

## Self-Persuasion: Evidence from Field Experiments at International Debating Competitions<sup>†</sup>

By PETER SCHWARDMANN, EGON TRIPODI, AND JOËL J. VAN DER WEELE\*

*Laboratory evidence shows that when people have to argue for a given position, they persuade themselves about the position's factual and moral superiority. Such self-persuasion limits the potential of communication to resolve conflict and reduce polarization. We test for this phenomenon in a field setting, at international debating competitions that randomly assign experienced and motivated debaters to argue one side of a topical motion. We find self-persuasion in factual beliefs and confidence in one's position. Effect sizes are smaller than in the laboratory, but robust to a one-hour exchange of arguments and a tenfold increase in incentives for accuracy. (JEL C93, D12, D72, D83, D91, I23)*

*It might be plausibly maintained that in almost every one of the leading controversies, past or present, in social philosophy, both sides were in the right in what they affirmed, though wrong in what they denied; and that if either could have been made to take the other's views in addition to its own, little more would have been needed to make its doctrine correct.*

—John Stuart Mill, *An Essay on Samuel Taylor Coleridge*

When asked to defend a particular point of view, people shift their private opinions in order to align them with the new arguments. Decades of research in the experimental laboratory have demonstrated this effect by having subjects argue in a randomly selected role (Janis and King 1954; O'Neill and Levings 1979), make counter-attitudinal statements (Festinger and Carlsmith 1959; Elliot and Devine

\*Schwardmann: Carnegie Mellon University (email [schwardmann@cmu.edu](mailto:schwardmann@cmu.edu)); Tripodi: University of Essex and JILAEE (email [egon.tripodi@essex.ac.uk](mailto:egon.tripodi@essex.ac.uk)); van der Weele: University of Amsterdam and Tinbergen Institute (email: [j.vanderweele@uva.nl](mailto:j.vanderweele@uva.nl)). Stefano DellaVigna was the coeditor for this article. We are grateful to the coeditor, a second coeditor and three anonymous referees for comments that greatly improved the paper. We also received valuable comments from numerous seminar participants and thank Saskia Bergmann, Andrea D'Souza, Yimin Ge, and Lena Martinovic as well as many other members of the debating community for their collaboration. We especially benefited from advice and support from Huyen Nguyen. Giacomo Manferdini, Hannah Rehwinkel, and Irene van Rooij provided excellent research assistance and we are grateful to numerous local field assistants for their support with the data collection. 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), the European University Institute and the Russell Sage Foundation is gratefully acknowledged. This study was approved by the Economics & Business Ethics Committee at the University of Amsterdam (ref. 20181017051000) and by the Ethics Sub Committee 1 at the University of Essex (ETH2021-1082), and was preregistered on the AEA RCT Registry (AEARCTR-0003922). Survey templates and replication files are available at <https://osf.io/u7ekr/>. We are convinced that this paper contains no errors.

<sup>†</sup>Go to <https://doi.org/10.1257/aer.20200372> to visit the article page for additional materials and author disclosure statements.

1994), advise others to buy inferior products (Chen and Gesche 2017; Gneezy et al. 2020), or convince others of their own ability (Smith, Trivers, and von Hippel 2017; Schwardmann and van der Weele 2019; Soldà et al. 2019). In experimental courtroom or bargaining settings where subjects argue a randomly selected side of a case, they adopt self-serving views of the underlying evidence that limit their willingness to compromise (Thompson and Loewenstein 1992; Babcock, Issacharoff, and Camerer 1995; Engel and Glöckner 2013).

This effect of persuasion goals on beliefs and attitudes, which we call self-persuasion, can have important implications. It limits the potential of communication in resolving costly disagreements in pretrial legal bargaining and labor disputes (Babcock and Loewenstein 1997) and helps explain why political polarization and partisanship are at record levels (Iyengar et al. 2019; Gentzkow 2016), even though the internet has made it cheaper than ever to communicate with people that differ in background and ideology. It has also inspired theories about motivated cognition (Taber and Lodge 2006; Bénabou and Tirole 2016) and the social origins of reasoning (von Hippel and Trivers 2011; Mercier 2011).

