivation," *The Quarterly Journal of Economics*, 2002, *117* (3), 871–915.

Bénabou, Roland and Jean Tirole, "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, September 2016, *30* (3), 141–164.

Bénabou, Roland, Armin Falk, and Jean Tirole, "Narratives, Imperatives and Moral Reasoning," 2019.

Benjamin, Daniel J, "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics-Foundations and Applications 2*, 2019, p. 69.

Bullock, John G, Alan S Gerber, Seth J Hill, and Gregory A Huber, "Partisan bias in factual beliefs about politics," Technical Report, National Bureau of Economic Research 2013.

Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 1980, *47* (1), 225–238.

Cheng, Ing-Haw, Sahil Raina, and Wei Xiong, "Wall Street and the Housing Bubble: Bad Incentives, Bad Models, or Bad Luck?," *American Economic Review*, 2015, *104* (9), 2797–2829.

Compte, Olivier and Andrew Postlewaite, "Confidence-enhanced performance," *American Economic Review*, 2004, *94* (5), 1536–1557.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang, "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 2007, *33* (1), 67–80.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg, "Culture, ethnicity, and diversity," *American Economic Review*, 2017, *107* (9), 2479–2513.

Di Tella, Rafael, Ricardo Pérez-Truglia, Andres Babino, and Mariano Sigman, "Conveniently Upset: Avoiding Altruism by Distorting Beliefs About Others," *American Economic Review*, 2015, *105* (11), 3416–3442.

Duclos, Jean-Yves, Joan Esteban, and Debraj Ray, "Polarization: Concepts, Measurement, Estimation," *Econometrica*, November 2004, *72* (6), 1737–1772.

Eliaz, Kfir and Rani Spiegler, "A Model of Competing Narratives," 2018.

Elster, Jon, *Deliberative Democracy* Cambridge Studies in the Theory of Democracy, Cambridge University Press, 1998.

Exley, Christine and Judd B Kessler, "Motivated Errors," 2018.

Exley, Christine L, "Excusing selfishness in charitable giving: The role of risk," *The Review of Economic Studies*, 2015, *83* (2), 587–628.

Fiorina, Morris P. and Samuel J. Abrams, "Political polarization in the American public," *Annual Revue of Political Science*, 2008, *11*, 563–588.

Fryer, Roland G., Philipp Harms, and Matthew O. Jackson, "Updating beliefs when evidence is open to interpretation: Implications for bias and polarization," *Journal of the European Economic Association*, 2018.

Gennaioli, Nicola and Guido Tabellini, "Identity, Beliefs, and Political Conflict," 2019.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber, "Motivated Bayesians: Feeling moral while acting egoistically," *Journal of Economic Perspectives*, 2016, *30* (3), 189–212.

Goetzmann, William N. and Nadav Peles, "Cognitive dissonance and mutual fund investors," *Journal of financial Research*, 1997, *20* (2), 145–158.

Gould, Eric D. and Esteban F. Klor, "Party hacks and true believers: The effect of party affiliation on political preferences," *Journal of Comparative Economics*, 2019.

Grossman, Zachary and Joël J. van der Weele, "Self-Image and Willful Ignorance in Social Decisions," *Journal of the European Economic Association*, 2017, *15* (1), 173–217.

Gutmann, Amy and Dennis Thompson, *Why Deliberative Democracy?*, student edition ed., Princeton University Press, 2004.

Habermas, Jürgen, *The theory of communicative action*, Vol. 1, Beacon Press, 1984.

Harrison, Glenn W, Jimmy Martínez-Correa, and J Todd Swarthout, "Eliciting subjective probabilities with binary lotteries," *Journal of Economic Behavior & Organization*, 2014, *101*, 128–140.

Hippel, William Von and Robert Trivers, "The evolution and psychology of self-deception," *Behavioral and Brain Sciences*, 2011, *34* (1), 1–16.

Hossain, Tanjim and Ryo Okui, "The binarized scoring rule," *Review of Economic Studies*, 2013, *80* (3), 984–1001.

Imai, Kosuke, Luke Keele, and Dustin Tingley, "A general approach to causal mediation analysis.," *Psychological methods*, 2010, *15* (4), 309.

_ , _ , and Teppei Yamamoto, "Identification, inference and sensitivity analysis for causal mediation effects," *Statistical science*, 2010, pp. 51–71.

Juslin, Peter, Anders Winman, and Patrik Hansson, "The naive intuitive statistician: a naive sampling model of intuitive confidence intervals.," *Psychological review*, 2007, *114* (3), 678.

Kahan, Dan, "The Politically Motivated Reasoning Paradigm," SSRN Scholarly Paper ID 2703011, Social Science Research Network, Rochester, NY December 2015.

Kuhn, Deanna, Victoria Shaw, and Mark Felton, "Effects of Dyadic Interaction on Argumentive Reasoning," *Cognition and Instruction*, 1997, *15* (3), 287–315.

Kunda, Ziva, "The case for motivated reasoning.," *Psychological bulletin*, 1990, *108* (3), 480.

Kurzban, Robert, *Why everyone (else) is a hypocrite: Evolution and the modular mind*, Princeton University Press, 2012.

le Cessie, Saskia, Jan Debeij, Frits R. Rosendaal, Suzanne C. Cannegieter, and Jan P. Vandenbroucke, "Quantification of bias in direct effects estimates due to different types of measurement error in the mediator," *Epidemiology*, 2012, pp. 551–560.

Lord, Charles G, Lee Ross, and Mark R. Lepper, "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.," *Journal of Personality and Social Psychology*, 1979, *37* (11), 2098.

Mercier, Hugo, "The argumentative theory: Predictions and empirical evidence," *Trends in Cognitive Sciences*, 2016, *20* (9), 689–700.

_ **and Dan Sperber**, "Why do humans reason? Arguments for an argumentative theory," *Behavioral and brain sciences*, 2011, *34* (2), 57–74.

_ **and Hélène Landemore**, "Reasoning is for arguing: Understanding the successes and failures of deliberation," *Political Psychology*, 2012, *33* (2), 243–258.

Mutz, Diana C., "Effects of "In-Your-Face" Television Discourse on Perceptions of a Legitimate Opposition," *American Political Science Review*, November 2007, *101* (4), 621–635.

Nickerson, Raymond S., "Confirmation bias: A ubiquitous phenomenon in many guises," *Review of General Psychology*, 1998, *2* (2), 175–220.

Petersen, Michael Bang, Martin Skov, Søren Serritzlew, and Thomas Ramsøy, "Motivated reasoning and political parties: Evidence for increased processing in the face of party cues," *Political Behavior*, 2013, *35* (4), 831–854.

Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth, "Measuring and bounding experimenter demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

Schlag, Karl H. and Joël J. Van der Weele, "Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming Risk Neutrality," *Theoretical Economics Letters*, 2013, *3* (1), 38–42.

Schlag, Karl H, James Tremewan, and Joël J Van der Weele, "A penny for your thoughts: A survey of methods for eliciting beliefs," *Experimental Economics*, 2015, *18* (3), 457–490.

Schwardmann, Peter and Joël J. van der Weele, "Deception and Self-Deception," *Nature Human Behavior*, 2019, *3*, 1055–1061.

Simler, Kevin and Robin Hanson, *The elephant in the brain: Hidden motives in everyday life*, Oxford University Press, 2017.

Smith, Megan K., Robert Trivers, and William von Hippel, "Self-deception facilitates interpersonal persuasion," *Journal of Economic Psychology*, December 2017, *63*, 93–101.

Solda, Alice, Changxia Ke, Lionel Page, and William von Hippel, "Strategically delusional," Technical Report, QUT Business School 2019.

Sunstein, Cass R., "The Law of Group Polarization," *Journal of Political Philosophy*, 2002, *10* (2), 175–195.

Taber, Charles S. and Milton Lodge, "Motivated Skepticism in the Evaluation of Political Beliefs," *American Journal of Political Science*, 2006, *50* (3), 755–769.

Thompson, Dennis F., "Deliberative democratic theory and empirical political science," *Annu. Rev. Polit. Sci.*, 2008, *11*, 497–520.

Trivers, Robert, *The folly of fools: The logic of deceit and self-deception in human life*, Basic Books (AZ), 2011.

Tversky, Amos and Daniel Kahneman, "Availability: A heuristic for judging frequency and probability," *Cognitive Psychology*, 1973, *5* (2), 207–232.

Vinokur, Amiram and Eugene Burstein, "Effects of partially shared persuasive arguments on group-induced shifts: A group-problem-solving approach," *Journal of Personality and Social Psychology*, 1974, *29* (3), 305–315.

A. British Parliamentary debating

Debates can take place in various formats. The most popular format, that features in the most prestigious tournaments (e.g. the World University Debating Championship), is the British Parliamentary (BP). For such format, debaters take part in debates in teams and each team is composed of two debaters. A debate is characterized by a motion, four teams of debaters, and a panel of experienced judges. Debates begin with the announcement of the motion that two teams, on the proposition (also called Government) side of the House have, to defend and two teams, of the opposition side of the House, have to contrast. BP debating exclusively feature *impromptu* debates, in which motions are revealed only 15 minutes ahead of debates and teams are randomly assigned to argue either in favor or against the given motion. Finally, while the order of teams speaking in each debate is also random, it is each team's choice to determine which team member speaks first. All speakers are given 7 minutes to present their arguments following a precise structure that we illustrate in Table A.1.

Table A.1: Debaters' Responsibilities by Role

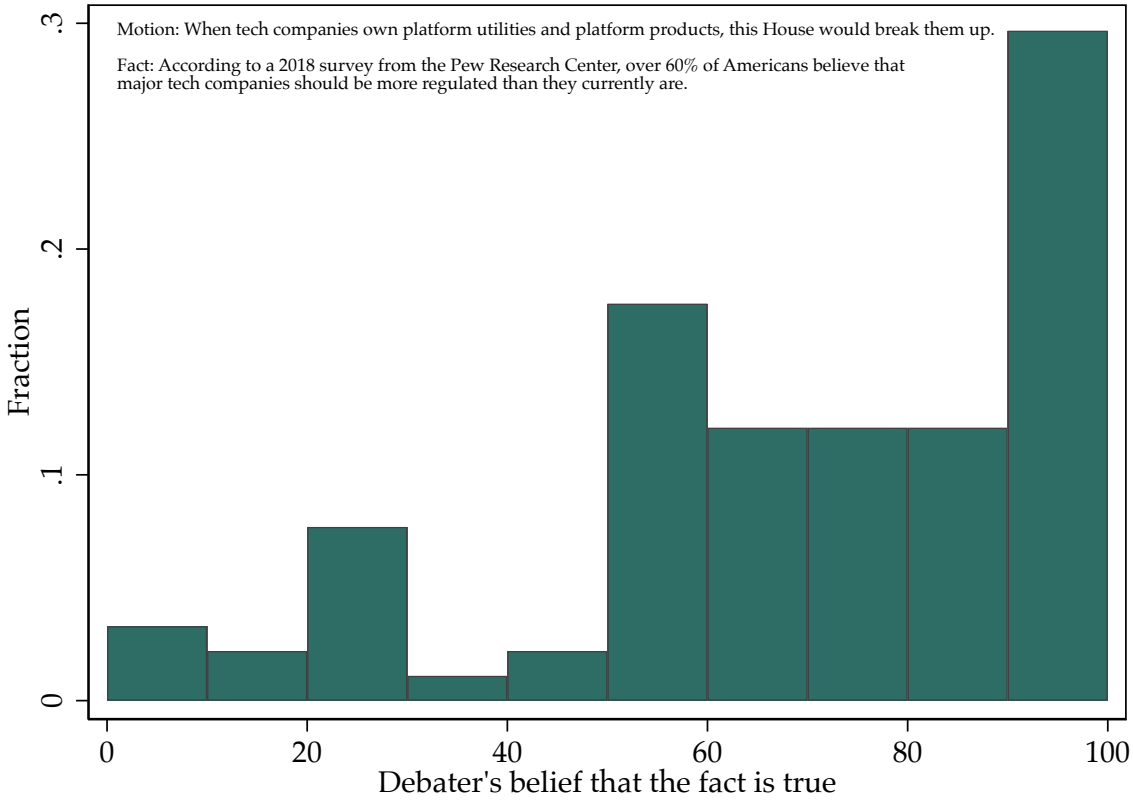
Team	Speaking role	Speaking order	Team	Speaking role	Speaking order
Opening Government (OG)	Prime Minister (PM)	First	Opening Opposition (OO)	Leader of the Opposition (LO)	Second
	• Defines and interprets the motion • Develops the case for the proposition			• Accepts definition of the motion • Refutes the case of OG • Constructs arguments against PM's interpretation of the motion	
	Deputy Prime Minister (DPM)	Third		Deputy Leader of the Opposition (DLM)	Fourth
	• Refutes the case of OO • Rebuilds the case of OG • May add new arguments to the case of the PM			• Continues refuting the case of OG • Rebuilds the case of OO • May add new arguments to the case of the LO	
Closing Government (CG)	Member of the Government (MG)	Fifth	Closing Opposition (CO)	Member of the Opposition (MO)	Sixth
	• Defends the general direction and case of OG • Continues refutation of OO • Develops a new argument that is different from but consistent with the case of OG			• Defends the general direction taken by OO • Continues general refutation of OG's case • Provides more specific refutation of CG's case • Provides new opposition arguments	
	Government Whip (GW)	Seventh		Opposition Whip (OW)	Eighth
	• Summarizes the entire debate from the point of view of the proposition, defending the general view point of both OG and CG with a special eye toward the case of CG • Does not provide new arguments			• Summarizes the entire debate from the point of view of the proposition, defending the general view point of both OO and CO with a special eye toward the case of CO • Does not provide new arguments	

B. Example Motion, Factual Belief Questions, and Attitudes Elicitation

For every motion, we devise four factual statements and two charitable donations tailored to the motion.

All facts are based on exact statistics from high quality research/reports/surveys. Instead of exact statistics, we report to subjects broad intervals, including values either above or below a given threshold, within which the exact statistic may or may not fall into. This allows us to formulate binary statements for which we ask debaters to predict whether the statement is true or false. Factual statements are devised in a way that truths that appear *convenient* on one side of the debate are instead *inconvenient* on the opposite side. Figure B.1 presents one of the four factual statements devised for a motion on breaking up big tech companies, and provides the distribution of elicited beliefs. This factual statement was devised expecting that it would be convenient for a speaker arguing *in favor* of the motion if the statement were true, and convenient for a speaker arguing *against* the motion if it were false. For both tournaments we collect 36 factual questions related to the motion. About half of these factual statements are favorable to the proposition (opposition) if true.