Given the robust self-persuasion effect in the laboratory and its potential implications, an important question is whether the phenomenon carries over to more natural settings where it may be reduced by expertise and higher stakes or drowned out by contextual factors (List 2003, 2006; Levitt and List 2007). The key difficulty in the field is to disentangle the causality between private views and persuasion goals. We confront this identification challenge by conducting preregistered field experiments at international debating competitions. The competitions feature parliamentary-style debates on topical motions like the freedom of movement in the European Union, investment in geoengineering, and the regulation of big technology companies. Because debaters are randomly assigned to persuasion goals just before the debate, comparing the beliefs and attitudes of the two sides yields clean estimates of self-persuasion. We survey debaters pre- and post-debate to measure three separate outcome variables: beliefs about motions-related facts, confidence in the relative strength of each debating position, and attitudes toward motion-related charities. Across two offline and two online competitions, 473 debaters from 58 countries filled in a total of 4,854 surveys relating to 19 different motions.

Several features of the debating competitions make them particularly well suited for our purpose. The setting is natural to debaters who participate in several similar competitions each year, are skilled at the task of persuading and are highly motivated to be persuasive. Performance in these prestigious international tournaments is scored by experienced and impartial adjudicators and confers status within the debating community. These incentives resemble those of professionals in politics and law, and many famous politicians and lawyers honed their skills in competitive debating.<sup>1</sup> At the same time, the competitions allow us to maintain a high degree of control. We are able to incentivize our elicitations to assure that they reflect true beliefs and attitudes. We also precisely time our surveys to measure self-persuasion

<sup>1</sup>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 Speech and Debate Association (<https://www.speechanddebate.org/alumni>) or <http://worldcollegiatefriends.blogspot.com/p/famous-former-debaters.html> for partial lists of famous former debaters.

pre-debate, i.e., after debaters know their persuasion goals but right before the start of the debate, as well as post-debate, i.e., right after the hour-long exchange of arguments.

We find strong evidence for self-persuasion before the start of the debate. First, debaters are more likely to believe that a factual statement is true if the statement strengthens an argument supporting their position. Second, debaters become more confident about the relative strength of their debating position, as measured by the subjective probability that teams arguing the same side of the motion in other debates will win. In our third outcome measure, a monetary allocation task between motion relevant charities, we find only weak evidence for self-persuasion.

We provide two benchmarks for the size of these effects. First, we find that the size of our estimates from the field is about 21 percent of the average effect in the laboratory studies. This may have to do with the selection of particularly polarizing issues in lab studies or with publication bias in the literature (Andrews and Kasy 2019; DellaVigna and Linos 2020). Second, we contrast the polarizing effect of self-persuasion with that of political partisanship, two effects that coincide in most settings but can be separated in this context. We find that the polarization predicted by the political attitudes of debaters is smaller than the self-persuasion gap on two out of three outcome measures.

Next, we investigate the impact of the debate itself on the polarization induced by self-persuasion. *Ex ante*, both an increase and a decrease in polarization are plausible outcomes. On the one hand, the debate gives both sides access to the same arguments, so impassionate reasoners should converge on the same conclusions. Previous research shows that explicit prompts to focus on the opposing side of the argument can lead to more balanced argumentation (Lord, Lepper, and Presto 1984; Perkins 2019). Moreover, Levy (2021) finds that online exposure to counter-attitudinal news reduces affective polarization. On the other hand, the act of debating may reinforce persuasion goals and lead to further self-persuasion. In line with this idea, exposure to opposing views has been shown to harden preexisting views and attitudes both in the laboratory (Lord, Ross, and Lepper 1979; Taber and Lodge 2006) and on Twitter (Bail et al. 2018).

Comparing pre- and post-debate outcomes, we do not find evidence for a decrease in polarization. Two of our three outcomes, i.e., confidence in one's position and attitudes toward charities, show a slight increase in polarization post-debate, while factual beliefs show a slight decrease. However, none of the effects are sizable, resulting in very similar self-persuasion effects post- and pre-debate, although we can not rule out modest increases or decreases. This is not because debaters are ignoring opposing arguments: we find that they report a higher tally of arguments for the other side than before the debate. We also find evidence for convergence among the partisan polarization of debaters, a dimension of disagreement that is not reinforced during the debate.