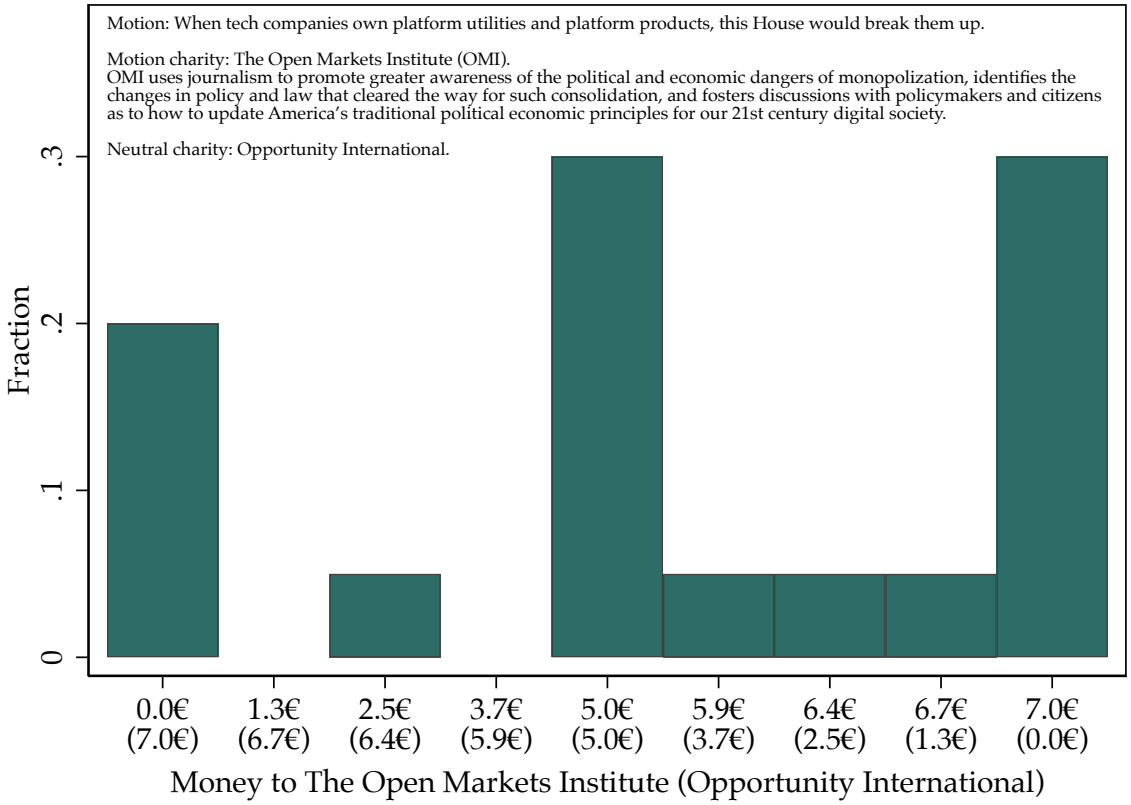
Figure B.1: Example Distribution of Reported Beliefs on a Factual Statement



All charities related to the motion are selected NGOs such that individuals on one side of the debate, who are truly convinced of the factual and moral merits of their persuasion goal, would tend to favor the charitable cause related to the motion. Figure B.2 presents one of the two motion charities devised for the motion on breaking up big tech companies, and provides the distribution of elicited monetary allocations. In this case, we expected individuals who would genuinely argue the proposition (opposition) side of the debate to display a relative preference for the motion charity (neutral charity). The choice of charities is restricted to NGOs that have no known (or alleged) relationship with terrorist organizations.[13]

[13]Non-trivial restriction given that two of the nine motions were explicitly related to terrorism.

Figure B.2: Example Distribution of Chosen Monetary Allocations Between a Motion-Specific Charity and a Neutral Charity



C. Belief and Attitude Convergence

In this section we present estimation of two quantitative measures of cultural polarization. First, we present estimates of an axiomatized index of polarization for continuous distributions (Duclos et al., 2004). Second, we present estimates of an index of cultural distance, borrowed by economists from population genetics, that incorporates socio-demographic information to assess distance along a particular dimension across cultural labels (Desmet et al., 2017).

The first measure of polarization, reflects an identification-alienation framework of conflict, in which polarization and conflict are intimately related, and conflict in society stems from alienation across individuals and proximity within groups of individuals that are alienated from the rest of society. This measure ignores cultural labels, but rather incorporates identities as modal observations of the variable of interest y.

$$P_\alpha(y) = \int \int f(y)^{1+\alpha} f(y') |y - y'| dy \, dy'$$

for $\alpha \in [0.25, 1]$ polarization sensitivity parameter.

The second measure of polarization Φ_{ST}, incorporates cultural labels to capture the extent to which, along the outcome of interest y, individuals within a certain group are similar to one other relative to overall similarity in the population. Such index is obtained as

$$\Phi_{ST}(y) = \frac{P_0(y) - \sum_{g \in G} w_g P_0(y)_g}{P_0(y)}$$

where $P_0(y)$ is the polarization index estimated at $\alpha = 0$, g denotes a cultural label in the set of cultural labels G, w_g is the share of individuals in the population with cultural label g, and $P_0(y)_g$ is the polarization index computed for the distribution of y among individuals in group g at $\alpha = 0$.

Table C.1: Cultural Distance and Polarization, by Question and Survey

Motion	Φ_{ST}				P^2			
	Base (B)	Pre (P)	Post (B)	Post (P)	Base (B)	Pre (P)	Post (B)	Post (P)
1	0.028	0.010	0.007	0.018	0.288	0.330	0.285	0.279
	0.011	0.008	0.005	0.015	0.315	0.300	0.300	0.313
2	0.014	0.022	0.021	0.032	0.284	0.309	0.292	0.310
	0.024	0.070	0.019	0.043	0.326	0.323	0.311	0.310
3	0.006	0.080	0.021	0.078	0.285	0.280	0.279	0.298
	0.006	0.035	0.008	0.019	0.297	0.316	0.294	0.299
4	0.005	0.018	0.012	0.005	0.295	0.281	0.291	0.272
	0.014	0.010	0.007	0.010	0.280	0.287	0.300	0.308
5	0.010	0.007	0.002	0.022	0.304	0.326	0.291	0.277
	0.004	0.010	0.039	0.019	0.309	0.301	0.275	0.286
6	0.023	0.050	0.016	0.016	0.322	0.300	0.288	0.289
	0.015	0.011	0.108	0.038	0.309	0.296	0.312	0.293
7	0.006	0.069	0.009	0.015	0.303	0.283	0.272	0.280
	0.025	0.033	0.035	0.052	0.315	0.306	0.292	0.282
8	0.015	0.036	0.045	0.061	0.286	0.299	0.311	0.300
	0.022	0.046	0.008	0.019	0.312	0.335	0.298	0.278
9	0.004	0.024	0.017	0.026	0.284	0.288	0.297	0.322
	0.008	0.075	0.030	0.011	0.305	0.294	0.292	0.281
Average	**0.013**	**0.034**	**0.023**	**0.028**	**0.301**	**0.293**	**0.303**	**0.293**
95% CIs	[0.010 – 0.017]	[0.023 – 0.046]	[0.012 – 0.034]	[0.019 – 0.037]	[0.294 – 0.308]	[0.288 – 0.299]	[0.295 – 0.311]	[0.286 – 0.300]

Notes: Confidence intervals around the average of each index across questions are obtained from 500 simulated bootstrap samples of the indices underlying the average. Base (B) [Post (B)] refers to indeces computed on answers collected from questions that are only asked at baseline [postdebate]. Pre (P) [Post (P)] refers to indeces computed on answers collected from questions that are only asked at predebate [postdebate].

Table C.1 shows relatively little cultural distance across proposition and opposition speakers, and moderate polarization along elicited beliefs.

The bottom raw of the table aggregates the indices computed at the question-survey level to make inference about how debates affect these measures. We find that on average polarization increases from baseline to postdebate, and remains constant from predebate to postdebate. This suggests that debates can increase polarization because of self-persuasion, and the exchange of views taking place during debates may be ineffective at driving a social consensus.

Cultural distance increases from baseline to postdebate, and decreases (by a somewhat smaller extent) from predebate to postdebate. These patterns confirm that self-persuasion drives beliefs apart between proposition and opposition speakers, and

show that the exchange of views can play some role in reducing divergence.[14]

Table C.2 shows that the debate helps speakers form beliefs that are closer to the truth ((1) and (2)). Columns (3) to (11) provide the simplest possible tests of beliefs and attitude convergence that were included in the pre-analysis plan. The results are largely consistent with the main analysis presented in Section 3: at the individual level, (i) distance from median belief is larger at postdebate than it is at baseline, (ii) distance from median belief is not statistically different between postdebate and predebate, and (iii) the same is for distance from median chosen charity allocation bundle. Columns (7), (8), and (11) indicate that even if we restrict the analysis to the half of the sample of subjects whose beliefs at baseline are aligned to the randomly assigned persuasion goal we observe similar qualitative patterns as for the full sample. This analysis is however only very suggestive as we are clearly under-powered to detect significant convergence/divergence in this sub-sample.

Table C.2: Fixed Effect Regression for Convergence in Beliefs and Attitudes

	Distance from Truth		Distance from Median								
			Beliefs						Charity allocation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Baseline survey	4.152**		-1.429*	-1.214			-1.708				
relative to Postdebate	(1.654)		(0.836)	(1.204)			(1.262)				
Predebate survey		1.998			0.953	0.810		1.907	0.002	0.001	-0.002
relative to Postdebate		(1.478)			(0.813)	(1.090)		(1.258)	(0.055)	(0.080)	(0.071)
Baseline survey × Heated debate				-0.402						0.000	
				(1.656)							
Predebate survey × Heated debate						0.267				0.002	
						(1.512)				(0.105)	
Heated debate				0.837		3.367***				0.027	
				(1.284)		(1.276)				(0.110)	
Observations	1753	1769	1753	1753	1769	1769	856	855	1766	1766	854

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$

Notes: Heated debate is a binary variable indicating, for each round of debate, the debates in which the average subjective heat score of speakers in a debate room is above the median. Standard errors in parentheses are clustered at the team level.

A recurrent finding in social psychology and political science is that the exchange

[14]Unfortunately, by design, we can only directly compare the estimates of these indices from baseline to postdebate and from predebate to postdebate, as the underlying factual statements on which beliefs are elicited differ for different debaters across these two sets of surveys.

of views can either polarize or unite individuals depending on the level of conflict that surrounds the conversation (see e.g. Mutz, 2007, and references therein). Hence, we interact a measure of conflict in a debate, based on how heated enumerators score single debaters in a debate room to be, with the timing of the outcome elicitation.[15] We would have expected more heated debates to possibly increase polarization and less heated debates to decrease it, but we do not find support for such interaction.

[15] If we instead use for this analysis an objective measure of conflict in a debate, given by the number of times that speakers in a debate are challenged by the opposing teams, we obtain qualitatively similar results.

D. Additional Figures and Tables

Table D.1: Debater Characteristics by Tournament

	Full sample	by tournament		
		Munich	Rotterdam	p-value
Female	0.351	0.427	0.261	0.017
	(0.035)	(0.049)	(0.047)	
Age	21.715	21.573	21.878	0.196
	(0.205)	(0.302)	(0.274)	
Time in debating	2.326	2.340	2.311	0.809
	(0.072)	(0.099)	(0.106)	
Past achievements	3.218	2.078	4.522	0.192
	(0.763)	(1.199)	(0.876)	
Local nationality	0.245	0.250	0.239	0.860
	(0.031)	(0.043)	(0.045)	
Left to right political ideology scale	3.372	3.294	3.461	0.734
	(0.134)	(0.173)	(0.208)	
Observations	196	104	92	196

Note: The last column reports the p-value from a one-way ANOVA on ranks (Kruskal-Wallis) test comparing the two tournaments.

Table D.2: Debaters' Baseline Beliefs and Characteristics, by Tournament and Side of the Motion

	Munich				Rotterdam			
	Full sample	Opposition	Proposition	p-value	Full sample	Opposition	Proposition	p-value
				(a) By motion				
Baseline belief motion 1	44.369	45.596	43.118	0.764	52.322	52.022	52.636	0.881
	(3.084)	(4.303)	(4.456)		(3.474)	(5.212)	(4.623)	
Baseline belief motion 2	39.794	36.314	43.275	0.193	51.378	46.854	56.548	0.131
	(3.131)	(4.652)	(4.181)		(3.084)	(4.537)	(4.008)	
Baseline belief motion 3	65.000	64.451	65.549	0.965	39.483	40.907	38.152	0.578
	(2.622)	(3.837)	(3.609)		(3.255)	(4.498)	(4.729)	
Baseline belief motion 4	52.363	51.667	53.059	0.820	56.989	58.444	55.500	0.684
	(2.818)	(3.996)	(4.010)		(3.173)	(4.525)	(4.489)	
Baseline belief motion 5	71.588	72.608	70.569	0.968				
	(2.645)	(3.403)	(4.079)					
Observations	104	52	52		96	48	48	
				(b) All motions				
Female	0.427	0.438	0.416	0.620	0.262	0.258	0.266	0.874
	(0.022)	(0.031)	(0.031)		(0.024)	(0.033)	(0.034)	
Age	21.573	21.519	21.626	0.948	21.877	21.847	21.909	0.703
	(0.134)	(0.183)	(0.197)		(0.137)	(0.194)	(0.193)	
Time in debating	2.340	2.341	2.339	0.981	2.315	2.279	2.352	0.464
	(0.044)	(0.062)	(0.063)		(0.053)	(0.074)	(0.074)	
Achievements	3.069	3.196	2.941	0.583	4.529	4.284	4.784	0.766
	(0.304)	(0.457)	(0.402)		(0.437)	(0.583)	(0.656)	
Local nationality	0.250	0.238	0.263	0.527	0.237	0.246	0.228	0.682
	(0.019)	(0.026)	(0.027)		(0.022)	(0.032)	(0.031)	
Political scale	3.294	3.271	3.318	0.843	3.462	3.497	3.425	0.612
	(0.077)	(0.108)	(0.110)		(0.104)	(0.143)	(0.151)	
Observations	519	259	260		367	175	192	

Note: P-value is from a one-way ANOVA on ranks (Kruskal-Wallis) test comparing the two groups. Each observation is a debater at each round of the tournament. For panel (a) we have a total of 104 observations for each Factual Beliefs relating to the motions of each round. For panel (b), where the outcomes are not round specific while treatment assignment is, the number of observations equals the number of debaters in each position across all rounds of the tournament.

Table D.3: Debaters' Baseline Characteristics, by Tournament

	Munich				Rotterdam			
	Full sample	Group 1	Group 2	p-value	Full sample	Group 1	Group 2	p-value
Female	0.427	0.451	0.404	0.630	0.261	0.349	0.178	0.069
	(0.049)	(0.070)	(0.069)		(0.047)	(0.074)	(0.058)	
Age	21.573	21.667	21.481	0.519	21.878	22.233	21.553	0.282
	(0.302)	(0.422)	(0.435)		(0.274)	(0.417)	(0.357)	
Time in debating	2.340	2.314	2.365	0.732	2.311	2.302	2.319	0.953
	(0.099)	(0.144)	(0.137)		(0.106)	(0.158)	(0.143)	
Achievements	3.069	2.255	3.882	0.223	4.522	4.488	4.553	0.880
	(0.682)	(0.557)	(1.243)		(0.876)	(1.133)	(1.331)	
Local nationality	0.250	0.269	0.231	0.652	0.239	0.227	0.250	0.800
	(0.043)	(0.062)	(0.059)		(0.045)	(0.064)	(0.063)	
Political scale	3.294	3.627	2.961	0.108	3.461	3.738	3.213	0.227
	(0.173)	(0.264)	(0.215)		(0.208)	(0.293)	(0.293)	
Observations	104	52	52		92	44	48	

Note: The two partitions of teams (Group 1 and Group 2) answer the same set of question, but answer sets of factual beliefs and attitude elicitations in different orders across surveys. P-value is from a one-way ANOVA on ranks (Kruskal-Wallis) test comparing the two groups.

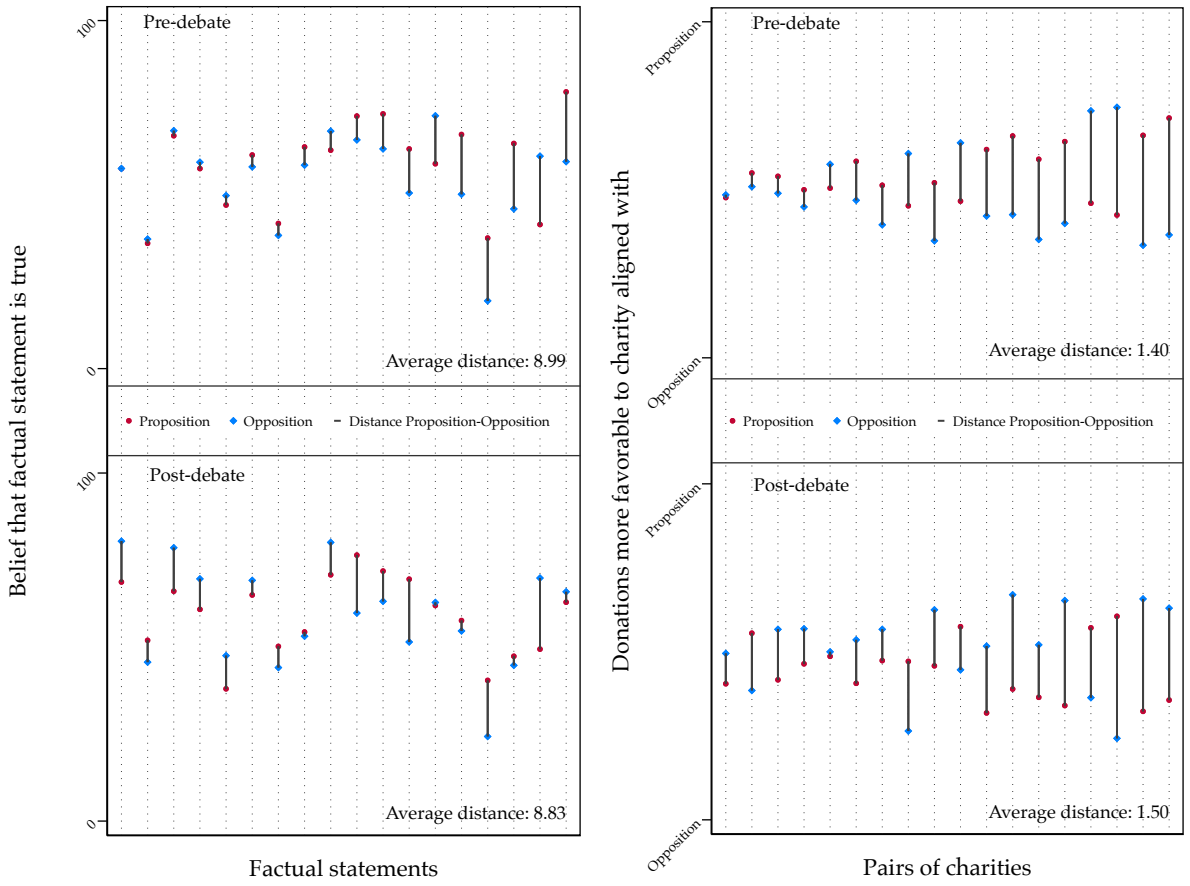
Table D.4: Ordered Logit Regressions for Effect of Persuasion Goals on the Allocation of Charitable Donations

	Donation bundle favorable to proposition charity		
	(1)	(2)	(3)
Speaker in proposition	0.271**	0.274**	0.282**
	(0.120)	(0.127)	(0.131)
Socio-demographic and experience controls		✓	
Debater fixed effects			✓
Round FEs	✓	✓	✓
Observations	883	850	883

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

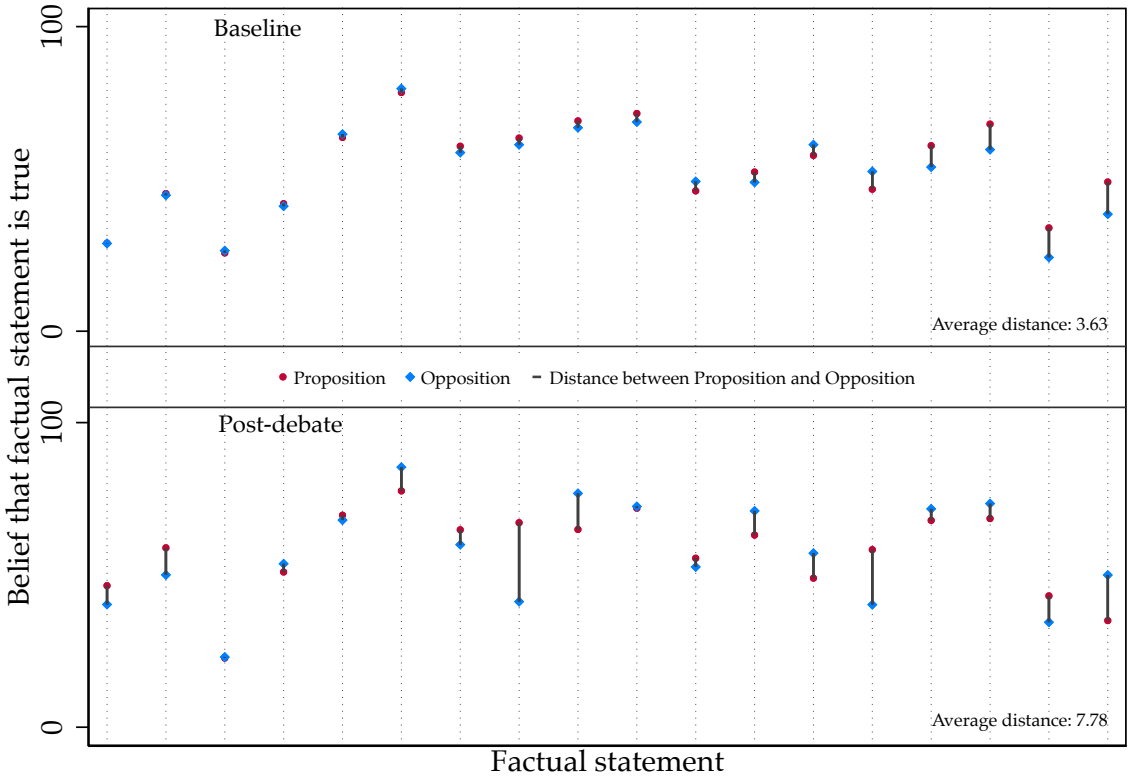
Notes: Standard errors in parentheses are clustered at the team level for the random effects estimates (columns (1) to (2)), and at the individual level for the fixed effects estimates (column (3)). Fixed effects estimates are obtained from the Baetschmann et al. (2015) estimator to overcome notorious under-identification problem of ordered logit models with fixed effects Chamberlain (1980). Socio-demographic controls include age, gender, and an indicator for whether the speaker's nationality is from the country that hosts the competition. Experience controls include the reported number of international tournaments in which the speaker has made it to semi-finals, and a categorical variable capturing the number of years the speaker has been actively debating.

Figure D.1: Distance in Beliefs and Attitudes, Pre- and Post- Debate



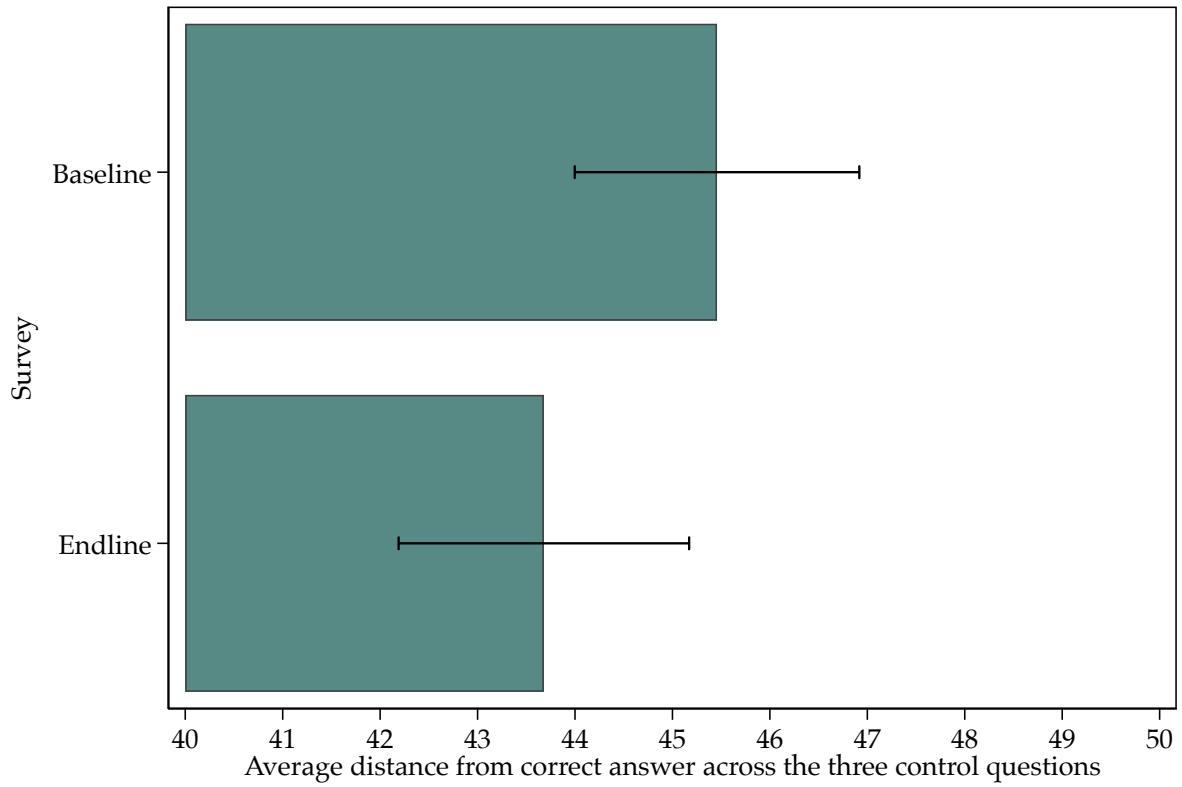
Note: Each vertical dotted line represents either a factual statement over which beliefs are elicited at predebate (top left panel) and postdebate (bottom left panel), or a pair of charities between which debaters allocate monetary endowments at predebate (top right panel) and postdebate (bottom right panel). In the left (right) panel, colored markers represent average report (chosen monetary allocation bundle) among speakers on each side of the debate. Black segments between each pair of colored markers represent the distance in the average position of speakers on the two sides of the debate. For each panel, for readability, factual statements and pairs of charities are sorted by distance between average proposition and opposition outcomes at the predebate stage. The four sets of outcomes are summarized in the bottom right corner by the average distance between the average positions of proposition and opposition.

Figure D.2: Distance in Beliefs, at Baseline and Post- Debate



Note: Each vertical dotted line represents a factual statement over which beliefs are elicited at baseline (top panel) and postdebate (bottom panel). Colored markers represent average report among speakers on each side of the debate. For readability, factual statements are sorted by distance between average proposition and opposition outcomes at the baseline stage. The two sets of outcomes are summarized in the bottom right corner by the average distance between the average positions of proposition and Opposition.

Figure D.3: Evidence on Learning of Correct Answers to Belief Elicitation Questions Through the Entire Tournament



Note: Mean distances of reported beliefs from correct answers are averaged at the individual level for the three control questions in each survey. This figure reports the survey average of such individual-survey level metrics and the corresponding error bars.

E. Predictors of Persuasiveness

Table E.1: Panel Regressions fo Correlation Between Persuasiveness and Alignment with the Motion (Standard Errors in Parentheses)

	Broad persuasiveness score				Quality of argumentation score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline belief aligned (binary outcome)	0.009	0.015			0.114	0.092		
	(0.075)	(0.076)			(0.216)	(0.218)		
Baseline belief alignment (continuous outcome)			-0.009	-0.011			0.129	0.109
			(0.034)	(0.036)			(0.110)	(0.110)
Debater FEs	✓		✓		✓		✓	
Socio-demographic and experience controls		✓		✓		✓		✓
Round FEs	✓	✓	✓	✓	✓	✓	✓	✓
Observations	869	848	869	848	869	848	869	848

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Heteroskedasticity robust standard errors in parentheses.