We investigate a number of additional aspects of self-persuasion. First, we find that more experienced and more successful debaters self-persuade less in factual beliefs, but not in confidence. Thus, while experience decreases the bias on some dimensions, it does not eradicate it. Second, we vary the cost of self-persuasion by implementing a tenfold increase in the incentives for accuracy on the belief elicitations. We do not find evidence that higher incentives impact self-persuasion

in factual beliefs, and while they reduce self-persuasion in the confidence in one's position, this effect is not statistically significant. Third, we find that expert adjudicators predict self-persuasion, generating an intriguing contrast between the persistence of the bias and the awareness of it in the community. The predictions also capture some of the substantial heterogeneity in the effect across motions, survey questions, and outcome variables, suggesting that contextual information may be used to anticipate polarization (DellaVigna and Pope 2018). Finally, we provide some insights into the "how" and "why" of self-persuasion. We find that self-persuasion is partly driven by a biased investment in arguments for one's own side, suggesting the phenomenon results from a failure to account for this biased investment (Thompson and Loewenstein 1992). We also find that debaters whose beliefs happen to be more aligned with their randomly assigned persuasion goal receive higher scores in the debate, in line with the existence of instrumental benefits of self-persuasion.

Our field experiments add to a small set of papers that study motivated reasoning in natural settings, which emphasize motives for belief distortion other than persuasion goals. Di Tella, Galiati, and Schargrotsky (2007) show that the quasi-random assignment of property rights to squatters results in heightened pro-market beliefs, which is suggestive of motivated reasoning. Oster, Shoulson, and Dorsey (2013) and Ganguly and Tasoff (2016) show that some people avoid getting tested for serious diseases, and link their findings to models of self-deception in the service of reduced anxiety. Finally, Huffman, Raymond, and Shvets (2019) show that managers distort their memories of past performance feedback to maintain overconfident beliefs.

Our results show that self-persuasion is a highly robust phenomenon outside of the laboratory: it occurs among highly motivated subjects with years of debating experience and is not eliminated by an hour-long exchange of arguments or a tenfold increase in incentives for accuracy. While the field effects are smaller than effects in the lab, we still find a sizable effect of about a quarter of a standard deviation for two of our three outcome variables. These findings suggest that self-persuasion is a significant and resilient contributor to polarization and disagreement on policy issues.

## I. Experimental Setting

Competitive debating has a long tradition as a platform for civil discussion on important and controversial topics. The format is based on parliamentary practices and features the random assignment of debaters to positions on a given issue. Therefore, in contrast to debates between experts or politicians, competitive debates require participants to take a stance that may not correspond to their original views. Today, many universities have debating societies that organize local or international tournaments, the most prestigious of which include the North American, European, and World Championships.

We conducted field experiments at four international debating competitions. The *Munich Research Open* and the *Erasmus Rotterdam Open* took place in the spring of 2019. We then collaborated with the *Amsterdam Open* in October 2020 and the *London School of Economics (LSE) Open* in February 2021 for a second wave of data collection. Due to the COVID-19 pandemic, these last two tournaments took

place online. Like most international tournaments, all four competitions follow the procedures of British Parliamentary debating. Debates feature two teams of two debaters each in the proposition, who argue in favor of a given motion, and two teams of two debaters each in the opposition, who argue against the motion. Debaters are randomly assigned either to the proposition or the opposition of a debate and to a speaking order.<sup>2</sup> They are not allowed to research the motion's topic and have only 15 minutes to prepare their speeches.

The motions are designed by "chief adjudicators," who tend to be members of the debating community that are highly regarded for having won or having been adjudicators of the final stages of continental and world championships. Chief adjudicators aim at designing motions that are balanced, with reasonable arguments on both sides, and that pertain to topical issues in politics, such as immigration, climate change and the regulation of new technology.