Table E.2: Pair-wise Correlation Between Persuasion Outcomes and Potential Predictors

	Broad persuasiveness	Quality of arguments
	(1)	(2)
(a) Pearson's correlation		
Achievements	0.475***	0.528***
	(0.000)	(0.000)
Factual knowledge at baseline	0.118	0.126*
	(0.102)	(0.080)
Predebate share of strong arguments for the other side of the debate	0.037	0.087
	(0.604)	(0.229)
Predebate share of arguments for the other side of the debate	0.017	0.042
	(0.814)	(0.564)
(b) Spearman's rank correlation		
Time in debating	0.549***	0.479***
	(0.000)	(0.000)
Observations	196	196

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: P-value for statistical significance in parentheses. All variables for this analysis are averaged across all rounds of debate. Broad persuasiveness of a debater is evaluated by each judge on the panel independently; for this analysis we use panel averages of broad persuasiveness. Factual knowledge at baseline captures, how close debaters' beliefs on the 5 motion related factual statements presented at baseline are to the truth. Predebate belief (attitude) alignment captures how close debaters' beliefs are to the response aligned with their persuasion goal.

F. Heat of Debates

Table F.1 summarizes our two measures of heat in a debate. The first is an objective proxy obtained by counting how many times a speaker is challenged by non-speaking debaters in the room. The second is a subjective heat score that the enumerator attributes to each speaker in the room. The average of these two individual outcomes at the round level are informative of how much heat each motion generates, and visual inspection of the table already indicated a positive correlation of these two outcomes.

Table F.1: Average Heat Score (Standard Errors in Parentheses)

Motion	Number of POIs	Subjective heat
	(1)	(2)
This House believes that governments should stop funding scientific programs that have	4.165	2.680
no immediate benefit for humankind (such as space travel and exploration, human cloning).	(0.300)	(0.123)
This House believes that Western States should permanently revoke the citizenship of	5.202	2.961
citizens who join terrorist organisations.	(0.362)	(0.111)
This House regrets the EU's introduction of freedom of movement	4.260	2.798
	(0.361)	(0.101)
This House would suspend trade union powers and significantly relax labour protection	4.260	2.721
laws in times of economic crisis.	(0.360)	(0.104)
This House believes that causing deliberate harms to enemy civilians, by the	4.337	2.817
weaker side, is a justified tactic in asymmetrical warfare.	(0.346)	(0.112)
Observations	104	104
During periods of national housing shortages, this House would forcibly take	4.054	3.033
ownership of privately owned homes which are not lived in by their owners).	(0.358)	(0.113)
This House believes that states should aggressively fund geoengineering projects	4.152	3.352
instead of attempting to mitigate the effect of climate change.	(0.305)	(0.126)
This House regrets the decision to let the FARC (i.e. The Revolutionary	4.272	3.033
Armed Forces of Colombia -People's Army) run as a political party.	(0.442)	(0.103)
When tech companies own platform utilities and platform products,	3.739	2.835
this House would break them up.	(0.361)	(0.123)
Observations	92	92

Note: Column (1) reports the number of Points of Information, the event of a non-speaking debater standing up to challenge the speaker, received by each speaker. Column (2) reports the score, on a scale from 1 "Not heated at all" to 5 "Very heated" that the enumerator assigns to each speaker for her performance.

Table F.2: Pair-wise Correlation Between Measures of Debate Heat and Baseline Alignment

	POIs above median	Subjective heat scores above median	Baseline belief alignment
	(1)	(2)	(3)
POIs above median	1.000***		
	(0.000)		
Subjective heat scores above median	0.281***	1.000***	
	(0.002)	(0.000)	
Baseline belief alignment	0.184*	0.036	1.000***
	(0.051)	(0.702)	(0.000)
Observations	114	114	114

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$

Note: P-value for Statistical Significance in Parentheses The unit of observation for this analysis is a debate. The number of Points of Information and the subjective heat scores are aggregated at the debate room level, and for each of these aggregate measures we construct a binary indicator variable to denote, within each round, the debate rooms with aggregate score above median.

At the individual level, the first measure is a poor predictor of how heated the speaker is, because in fact the measure captures how heated the non-speaking debaters in the room are. Aggregating each of the two individual level measures at the debate room level allows us to obtain two outcomes that lend themselves to an interpretation in terms of heat. Table F.2 quantifies the correlation between the measures of heat of a debate: such correlation of 0.28 is substantial, but far from perfect. To complement the set of result on the correlation between alignment and persuasiveness, we show some evidence that the more debaters' beliefs turn out to be aligned with their persuasion goals, the more heated the debate turns out to be. This is interesting, because it suggests that debaters who truly believe in their position act more forcefully during the debate. Though, as shown in section 3, such additional energy does not translate into significantly better persuasion outcomes.

G. Robustness to Experimenter Demand Effects

When subjects of experimental work are able to infer the research hypotheses under investigation, we often worry that they may distort their reports to help the researchers prove their hypotheses. To reduce such concerns, one can raise the costs for subjects to distort their reports to conform to the researchers' hypotheses. This is what we achieve in our experiment by eliciting incentivized beliefs, and by asking subject to distribute monetary endowments between causes that generate real social returns.

By definition, for experimenter demand effects to potentially drive the results, it is necessary that subjects are able to infer the research hypotheses under investigation. To establish the extent to which they can, at the end of our study, we ask subjects of our experiment to write down in an open field text box what they thought the research was trying to demonstrate.

Table G.1: Categorization of Debaters' Response

(a) Having to argue for a given position alters the perception of empirical facts	0.227
	(0.032)
(b) Having to argue for a given position alters the perception of values	0.125
	(0.025)
(c) Having to argue for a given position makes individuals relatively more confident about the merit of their position	0.091
	(0.022)
(d) Positive correlation between private beliefs aligned with the persuasion goal and persuasiveness	0.142
	(0.026)
(e) Convergence of opinions through the debate	0.131
	(0.025)
(f) Other research questions	0.284
	(0.034)
(g) Overly generic answer	0.301
	(0.035)
Answered question	176
Left field blank	20
Observations	196

Notes: Open-field answers are categorized by a research assistant to be either an overly generic answer, or to reflect at least one of the research hypotheses (a) to (e) and possibly other potential research hypotheses. We report shares of respondents (and standard errors) in each category among the 90 percent of respondents who did not leave the open-field question unanswered.

The majority of subjects reported fairly sophisticated guesses.[16] In Table G.1 we

[16]Some responses were fairly accurate in capturing many of the research hypotheses (e.g. *"1. See how*

report the result of our manual categorization of non-blank responses (90 percent of the sample). Among these, only 30 percent give an overly generic answer, while the rest seem to have in mind some concrete research hypotheses. The most frequent category is our residual category "Other research questions", that includes questions that were not part of our pre-registered hypotheses. Relatively frequently, subjects also seem to appreciate some reasonably close version of our primary research hypothesis of self-persuasion on facts.

Studies that try to bound the extent to which experimenter demand effects can explain experimental results, assess how sensitive results are to increasing awareness among subjects of the experimenters' research hypotheses De Quidt et al. (2018). In the absence of such exogenous variation of awareness of research hypotheses, an imperfect but informative exercise that we can conduct is to provide evidence of how results change when we exclude from the test of a specific hypothesis the responses of subjects who were able to figure out that hypothesis. In Table G.2 we do exactly that to consolidate our self-persuasion results obtained by comparing belied, attitude, and confidence alignment with the persuasion goal. Reassuringly, we find that the magnitudes of the differences in all three outcomes between proposition and Opposition speakers, estimated for the subset of "unaware subjects", are very similar to the ones estimated in the full sample.

engaging with motion from a certain assigned point of view influences perception of facts in accordance to position in debate 2. how belief/being convinced of position in debate affects debaters persuasiveness (that's why you gave us scores on persuasion and rhetoric as well) –> How debating from assigned point of view affects opinion and how that affects performance in debate", some others completely miss the main hypotheses (e.g. *"Connection between knowledge and persuasivness? - Not sure, would love to find out!"*), and some others are overly generic (e.g. *"Game-theory"*).

Table G.2: Replication of Main Results Excluding Subjects Who Could Guess The Research Hypothesis at the End of the Tournament

	Beliefs aligned with proposition	Attitudes aligned with proposition	Confidence in proposition
	(1)	(2)	(3)
Speaker in proposition	0.235***	0.243*	4.325***
	(0.065)	(0.127)	(1.581)
Debater fixed effects	✓		✓
Round FEs	✓	✓	✓
Observations	698	779	813

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$

Note: Column (1) replicates analysis in column (1) of Table 3 excluding subjects who guessed the research hypothesis of self-persuasion on facts. Column (2) replicates analysis in column (1) of Table 4 excluding subjects who guessed the research hypothesis of self-persuasion on the values of social causes. Column (3) replicates analysis in column (1) of Table 5 excluding subjects who guessed the research hypothesis that debaters who be relatively more confident of the merits of their own position.

H. Mechanisms

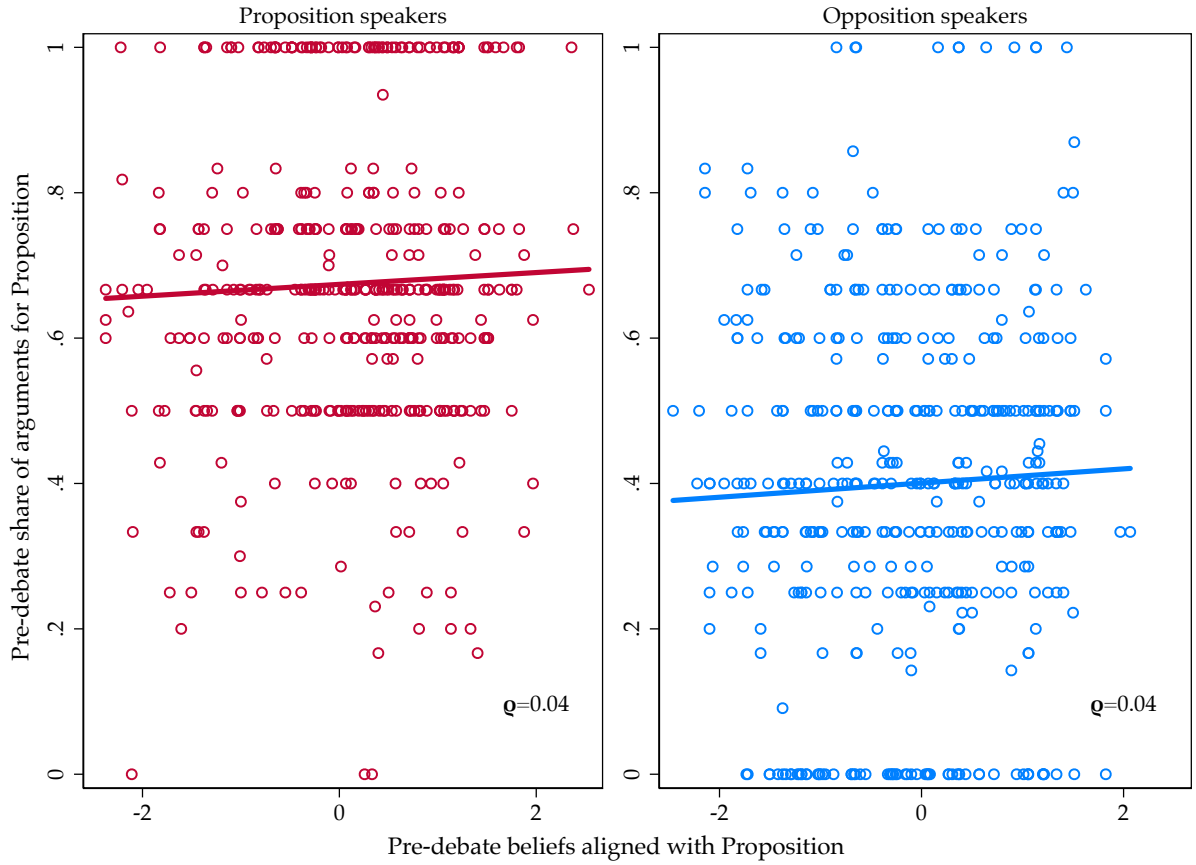
Our discussion proposes that persuasion goals can have both a direct effect on belief alignment due to strategic choice of beliefs and an indirect effect due to the cognitive constraints that generate bias when debaters sample an unbalanced set of arguments to prepare their speech. In a linear framework, such direct and indirect effects can be assessed through the following system of structural equations

$$
\begin{aligned}
Y_i &= \alpha_1 + \beta_1 T_i + \phi_1 X_i + \epsilon_{i1} \\
M_i &= \alpha_2 + \beta_2 T_i + \phi_2 X_i + \epsilon_{i2} \\
Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \phi_3 X_i + \epsilon_{i3}
\end{aligned}
\tag{H.1}
$$

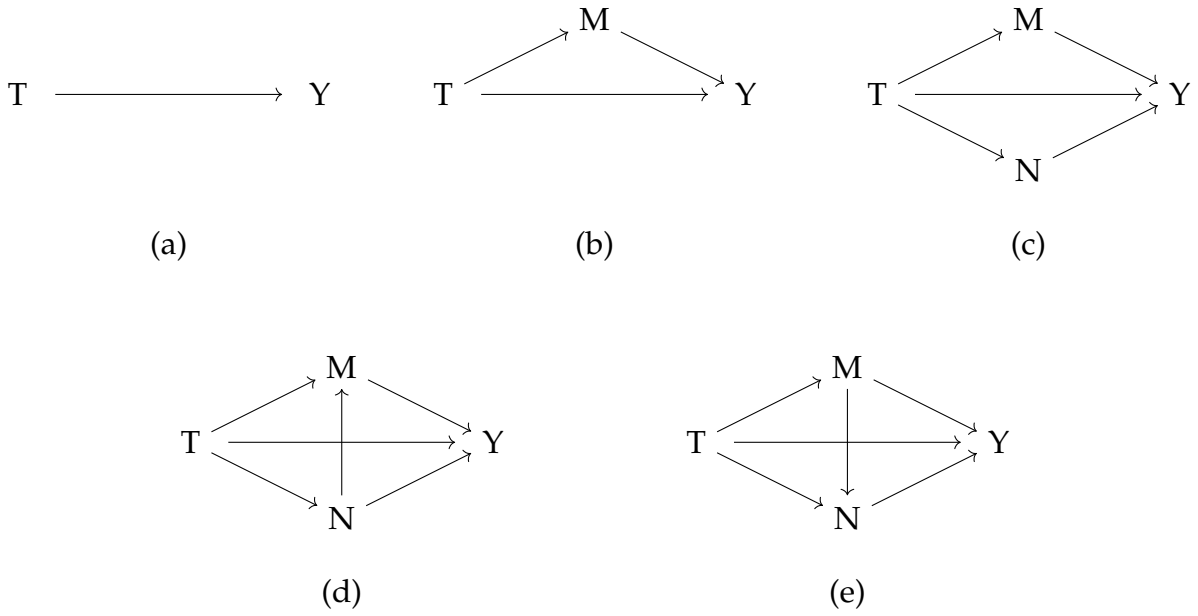
where standard notation is used for expositional purposes: Y_i is the outcome of interest, T_i is the treatment variable, M_i is the intermediate outcome measure after treatment that mediates the treatment effect, and X_i is a vector of controls. β_1 represents the average treatment effect (ATE), which includes both direct and indirect effects of the treatment on the main outcome of interest. If the structural equations are correctly specified, a *sequential ignorability* assumption allows to interpret $\gamma\beta_2$ as the *causal* indirect effect of T_i, mediated through M_i, on Y_i Imai et al. (2010b).

Sequential ignorability requires that (i) conditional on X_i, the outcome and the mediator are distributed independently of the treatment, and (ii) conditional on T_i and X_i, the outcome is distributed independently of the mediator. Both conditions are fairly strong. Because our treatment assignment is randomized, the first condition is met by design. However, the second condition does not directly follow from random assignment, and is hard to test. If the second condition is met, we would expect that the outcome and the mediator are uncorrelated within treatment. Figure H.1 provides supporting evidence of the lack of such correlation.

Figure H.1: Correlation Between Share of proposition Arguments and Predebate Belief Alignment, Within Each Side of the Debate



In Figure H.2 we include diagrams that illustrate potential causal links between the treatment, mediating factors, and the outcome. Assuming sequential ignorability rules out causal links between mediators (sub-figures (d) and (e)), but allows for multiple downstream causal relationships from treatment, through mediators, to the outcome of interest (sub-figures (a) to (c)), so that by estimating $\gamma\beta_2$ from H.1 we could directly obtain a valid estimate of the causal effect of the treatment mediated through M_i.



Note: In (a), the outcome can only be affected directly by the treatment variable. In (b), the treatment affects both the outcome directly and an intermediate mediator; the mediator in turn affects the outcome. In (c), the treatment affects both the outcome directly and two intermediate mediators; both mediators in turn affect the outcome. In (d) and (e), the treatment affects both the outcome directly and two intermediate mediators; both mediators in turn affect the outcome, and mediators also affect one another.

Figure H.2: Diagrams Representing Possible Causal Mechanisms Between Treatment, Mediating Outcomes, and Main Outcome

In the potential outcome framework with binary treatment $t \in \{0, 1\}$ and one mediator it is straightforward to derive the causal mediated effect directly as a component of the average treatment effect $\tau_i = Y_i(1) - Y_i(0)$, which can be equivalently written as $Y_i(1, M_i(1)) - Y_i(0, M_i(0))$. With some algebra, it is simple to obtain that

$$2[Y_i(1, M_i(1)) - Y_i(0, M_i(0))] = \overbrace{Y_i(1, M_i(1)) - Y_i(1, M_i(0))}^{\delta_i(1)} + \overbrace{Y_i(0, M_i(1)) - Y_i(0, M_i(0))}^{\delta_i(0)} +$$
$$+ \overbrace{Y_i(1, M_i(1)) - Y_i(0, M_i(1))}^{\zeta_i(1)} + \overbrace{Y_i(1, M_i(0)) - Y_i(0, M_i(0))}^{\zeta_i(0)}$$

where $\delta(t)$ defines the indirect effect of the treatment in treatment t, and $\zeta_i(t)$ defines the direct effect of the treatment holding constant the level of the mediator at the treat-

ment t level. When $\delta_i(t) = \delta_i$ and $\zeta_i(t) = \zeta_i$ for any t, there is no interaction between treatment and mediator, and the ATE can simply be expressed as $\tau_i = \delta_i + \zeta_i$, yielding a simple decomposition of the ATE in average causal mediated effect (ACME) and average direct effect (ADE).

To identify the ACME of persuasion goals on belief alignment with proposition b_i through the share of proposition arguments considered during preparation period s_i, we estimate the following random effects models with standard errors clustered at the team level

$$\begin{aligned}
&\textit{Model 1:} && b_{i,m} = \alpha_1 + \beta_1 proposition_{i,m} + \phi_1 X_i + \epsilon_{i1,m} \\
&\textit{Model 2:} && s_{i,m} = \alpha_2 + \beta_2 proposition_{i,m} + \phi_2 X_i + \epsilon_{i2,m} \\
&\textit{Model 3:} && b_{i,m} = \alpha_3 + \beta_3 proposition_{i,m} + \gamma s_{i,m} + \phi_3 X_i + \epsilon_{i3,m}
\end{aligned}$$

and use sampling distributions of the parameter estimates from *model 1* to simulate potential outcomes $b_{i,m}(proposition_{i,m} = 1)$ and $b_{i,m}(proposition_{i,m} = 0)$, from *model 2* to simulate potential outcomes $s_{i,m}(proposition_{i,m} = 1)$ and $s_{i,m}(proposition_{i,m} = 0)$, and from *model 3* to simulate potential outcomes $b_{i,m}(1, s_{i,m}(1))$, $b_{i,m}(0, s_{i,m}(1))$, $b_{i,m}(1, s_{i,m}(0))$, and $b_{i,m}(0, s_{i,m}(0))$. Table 6 in the main text reports the results from this exercise.

I. Surveys

I.1. General instructions

A two-page general instructions document includes relevant information for answering the surveys throughout the tournament. In particular this explains how belief elicitations are incentivized using the Quadratic Scoring Rule for binarized outcomes (Harrison et al., 2014), how charitable allocations are paid out, and general payment procedures. All subjects are given 10 minutes to carefully read these general instructions right before the baseline survey begins. To make sure that procedures are adequately understood, if subjects miss their opportunity to read the general instructions we exclude them from the study.[17] The original content of these instructions is provided below.

.....

General Instructions

Please read the following instructions carefully and keep them in mind, as they contain information that is relevant for the surveys we will ask you to complete during the next two days. We kindly ask you to use the time allocated to each survey to focus exclusively on answering the questions in front of you; throughout these times no information regarding the debates will be provided. Please answer each question carefully, don't use your phone and don't interact with others. Our instructions are never deceptive. All of your answers are treated confidentially and used for research purposes only.

Assessing factual statements

Spread across the various surveys, there are 34 questions that are marked by an "$", for which you can earn money. After you completed the last survey, we will pay you based on one randomly selected answer. While you will get paid for only one of your answers, every question might be the one that counts.

[17] They are allowed to answer the surveys, but their data is discarded.

Questions marked by an "$" ask you to state the likelihood (in percent) that a given statement is true. Most such statements are designed to assess your factual knowledge. There will be no trick questions. Moreover, all sources we refer to actually exist and are of high quality, but the actual fact may be either true or not true. As an example, consider the following statement.

According to Eurostat, more than 30 percent of live births in Germany in 2016 were outside of marriage.

This statement is true if Eurostat indeed reported this finding. It is false if Eurostat reported a different finding. You will be asked to provide your belief as to how likely you think it is that this statement is true. If this answer is selected for payment, you will earn either 30 euros or nothing. The procedure that determines how likely it is that you win the 30 euros assures that the closer you are to the correct answer (either 0 or 100 percent), the higher is your probability of winning the money.

Moreover, the procedure assures that you maximize your chance of winning money by stating your true belief (between 0 and 100 percent). So if you are almost certain that a given statement is true, then you should state a belief that is very high. If you are almost certain that a given statement is false, then you should state a belief that is very low. If you are completely uncertain, you maximize your chance of winning by stating a belief that is close to 50 percent.

The Procedure Box below provides more comprehensive information about the exact payment mechanism. But note that it is not important that you understand the procedure in detail. What matters is that you know that you maximize your probability of winning when you report your true belief - if you under- or overstate your belief, you will reduce your chance of winning the 30 euros.

Donating to Charities

For some questions in the survey, you will be able to allocate monetary endowments between different charities. This is money that we make available from our budget for you to allocate, according to your preferences, to charities that have different missions. One of the allocations you make will be selected at random and we will transfer the money to the relevant charities. While we will implement only one of your allocations,

every allocation might be the one that counts.

The surveys will also feature further questions that allow you to earn more money for yourself. The instructions for these questions are simple and will be provided above the relevant question.

Procedure Box

How a given answer maps into your chance of winning 30 euros is based on a formula. This formula is designed to make sure that you maximize your chance of winning if you report your true belief that a given statement is true.

Suppose that the correct answer is given by R, which is equal to 1 if the statement is true and 0 if the statement is false. The variable *r* is your report–the likelihood that you attribute to the statement being true (from 0 to 100 percent). The winning probability for the prize is then given by:

$$\text{winning probability} = 100 - 100 \times (R - r/100)^2$$

Example: Suppose again that you are tasked with assessing the following statement: *According to Eurostat, more than 30 percent of live births in Germany in 2016 were outside of marriage*. And suppose that your belief that the statement is true is 63 percent. The following table shows your winning probability based on the formula. The columns represent a number of hypothetical answers you may give. As you can see, you maximize your chance of winning by reporting your true belief.

	Report 1	Report 2	**Report 3**	Report 4
Hypothetical report	22	35	**63**	89
Expected winning probability if **your belief** that the statement is true is **63%**	59.9%	68.9%	**76.7%**	69.9%

Payment

On Sunday, we will pay out your earnings in cash. To determine your earnings for the assessment of factual statements, we first randomly draw the question that is relevant for your payment. We then determine your winning probability based on the true answer and your reported answer. Finally, a computer program constructs a virtual urn

with only white and black balls, where the share of white balls equals your winning probability. If the computer then draws a white ball from the urn, then you will win the 30-euro prize. This is a fair and transparent procedure to pay you the prize with the winning probability you have earned based on the quality of your answers.

If the question that is drawn for payment is from a round that you missed, then there will be no new draw and you will not earn any money for this type of question. If you would like us to send you receipts of the charity donation based on your choice, then please leave us your email address when you collect your payment.

I.2. General remarks

We take several steps to collect high quality data in a confidential manner.

First, all surveys that debaters fill out begin with a cover page containing brief instructions to (i) inform subjects how much time they have to complete the survey, and (ii) remind subjects of the procedure to collect incentive compatible beliefs. The cover page does not contain any question, and enumerators are instructed to not turn the cover page after surveys are filled out and read the answers provided by debaters.

Second, each survey is linked to the individual who filled it through a personal identifier. Debaters are assigned S### IDs, Judges J## IDs, and Enumerators E## IDs. These IDs allow data to be collected and payments to be carried out confidentially. We ask debaters to enter their S IDs on the cover page of each of their surveys.

Every study participant (debaters, judges, and enumerators) wears a name tag that includes their ID. Before collecting the survey, enumerators double-check that the S ID entered by each debater on the cover page of their survey matches the one on the name tag.

I.3. Baseline survey

A 25-minute baseline survey includes the following items:

- Age (open field, suggested to provide a numeric answer).
- Gender (open field).

- Nationality (open field).
- Political ideology scale: "*In politics people sometimes talk of "left" and "right". Where would you place yourself on this scale, where 0 means the left and 10 means the right?*" (check box).
- Years actively debating on a regular basis. Options: "Less than a year", "1 to 2 years", "3 to 4 years", "At least 5 years". (check box)
- Times debater got to semifinals in Open/IV tournaments (open field).
- "*What do you think makes a good debater*". Options: "Choosing arguments strategically", "Confidence in own position", "Debating experience", "Factual knowledge", "Eloquence" (ranking).
- Incentivized belief elicitation on fifteen factual statements: for each such statement subjects state how likely it is that the fact is true (open field, suggested to provide a numeric answer from 0 to 100).
- "*Did you take part as a speaker at the Munich Research Open 2019?*". Options: "Yes", "No" (check box).[18]

A key component of this survey was to gather beliefs at baseline regarding the motions that subjects were going to debate. At the same time, we had to be careful in not revealing, through our questions, the motion of the debates – which are meant to be secret. To obfuscate the relation of these belief elicitations and the motions we elicit beliefs over whether 15 factual statements are true: 5 such statements relate to the in-round motions, 7 are decoy questions, and 3 are control questions.[19] For each team of debaters, control questions are drawn from a pool of 6 questions, and the questions that were not selected for the baseline survey are then included in the endline survey. Comparing responses to the control questions at baseline and endline by different debaters helps uncover to what extent debaters discuss the contents of the surveys among themselves.

[18]Only in Rotterdam.

[19]In Rotterdam, 4 statements relate to the in-round motions, and 8 are decoy questions.

Decoy questions are designed to look like they could relate to plausible motions for debate. Control questions are facts that not necessarily relate to typical debate topics.

For each motion, we devise multiple factual statements that we phrase as binary states to capture alignment of beliefs with the persuasion goal. Any given question may not have a tight enough link to the motion in debaters' minds or give rise to a high degree of certainty in debaters' beliefs and may therefore be ill-suited to pick up a treatment effect. To diversify this risk, we come up with 4 questions (A, B, C, D) for each motion and administer them as illustrated in the table below: at baseline, debaters are asked either about fact A or B; predebate, debaters are asked either about fact D or C; postdebate debaters are asked either about fact B and C or A and D.

This approach also ensures that (i) no debater is asked the same question twice, and (ii) we protect the baseline and predebate belief elicitations from any potential information spillovers.

Timing:	Beginning of Day 1	Day 1 or Day 2	
	Baseline	Predebate	Postdebate
Subgroup 1	A	D	B, C
Subgroup 2	B	C	A, D

I.4. Predebate survey

This 5 minute survey is handed out before each debate begins and after the preparation time. It includes:

- Incentivized belief elicitation on two factual statements: for each such statement subjects state how likely it is that the fact is true (open field, suggested to provide a numeric answer from 0 to 100).
- Choose one of 9 monetary allocations, along a concave budget, between a baseline charity (either Oxfam or Opportunity International) and a charity aligned with one of the sides represented in the debate. For an illustration see Figure I.1.
- Questions on the number of arguments considered during preparation time in favor of the proposition:

i *How many good arguments did you come up with during the preparation time in favor of the proposition?* (open field, suggested to provide a numeric answer)

ii *How many of these arguments would you consider to be very strong?* (open field, suggested to provide a numeric answer between zero and the answer to the previous question)

- Questions on the number of arguments considered during preparation time against the proposition:

  i *How many good arguments did you come up with during the preparation time against the proposition?* (open field, suggested to provide a numeric answer).

  ii *How many of these arguments would you consider to be very strong?* (open field, suggested to provide a numeric answer between zero and the answer to the previous question).

Figure I.1: Illustration of charitable donations allocation question

Below you see nine potential ways in which you could allocate charitable donations—that are paid by us on your behalf—between two charitable organizations: Oxfam and The Planetary Society .

Oxfam is a major nonprofit group with an extensive collection of operations. Oxfam's programs address the structural causes of poverty and related injustice and work primarily through local accountable organizations, seeking to enhance their effectiveness

The Planetary Society is the world's largest and most influential non-profit space organization. The society advocates for space and planetary science funding in government, invests in inspiring educational programs, and funds groundbreaking space science and technology

How would you like to allocate these donations? (check only one box)

Choose one option	☐	☐	☐	☐	☐	☐	☐	☐	☐	
You want to give	0.0	1.3	2.5	3.7	5.0	5.9	6.4	6.7	7.0	euro to Oxfam
and	7.0	6.7	6.4	5.9	5.0	3.7	2.5	1.3	0.0	euro to the Planetary Society
A total of	7.0	8.0	8.9	9.6	10.0	9.6	8.9	8.0	7.0	*euro goes to charity*

Both factual statements are meant to capture whether beliefs are aligned with the motion after the debate. The first statement features a real-world fact. The second statement elicits confidence in the arguments of the proposition side of the debate by asking:

Statement: Excluding the debate happening in this room, in at least half of the parallel debates of this round, one of the two teams on the Government side of this motion will rank 1st.

Q2$: How likely do you think it is that the above statement is true? ___% (write a number from 0 to 100)

For each motion, we select two charities that we expect to be either positively or negatively aligned. We randomly determine which of these two charities features in the predebate survey. The other charity features in the postdebate survey. In Rotterdam, the baseline charity is always Opportunity International, whereas in Munich we also randomize between Oxfam and Opportunity International to be the baseline charity.

I.5. Postdebate survey

This 5 minute survey is handed out right after each debate. It includes:

- Incentivized belief elicitation on two factual statements: for each such statement subjects state how likely it is that the fact is true (open field, suggested to provide a numeric answer from 0 to 100).
- Subjective ranking of team performance in the debate.
- Choose one of 9 monetary allocations, along a concave budget, between a baseline charity (either Oxfam or Opportunity International) and a charity aligned with one of the sides represented in the debate. For an illustration see Figure I.1.

Both factual statements are meant to capture whether beliefs about real-world facts are aligned with the motion after the debate.

For each motion, we select two charities that we expect to be either positively or negatively aligned. We randomly determine which of these two charities features in the predebate survey. The other charity features in the postdebate survey. In Rotterdam, the baseline charity is always Opportunity International, whereas in Munich we also randomize between Oxfam and Opportunity International to be the baseline charity.

I.6. Endline survey

This 20-minute survey takes place right after the fifth (fourth in Rotterdam) round of debates.

It includes:

- A question that we use to assess how debaters think that beliefs about facts that we ask and charities they can donate to relate to alignment with the motions. An illustration of the precise wording of this question is provided in Figure I.2.

- Incentivized belief elicitation on four factual statements: for each such statement subjects state how likely it is that the fact is true (open field, suggested to provide a numeric answer from 0 to 100).

- Open text box in which subjects are asked to tell us what they think the research was about.[20]

Three of the four factual statements are control questions of the kind included in the baseline survey. One fact pertains the performance of two actual debaters in the Munich Research Open, and had a longer preamble than other belief elicitation questions:

The next question is about the performance of two actual debaters in a different tournament: the Munich Research Open that took place two weeks ago[21]*. We will call them debater A and debater B. Both debaters were representing the Government in the motion that "THBT governments should stop funding scientific programmes that have no immediate benefit for humankind (such as space travel and exploration, human cloning)", but they gave different responses to the factual question in the predebate survey:*
Debater A believed that the statement "More than 10 of the following 15 innovations are a consequence of inventions made in the pursuit of space travel: camera phones, scratch resistant lenses, electric light, CAT scans, LEDs, land mine removal, athletic shoes, penicillin, water purification systems, the internet, home insulations, wireless headsets, baby formula, portable computers" was true with 75% chance. Debater B believed that the same statement was true with 10% chance.
We asked judges to provide a broad measure of each debaters' persuasiveness. Now consider the following statement.

[20] We felt that the alignment question was revealing too much of what the study was about, so to get a better sense of whether subjects understood what hypotheses were being tested with the data collected in predebate and postdebate surveys, in Rotterdam, we decided to move this question to the last postdebate survey.

[21] In Rotterdam. In Munich, the orange text is replaced by "this tournament".

Statement: Debater A obtained a higher persuasiveness score than Debater B in the relevant debate.

Q6$: How likely do you think it is that the above statement is true? ___ % (write a number from 0 to 100)

Figure I.2: Example of Aligment Question in the Endline Survey

Q1: For this question, you can earn up to 5 euros. Consider the following motion, which was debated during the event:

"During periods of national housing shortages, this House would forcibly take ownership of privately owned homes which are not lived in by their owners".

Now consider someone who is strongly in favor of this motion, i.e. someone whose personal views are strongly aligned with the motion. For each statement in the table below, please indicate whether such a person, who is strongly aligned with the motion, is more likely to believe that the statement is true or more likely to believe that the statement is false. For each statement (each row of the table), give your answer by entering one (and only one) cross in the appropriate box.

We will randomly select one of these six statements and pay you based on your selection as follows: You will earn 5 euros for sure if your response is the same as the response that is selected most frequently by all other participants answering the same question.

	Someone aligned with the motion is…		
	… more likely to believe that this statement is **true**	… more likely to believe that this statement is **false**	.. equally likely to believe that this statement is **true or false**
Statement 1: According to the English Housing Survey, the number of second homes in the UK more than doubled between 1995 and 2013			
Statement 2: Under current UK regulation, squatters who live in and maintain unoccupied buildings enjoy protection under the law and can never be evicted without a court order			
Statement 3: According to an academic study published this year, over 5 percent of properties in England and Wales are low-use properties, defined as a property that is not registered as the primary residence of any individual			
Statement 4: According to research by the newspaper the Independent in 2018, more than one third of new-build luxury apartments and houses in Central London lies empty			
Statement 5: Action on Empty Homes* is an NGO supporting a cause that is especially important.			
Statement 6: The Land Is Ours** is an NGO supporting a cause that is especially important.			

*Action on Empty Homes is a UK NGO campaigning for more empty homes to be brought into use for people in housing need. It raises awareness of the waste of long-term empty homes and campaigns for changes to national policy to bring more homes into use..

**The Land Is Ours campaigns peacefully for access to the land, its resources, and the decision-making processes affecting them. Among other things, it advocates 'Use It Or Lose It' programme where empty buildings are forfeit or put on a tax escalator, where the owner can lose title after one year.

I.7. Judge survey

Judges are asked to independently provide individual scores of each debater's overall persuasiveness before filling out the shared score sheet with other judges.

Judges are asked to provide a broad persuasiveness score, on a scale from 1 to 10 where 1 is "Not at all persuasive" and 10 "Extremely persuasive". The original instructions given to judges on how to answer and interpret this question are provided below:

Without discussing with the other judges, please evaluate the persuasiveness of each debater. We consider a debater persuasive, if she would do well at convincing a general audience of her position. Therefore, please provide a broad measure of persuasiveness that captures the quality of arguments as well as speaking ability, body language and any other attribute that makes a speech persuasive to a general audience.

To ensure that the judges provided independent persuasiveness scores, we asked them to fill out these surveys during the debate. Judges on the panel painstakingly take notes of each speech and generally do not interact with each other during the debate. We collected the surveys before any deliberation of the panel took place.

I.8. Enumerator survey

A survey that the enumerator answers during the debate includes the following items:

- A count of the times not speaking debaters try to interrupt the speaker (through Points of Information).
- A subjective rating of how heated each debaters' argumentation is coming across (on a scale from 1 to 5).[22]
- For each of the four facts related to the motion over which we elicit debaters beliefs, and for both the motion related charities, note whether these were mentioned during the debate.

[22] Enumerators were instructed to write down this score for each debater at the end of the speech. They could however revise this score for debaters that acted particularly heatedly during other debaters' speeches.

I.9. Ballot

The ballot is the official module that debating tournament have panels of judges fill out to evaluate a debate. This form includes:

- Name and position of each team in the debate
- Ranking of the four teams in the debate (from First to Fourth, with no possibility for ties)
- Individual speaker scores (on a scale from 50 to 100)

After a debate is over, speakers leave the room to let judges on the panel privately discuss the performance of each debater. This discussion takes approximately 15 minutes during which the arguments presented by each debater are technically analyzed. A technical analysis is particularly relevant to the assignment of individual speaker scores, which are supposed to be assigned an a objective scale that applies to any British Parliamentary performance.[23] The ballot is filled out at the end of this discussion.

[23] An example of such scale can be found at https://debate.uvm.edu.

J. Motion Facts and Charities

Table J.1: Decoy and Control Belief Elicitations for Baseline Survey in Munich

Fact
Decoy questions
1. The US has more nuclear weapons than any other country.
2. A paper recently published in a leading economics journals finds that the decriminalization of prostitution in Rhode Island in 2003 caused reported rape offences to fall by over 20%.
3. A recent randomized controlled trial with almost 3000 social media users finds that individuals that are paid to stay off of Facebook for four weeks watch more TV and are less informed about current events.
4. As measured by the Eurobarometer survey, a majority of Europeans are not interested in receiving information about treatment conditions of farm animals.
5. According to a review published in a prominent public health journal in 2011, nutrition labels are a cost effective intervention to promote healthier diets.
6. A paper published in a leading economics journal in 2009 finds that violent crime increases on days with larger theater audiences for violent movies.
7. According to a 2019 review study in a prominent scientific journal, the well-being of teenagers has a stronger relation with having regular breakfast habits than with the use of digital technologies.
Control questions
1. The corporate income tax is higher in the US than in Finland.
2. In France, government spending was over half of GDP in 2017.
3. More than half of children in the United States were overweight or obese as of 2014 (BMI of 25 or greater).
4. Less than 30% of all Nobel prizes in Chemistry were awarded to U.S. citizens.
5. The PISA is a worldwide exam administered every three years that measures science, reading and math skills of 15-year-olds. In 2015, at least 4 Asian countries were in the top 10 in each category of the exam.
6. According to the UNESCO, the global literacy rate is under 90%.

Note: All decoy questions are included in the baseline survey. For each subject we randomize whether only the first three control question or the last three control questions are included in the baseline survey; the other three questions are included in the endline survey.

Table J.2: Alignment of facts with motions in Munich

	Fact	Alignment predicted by: Authors	Alignment predicted by: Debaters
	This House believes that governments should stop funding scientific programs that have no immediate benefit for humankind (such as space travel and exploration, human cloning)		
Motion 1	A. The European Space Agency's annual budget exceeds 4% of the EU budget	proposition	proposition (65%)
	B. More than six Europeans out of ten agree that space technologies have a role to play in avoiding threats like asteroids, comets, and space debris collisions	Opposition	Opposition (50%)
	C. More than 10 of the following 15 innovations are a consequence of inventions made in the pursuit of space travel: camera phones, scratch resistant lenses, electric light, CAT scans, LEDs, land mine removal, athletic shoes, penicillin, water purification systems, the internet, home insulation, wireless headsets, baby formula, portable computers	Opposition	Opposition (70%)
	D. A study in a leading economics journal in 1998 finds that the majority of R&D spending by the US government goes into wages for scientists, which in turn does little to increase the number of scientists	proposition	proposition (65%)
	This House believes that Western States should permanently revoke the citizenship of citizens who join terrorist organisations		
Motion 2	A. From 2015 to 2017 there were more than 50 separate Islamic terrorist attacks in the EU	proposition	proposition (90%)
	B. According to the UN's Basic Human Right's Reference Guide on the right to a fair trial and due process in the context of countering terrorism, the citizenship of people suspected or proven of having been part of terrorist organisations may never be revoked	Opposition	Opposition (60%)
	C. In Germany, the law governing citizenship already permits to strip those with dual citizenship of their German citizenship if they join a foreign army	proposition	proposition (70%)
	D. The two main perpetrators of the 2015 attacks on the offices of the satirical newspaper Charlie Hebdo had fought with IS in Syria prior to the attack	proposition	proposition (85%)
	This House regrets the EU's introduction of freedom of movement		
Motion 3	A. In a much-cited academic article from 2012, researchers from University College London found that immigration increased wages in the UK, both at the bottom and at the top of the income distribution	Opposition	Opposition (65%)
	B. More than 35% of UK citizens interviewed for the Eurobarometer in 2018 think that the Schengen Area has more disadvantages than advantages for the UK	proposition	proposition (80%)
	C. According to a 2018 paper by researchers from the University of Munich, emigration within Europe positively contributes to innovation in source countries, i.e. countries people emigrate from	Opposition	Opposition (70%)
	D. Less than half of Europeans agree that integration of immigrants has been a success in their local area, city or country	proposition	proposition (65%)
	This House would suspend trade union powers and significantly relax labour protection laws in times of economic crisis		
Motion 4	A. A 2015 survey by the Employment Policies Institute shows that a majority of economists thinks that a USD 15 minimum wage will reduce the number of jobs	proposition	proposition (73%)
	B. According to a study of 21 Eastern European economies published in 2017, members of labor unions are less likely to lose their job during an economic crisis	Opposition	Opposition (46%)
	C. The fraction of the U.S. population that approves of labor unions dropped by more than one third from the mid-50s to 2009, according to the Gallup poll	proposition	proposition (86%)
	D. In a 2005 study of OECD countries, economists from Harvard University and the University of Bonn find that greater labor market flexibility (e.g. due to weaker trade unions) is associated with greater employment	proposition	proposition (59%)
	This House believes that causing deliberate harms to enemy civilians, by the weaker side, is a justified tactic in asymmetrical warfare		
Motion 5	A. Research on the psychological effect of violence against Israeli civilians shows that such violence caused a hardening of attitudes, stronger opposition to political reconciliation with perpetrators, and an increase in support for counter-terrorist measures	proposition	proposition (45%)
	B. Nelson Mandela, who was awarded the Nobel Peace Prize for facilitating South Africa's peaceful transition into democracy, was also the co-founder of the violent paramilitary wing of the African National Congress and was classified as a terrorist by the US until 2008	proposition	proposition (73%)
	C. Willingness to resort to violence was an integral part of the Birmingham campaign, which is widely credited with bringing about civil rights and desegregation in the United States of the 1960s	proposition	proposition (73%)
	D. Studies in political science consistently find that that rebel groups that use indiscriminate violence against civilians are more likely to achieve political goals	proposition	proposition (77%)

Table J.3: Alignment of charitable causes with motions in Munich

	Charitable cause	Alignment predicted by: Authors	Alignment predicted by: Debaters
	This House believes that governments should stop funding scientific programs that have no immediate benefit for humankind (such as space travel and exploration, human cloning)		
Motion 1	**The International Space University** develops the future leaders of the world space community. It encourages the innovative development of space for peaceful purposes: to improve life on Earth and advance humanity into space	Opposition	Opposition (80%)
Motion 1	**The Planetary Society** is the world's largest and most influential non-profit space organization. The society advocates for space and planetary science funding in government, invests in inspiring educational programs, and funds groundbreaking space science and technology	Opposition	Opposition (65%)
	This House believes that Western States should permanently revoke the citizenship of citizens who join terrorist organisations		
Motion 2	**The Active Change Foundation** is based in the UK and provides a holistic approach to neutralising extremism and violence on both an individual and community level. Its chief executive is an outspoken critic of those actors within the UK that favor stripping individuals of their citizenship for being involved with terrorist organisations	Opposition	Opposition (65%)
Motion 2	Human Rights Watch defends the rights of people worldwide. It scrupulously investigates abuses, exposes the facts widely, and pressures those with power to respect rights and secure justice. It has been a vocal defender of the right to citizenship for all people	Opposition	No relation (50%)
	This House regrets the EU's introduction of freedom of movement		
Motion 3	**The European Movement UK** is a grass-roots, independent, pro-European organisation. One of its main goals is to safeguard the freedom of movement made possible by membership of the EU, both for UK citizens who want to travel and work abroad and for citizens of other EU countries who want to come to the UK to work and to live	Opposition	Opposition (85%)
Motion 3	**ACT4FreeMovement** stands for Advocacy, Complaints, Trainings for Freedom of Movement. The organization campaigns for freedom of movement with EU citizens. The goal is to increase the capacity of EU citizens to effectively secure access to and knowledge of their rights, as well as build public awareness and political support for mobile citizen rights	Opposition	Opposition (75%)
	This House would suspend trade union powers and significantly relax labour protection laws in times of economic crisis		
Motion 4	**The European Trade Union Confederation** speaks with a single voice on behalf of European workers to have a stronger say in EU decision-making. It aims to ensure that the EU is not just an economic union but also a Social Europe, where improving the well-being of workers and their families is an equally important priority	Opposition	Opposition (64%)
Motion 4	**The Living Wage Foundation** is a campaigning organization in the United Kingdom, which aims to persuade employers to pay a Living Wage, an independently calculated and recommended minimum wage to cover workers' basic needs	Opposition	No relation (50%)
	This House believes that causing deliberate harms to enemy civilians, by the weaker side, is a justified tactic in asymmetrical warfare		
Motion 5	**The Israel Trauma Center** for Victims of Terror and War is an apolitical organization providing multidisciplinary treatment and support to direct and indirect victims of trauma due to terror and war in Israel	proposition	Opposition (50%)
Motion 5	**Muslim Aid** is an Islamic Charity, which has been actively working in Gaza since 2006. It helps vulnerable people to obtain essentials like food and medical supplies, which are scarce as importing and exporting has been made difficult	proposition	No relation (45%)

Table J.4: Decoy and Control Belief Elicitations for Baseline Survey in Rotterdam

Fact
Decoy questions
1. In 2016, from an estimated pre-war population of 22 million the UN estimates that more than 10 million people have been displaced internally as well as abroad.
2. A paper recently published in a leading economics journals finds that withdrawing legal access to cannabis improves academic performance of foreign university students affected by the policy in the Netherlands.
3. A recent The Lancet article finds that from the 15.6 million abortions that took place in India in 2015 over 10 percent were carried out outside of health facilities using unsafe methods.
4. A paper published in a leading economic journal estimates that juvenile incarceration in the US increases incarceration rates of individuals when they become adults.
5. A large representative survey published in a leading economic journal this year finds that over 30% of Americans would support a policy that allows recipients of kidney transplants to compensate living donors 100âĂŹ000 USD in cash.
6. In the United States, more than half of all guns are sold without background checks.
7. A paper published in a leading economics journal in 2009 finds that violent crime increases on days with larger theater audiences for violent movies.
8. According to a 2019 review study in a prominent scientific journal, the well-being of teenagers has a stronger relation with having regular breakfast habits than with the use of digital technologies.
Control questions
1. Americans drink more alcohol per person than Europeans.
2. More than 30% of Europeans are smokers.
3. The PISA is a worldwide exam administered every three years that measures science, reading and math skills of 15-year-olds. In 2015, at least 4 Asian countries were in the top 10 in each category of the exam.
4. According to the 2015 Eurobarometer, more than 50% of Europeans feel that diversity is sufficiently reflected in the media in terms of religion or beliefs.
5. According to the 2015 Eurobarometer, more than 90% of Europeans say that they would feel comfortable with having a woman in the highest elected position in their country.
6. According to the UNESCO, the global literacy rate is under 90%.

Note: All decoy questions are included in the baseline survey. We included in the survey one more decoy question than we had in Munich to balance for the one fewer motion question (the experiment in Rotterdam covers only four rounds of debate). For each subject we randomize whether only the first three control question or the last three control questions are included in the baseline survey; the other three questions are included in the endline survey.

Table J.5: Alignment of facts with motions in Rotterdam

	Fact	Alignment predicted by: Authors	Alignment predicted by: Debaters
	During periods of national housing shortages, this House would forcibly take ownership of privately owned homes which are not lived in by their owners)		
Motion 1	A. According to the English Housing Survey, the number of second homes in the UK more than doubled between 1995 and 2013	proposition	proposition (72%)
Motion 1	B. Under current UK regulation, squatters who live in and maintain unoccupied buildings enjoy protection under the law and can never be evicted without a court order	Opposition	Opposition (57%)
Motion 1	C. According to an academic study published this year, over 5 percent of properties in England and Wales are low-use properties, defined as a property that is not registered as the primary residence of any individual	proposition	proposition (74%)
Motion 1	D. According to research by the newspaper the Independent in 2018, more than one third of new-build luxury apartments and houses in Central London lies empty	proposition	proposition (74%)
	This House believes that states should aggressively fund geoengineering projects instead of attempting to mitigate the effect of climate change		
Motion 2	A. Germany's experience with renewable energy promotion (i.e. its Renewable Energy Sources Act (EEG)) is often used as a model to be replicated elsewhere. Instead, a widely cited scientific study from 2010 argues that the German government's support of renewables has resulted in massive expenditures (annual feed-in tariffs of over 7 billion euros) that show little long-term promise for stimulating the economy, protecting the environment, or increasing energy security	proposition	proposition (74%)
Motion 2	B. According to recent data from the Climate Action Tracker, more than one third of the surveyed countries are well on track to meet the CO2 emission targets they imposed on themselves under the Paris agreement	Opposition	Opposition (50%)
Motion 2	C. Even the US, which has not supported recent global efforts to fight climate change by means of reducing CO2 emissions, has been enthusiastic in its support for geoengineering projects, as evidenced by its support for the U.N. resolution on geoengineering	proposition	proposition (70%)
Motion 2	D. A 2018 study by two prominent economists from MIT argues that increased investments in geoengineering may also increase efforts to improve clean energy technologies	proposition	proposition (78%)
	This House regrets the decision to let the FARC (i.e. The Revolutionary Armed Forces of Colombia - People's Army) run as a political party.		
Motion 3	A. Shortly after the 2016 peace deal with FARC, Colombia has been experiencing a resurgence of violence. The number of homicides is up by more than 7% in 2018 compared to the previous year	proposition	proposition (92%)
Motion 3	B. In 2016, the Nobel peace prize was jointly awarded to Colombian president Santos and the leader of FARC, Rodrigo Londoño, for their "resolute efforts to bring the country's more than 50-year-long civil war to an end"	Opposition	Opposition (60%)
Motion 3	C. In March 2017, the Colombian government reported that more than 25% of the estimated 6'900 FARC fighters refused to disarm	proposition	proposition (77%)
Motion 3	D. Towards the end of the peace deal negotiations between the Colombian government and FARC, NGOs like Amnesty International and Human Rights Watch as well as the Colombian Conservative party criticized the peace deal for being too lenient on perpetrators of human rights violations	proposition	proposition (90%)
	When tech companies own platform utilities and platform products, this House would break them up.		
Motion 4	A. According to a 2018 survey from the Pew Research Center, over 50% of Americans believe that major tech companies have too much power and influence in today's economy	proposition	proposition (88%)
Motion 4	B. The UK government's digital competition expert panel, chaired by Professor Furman who was chief economic advisor in Obama's presidency, issued a report just two weeks ago rejecting the widely held view that "digital platforms are natural monopolies where only a small number of firms can succeed"	Opposition	Opposition (77%)
Motion 4	C. According to a 2018 survey from the Pew Research Center, over 60% of Americans believe that major tech companies should be more regulated than they currently are	proposition	proposition (54%)
Motion 4	D. A 2018 survey of 1200 sellers on the Amazon platform, conducted by the independent market research firm Feedvisor, finds that over 40% of private sellers on Amazon fear that the company will take away their seller privileges and over 60% of them fear Amazon competing directly with them	proposition	proposition (92%)

Table J.6: Alignment of charitable causes with motions in Rotterdam

	Charitable cause	Alignment predicted by: Authors	Alignment predicted by: Debaters
	During periods of national housing shortages, this House would forcibly take ownership of privately owned homes which are not lived in by their owners)		
Motion 1	**Action on Empty Homes** is a UK NGO campaigning for more empty homes to be brought into use for people in housing need. It raises awareness of the waste of long-term empty homes and campaigns for changes to national policy to bring more homes into use.	proposition	Opposition (52%)
Motion 1	**The Land Is Ours** campaigns peacefully for access to the land, its resources, and the decision-making processes affecting them. Among other things, it advocates 'Use It Or Lose It' programme where empty buildings are forfeit or put on a tax escalator, where the owner can lose title after one year	proposition	Opposition (37%)
	This House believes that states should aggressively fund geoengineering projects instead of attempting to mitigate the effect of climate change		
Motion 2	**Geoengineering Monitor** aims to be a timely source for information and critical perspectives on climate engineering. The goal is to serve as a resource for people around the world who are opposing climate geoengineering and fighting to address the root causes of climate change instead	Opposition	proposition (54%)
Motion 2	**The Environmental Defense Fund** addresses today's most urgent environmental challenges by focusing on the solutions that will have the biggest impact, such as removing obsolete rules that hamper the clean energy market in the U.S. It favors a strategy of reducing CO2 emissions over geoengineering	Opposition	proposition (74%)
	This House regrets the decision to let the FARC (i.e The Revolutionary Armed Forces of Colombia - People's Army) run as a political party.		
Motion 3	**Justice for Colombia** is a British NGO whose primary goal is to give a political voice internationally to Colombian civil society. It has been campaigning to help Jesús Santrich, a lead FARC negotiator of the peace deal who was going to take a seat into parliament in 2018, get justice. The US incarcerated him without providing any evidence of Santrich's crime to the Colombian government	Opposition	proposition (69%)
Motion 3	**Strangers to Peace** is a documentary project of film maker Noah DeBonis which follows the life of ex-FARC guerrillas during their reintegration process. If funded, the film aims to enrich viewer's understanding of a marginalized community through tales of personal and social redemption	Opposition	proposition (63%)
	When tech companies own platform utilities and platform products, this House would break them up.		
Motion 4	**Elizabeth Warren** is a candidate for the President of the United States in 2020. Among other causes, she runs on a platform breaking up big tech firms such as Google and Amazon in a platform component and a supplier component. Donations go towards her campaign for the presidency	proposition	proposition (46%)
Motion 4	**The Open Markets Institute** uses journalism to promote greater awareness of the political and economic dangers of monopolization, identifies the changes in policy and law that cleared the way for such consolidation, and fosters discussions with policymakers and citizens as to how to update America's traditional political economic principles for our 21st century digital society	proposition	Opposition (42%)

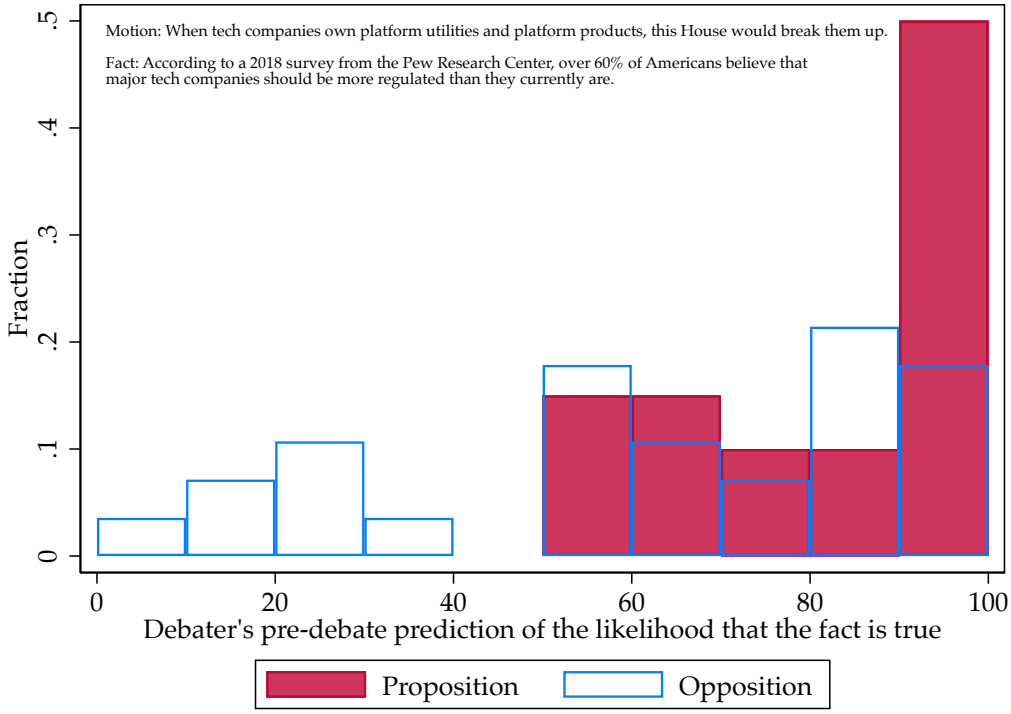
K. Variable Transformations

K.1. Beliefs regarding topics of the motions

The beliefs that we elicit for facts that are relevant to a motion are expected to capture alignment with either side of the motion. While in some cases we expect that someone who is aligned with the proposition is more likely to believe that a fact is true, in some other cases alignment with proposition is expected to be associated with a belief that a fact is false. Figure K.1 illustrates the example of a fact that we were expecting to capture alignment with the proposition. To half of the debaters in Rotterdam we asked this question just before the debate (predebate), and to another half after the debate. As the figure illustrates, in the predebate survey proposition speakers are more likely than Opposition speakers to believe that a survey conducted by the Pew Research Center in 2018 found that over 60% of Americans want major tech companies to be more regulated. The motion of this debate was that "When tech companies own platform utilities and platform products, this House would break them up.".

In order to make belief elicitations comparable across motions, we conduct a normal standardization of the reported belief (separately for each factual question asked at each survey), and we adjust the sign of the standardized belief in such a way that a positive (negative) sign of the standardized outcome captures alignment with the proposition (opposition) side of the motion. While we had a strong prior on the direction of alignment that each fact would capture, to make this sign correction objective and transparent we use the modal alignment predicted by debaters in the endline survey. Our predicted alignment and debaters' are reported in Table J.2 and Table J.5.

Figure K.1: Example of Reported Predebate Beliefs, by Side of the Debate



K.2. Attitudes regarding topics of the motions

Attitudes towards the motion are measured through an allocation of donations that individual debaters can make between a neutral charity – a charity that is used for every motion with an agenda that is relatively orthogonal to alignment with the motion, and a motion charity – a charity that is specific to each motion with an agenda that is expected to be particularly valued by an individual who is aligned with a particular side of the motion.

We had planned to follow a similar procedure as for beliefs to harmonize attitudes across motions. We diverge from that plan for two reasons: First, possible charitable allocations follow a discrete distribution, which clearly strongly violates normality. Second, due to poor phrasing of the mapping alignment question, answers to this question were very noisy and often conflicted with our prediction of alignment of the charity to the motion in ways that are hard to rationalize. In Table J.3 and Table J.6 we list for each charitable cause our predicted alignment with the motion as well as the

debaters'.

Figure K.2: Example of Charity Allocations Chosen Predebate, by Side of the Debate

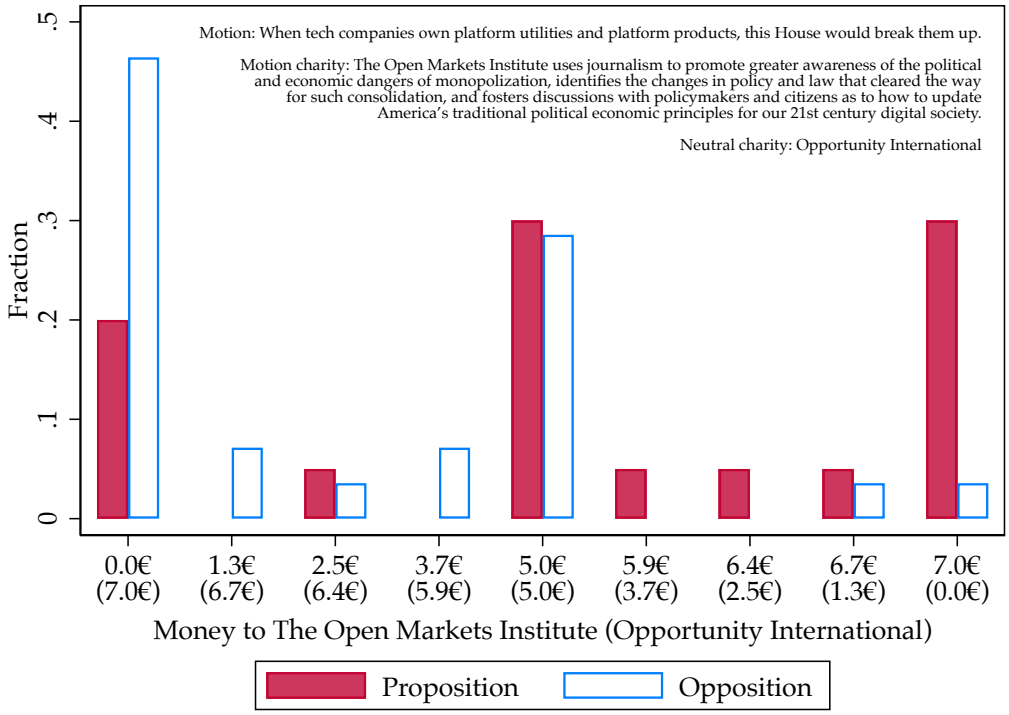


Figure K.2 illustrates an instance in which our prediction of alignment of the charity contradicts the debaters' as captured by the mapping question at endline: We predicted alignment of the motion charity with the proposition, while debaters predicted alignment of the motion charity with the Opposition. In this instance, debaters can choose to allocate money between a neutral charity, Opportunity International, and The Open Markets Institute, an NGO promoting awareness on the dangers of monopolization in the tech sector. From behavioral outcomes elicited predebate, we find that debaters tend to give more to The Open Markets Institute when they *propose* a motion that would break up big tech companies: the alignment that we predicted.

We decide to construct an harmonized ordinal variable that captures alignment with the proposition side of the motion using our predicted alignments. Such variable, for each question, simply takes the nine categories of increasing monetary amounts that are given to the baseline charity (and subtracted to the motion charity), and adjusts

the order in such a way that if the motion charity is aligned with the proposition (opposition) the order is reversed (kept as it is).

L. Mapping Pre-Analysis Plan to Paper

This study was pre-registered the week prior to the first debating competition. Relative to the pre-registered sample size and survey items we report the following substantial changes:

- We expected to have 104 teams of debaters across the two tournaments. We end up with 4 teams fewer in Rotterdam due to last minute cancellations.
- Dropped debaters' attractiveness score from the enumerator survey.

Pre-registration included a pre-analysis plan. In this appendix we spell out the analysis planned and the results of the planned analyses, which are sometimes replaced in the main paper with analyses that are now considered superior by the authors for statistical and expositional reasons.

L.1. Pre-registered Hypotheses

We formulated a first set of hypotheses deriving from strategic self-deception, and a second set of hypotheses on the role of debating for belief convergence.

L.1.1. Self-Persuasion

Hypothesis 1. *Debaters predebate factual beliefs are biased in the direction of their persuasion goal.*

The pre-registration specifies how beliefs are standardized and sign-adjusted to obtain a metric $b_{i,m}$ and conduct a fixed effects panel analysis to identify the causal effect of persuasion goals. Sign adjustment is determined by Endline responses to mapping questions in which, for each factual question and charity related to the motion, we ask subgroups of debaters to predict what the majority of respondents would believe the alignment to be between proposition/opposition/No alignment. When at least 51 percent of debaters correctly predict the reported modal alignment, we use that alignment to determine the sign adjustment of standardized beliefs.[24] We test the hypothesis by

[24] If the alignment of a belief distribution is proposition (opposition), then we change the sign of standardized beliefs for opposition (proposition) speakers.

estimating the following fixed effects model

$$b_{i,m} = \alpha_i + \beta proposition_{i,m} + \delta_m + \epsilon_{i,m}$$

in which δ_m are motion fixed effects, and $\epsilon_{i,m}$ is the error term allowing for a team component. Column (1) of Table 3 reports the estimated β from such model that confirms the original hypothesis, along with multiple additional specifications to assess the robustness of the result.

Hypothesis 2. *Debaters predebate attitudes are biased in the direction of their persuasion goal.*

The pre-registration specifies a similar standardization and sign-adjustment for our measure of attitudes, and a similar analysis of the causal effect of persuasion goals. Here we need to deviate from the pre-analysis plan. First, in the pre-analysis plan we failed to account for the ordinal nature of our attitudinal outcome, which does not warrant standard normalization. Therefore, we decide to conduct sign-adjustment, but not standardization. Second, we failed to adequately formulate the endline alignment question for charities. This led to puzzling alignment predictions presented in Table J.3 and Table J.6, that often conflict with our own prediction of alignment. Therefore, we decide to use the prediction of alignment formulated by us – that guided the choice of motion related charities in the first place. We test the hypothesis by estimating the following random effects model for the latent variable underlying our sign adjusted attitudinal outcome $a_{i,m}$:

$$\tilde{a}_{i,m} = \alpha_i + \beta proposition_{i,m} + \gamma X_i + \delta_m + \epsilon_{i,m}$$

in which X_i includes all socio-demographic and experience controls, δ_m are motion fixed effects, and $\epsilon_{i,m}$ is the error term allowing for a team component. Random effects models are used because standard fixed effects models for ordinal categorical variables are under-identified. Column (1) of Table D.4 reports the estimated β from such model without controls, column (2) reports estimates from the model with controls. Both estimates confirm the original hypothesis. We also report additional results from Chamberlain-like fixed effects estimators (column (3)) to assess the robustness of the result.

Hypothesis 3. *Debaters have more confidence in the arguments favoring their side than in the other side's arguments.*

The pre-registration specifies a straightforward fixed effects regression model to test this hypothesis using the prediction that the majority of debates in parallel debates will be won by proposition teams $c_{i,m}$:

$$c_{i,m} = \alpha_i + \beta proposition_{i,m} + \delta_m + \epsilon_{i,m}$$

in which δ_m are motion fixed effects, and $\epsilon_{i,m}$ is the error term allowing for a team component. Column (1) of Table 5 reports the estimated β from such model that confirms the original hypothesis, along with multiple additional specifications to assess the robustness of the result.

Hypothesis 4. *When persuasion goals are more aligned with private beliefs at baseline, debaters obtain higher persuasiveness ratings by judges.*

The pre-registration specifies a fixed effects regression model to test the correlation between *baseline alignment* and *persuasiveness*, where baseline alignment is defined as standardized and sign-adjusted baseline belief above 0 (below 0) if for speakers that will be assigned to proposition (opposition), and *persuasiveness* as the panel average of the independent scores that each judges gives for broad persuasiveness of speaker's performance $P_{i,m}$:

$$P_{i,m} = \alpha_i + \beta(\mathbb{1}_{y_{i,m}^{baseline} \geq 0} \mathbb{1}_{proposition_{i,m}} + \mathbb{1}_{b_{i,m}^{baseline} < 0} \mathbb{1}_{Opposition_{i,m}}) + \delta_m + \epsilon_{i,m}$$

in which δ_m are motion fixed effects, and $\epsilon_{i,m}$ is the error term allowing for a team component. Column (1) of Table E.1 reports the estimated β from such model that lends no support for such hypothesis.

L.1.2. Debating and Convergence

Hypothesis 5. *Postdebate attitudes are less dispersed than predebate attitudes.*

The pre-registered analysis proposes to assess whether an individual level measure distance from the median ordinality of sign-adjusted bundle $d(a)_{i,m,p,s}$ is lower at post-

debate than it is predebate.[25] We test for convergence of attitudes in the following fixed effects regression framework:

$$d(a)_{i,m,p,s} = \alpha_i + \beta Predebate_{i,m,p} + \delta_p + \delta_m + \epsilon_{i,m,p,s}$$

in which δ_m are motion fixed effects, δ_p are charity-pair fixed effects, and $\epsilon_{i,m,p,s}$ is the error term allowing for a team component. We would say that there is convergence in attitudes from predebate to postdebate if β were positive and significant. Column (9) of Table C.2 reports the estimated β from such model that lends no statistically significant support for such hypothesis.

Hypothesis 6. *Postdebate factual beliefs are less dispersed than predebate and baseline beliefs.*

The pre-registered analysis proposes to assess whether an individual level measure distance from the median ordinality of sign-adjusted bundle $d(b)_{i,m,q,s}$ is lower at postdebate than it is at predebate and baseline.[26] We test for convergence of beliefs from predebate to postdebate in the following fixed effects regression framework:

$$d(b)_{i,m,q,s_1} = \alpha_i + \beta_1 Predebate_{i,m,p} + \delta_p + \delta_m + \epsilon_{i,m,p,s_1}$$

and for convergence of beliefs from baseline to postdebate in the following fixed effects regression framework:

$$d(b)_{i,m,q,s_2} = \alpha_i + \beta_2 Predebate_{i,m,p} + \delta_p + \delta_m + \epsilon_{i,m,p,s_2}$$

in which $s_1 \in \{Predebate, Postdebate\}$, $s_2 \in \{Baseline, Postdebate\}$, δ_m are motion fixed effects, δ_p are charity-pair fixed effects, and $\epsilon_{i,m,p,s}$ is the error term allowing for a team component. We would say that there is convergence in attitudes from Predebate (Baseline) to Postdebate if β_1 (β_2) were positive and significant. Column (3) and (5) of Table C.2 report the estimated β_1 and β_2 from such models, respectively. The estimate of β_1 rejects the null hypothesis of convergence in a one-sided t-test, and provides evidence that beliefs in fact polarize from Baseline to Postdebate. The estimate of β_2 is qualitatively in line with convergence, but not statistically different from zero.

[25] For a sign-adjusted distribution of monetary donations to charitable organizations taking place at survey s of motion m for pair of charities p, $d(a)_{i,m,p,s} = |a_{i,m,p,s} - median(a_{i,m,p,s})|$.

[26] For a distribution of beliefs elicited at survey s of motion m for factual question q, $d(b)_{i,m,q,s} = |b_{i,m,q,s} - median(b_{i,m,q,s})|$.

Hypothesis 7. *Postdebate factual beliefs are less dispersed than predebate and baseline beliefs, looking at only those debaters who got to argue their baseline position.*

The plan for testing this hypothesis was to exactly replicate the analysis for Hypothesis 6, including in the analysis only the distance in beliefs from the median belief for debaters that have at baseline standardized and sign-adjusted beliefs aligned with their persuasion goal. Column (6) and (7) of Table C.2 report the estimated β_1 and β_2 from the estimates of the regression models for such sub-sample, respectively. The estimate of β_1 rejects the null hypothesis of convergence in a one-sided t-test, and provides evidence that beliefs in fact polarize from Baseline to Postdebate. The estimate of β_2 is qualitatively in line with convergence, but not statistically different from zero.

Hypothesis 8. *Heated debates are less likely to favor the formation of a consensus around facts and attitudes, and may even increase polarization.*

The plan for testing this hypothesis was to exactly replicate the analysis for Hypothesis 5 and Hypothesis 6, including in regression analysis an interaction term between the timing of the elicitation (the survey dummy) and a binary indicator for whether a debater was heated or not.