Debaters at our tournaments are predominantly undergraduate and graduate students that are members of debating societies, but also include former students that have entered professional careers of various kinds. Most debaters participate in regular meetings of their debating societies and travel to many tournaments each year. They tend to have strong analytical skills, an ability to think on their feet and a breadth of knowledge.<sup>3</sup> The most illustrious debaters at our tournaments have had successful runs at the European and World Championships.

Our four tournaments were organized yearly by university debating societies. The typical costs of organizing a tournament include the compensation of the adjudicators and the technical team, location rental, and catering. These costs are usually covered through registration fees and external sponsorship. We were able to recruit these prestigious tournaments for our research by offering sponsorship that covered a significant share of the tournament's organization costs. Moreover, our survey payments to debaters helped attract a larger number of teams, by lowering the effective registration fee for debaters. In order to remain attractive to elite debaters, it was essential that our research design had an especially light touch and did not interfere with standard debating rules and procedures. The only significant departure from standard protocols we needed to negotiate was the administration of our crucial pre-debate survey after the debaters had prepared and just before they started speaking.

### *A. Research Design*

We collected data in the five preliminary rounds of each competition, except in Rotterdam where we skipped the fifth round for logistic reasons. Each debater participates in all preliminary rounds, except for rare case where someone feels unwell or particularly uncomfortable with a motion. Debaters answered three main surveys: a baseline survey at the beginning of the tournament, a pre-debate survey right after preparation time and right before the start of the debate, and a post-debate survey right after the end of each debate but before adjudicators' ratings are announced.

<sup>2</sup>Online Appendix Table A.1 describes the eight roles in a debate and the order in which debaters speak.

<sup>3</sup>Further discussion of the characteristics of debaters that take part in this format can be found on the website of the American Parliamentary Debate Association: <http://apda.online/about/>.

Pre-debate and post-debate surveys are collected in each of the preliminary rounds. The random allocation of persuasion goals allows us to identify self-persuasion by comparing the outcomes of the pre-debate survey between the two sides of the debate. We then measure the same outcomes post-debate to study how debates affect the polarization due to self-persuasion.

*Outcome Variables.*—The main outcomes collected in our surveys are the following.

- **Factual Beliefs:** We elicited the probabilistic beliefs in factual statements related to the motion. Factual statements were constructed such that, if they were true, one side of the debate would find them “convenient” in support of their arguments. To interpret this belief as a measure of factual belief alignment with the proposition, we keep the raw reported belief for facts that favor the proposition and compute the complement belief for facts that favor the opposition. Thus, higher values of the resulting outcome reflect a stronger alignment with the proposition.
- **Confidence in Proposition:** We elicited the subjective probability that a majority of parallel debates in the round (excluding the debater’s own debate) will be won by the proposition side of the debate. In excluding the debater’s own debate, we elicit the confidence in the strength of the case for the proposition, rather than confidence in the own ability to persuade. Higher values of this outcome thus capture the perceived advantage of the persuasion goal of the proposition, independent of speakers’ confidence in their own ability.
- **Revealed 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. Debaters choose their preferred allocation out of nine possible allocations, displayed in order from least favorable to the neutral charity to most favorable. To interpret this choice as a measure of attitudinal alignment with the proposition, we keep the raw order of the debater’s choice when the motion-specific charity is aligned with the opposition and invert the order when the motion-specific charity is aligned with proposition. Higher values of the resulting outcome capture alignment with the proposition.

For concreteness, consider the following example of a motion and the associated factual statement, charity and confidence question.

*Example of Motion:* This house regrets the European Union’s introduction of the freedom of movement.

*Factual Statement:* More than 35 percent of UK citizens interviewed for the Eurobarometer in 2018 think that the Schengen Area has more disadvantages than advantages for the United Kingdom.

*Charity:* ACT4FreeMovement campaigns for freedom of movement with EU citizens. Its 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.



*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 first.

In addition to our three main outcomes, we elicited several other variables. The baseline survey collected background information of debaters, including experience with debating, past achievements, political orientation, and basic sociodemographics.<sup>4</sup> In the pre-debate and (online) post-debate surveys, we also asked debaters to report the number of arguments available for each side of the debate. Among these arguments, we asked them to indicate how many can be considered very strong.

*Incentives for Accuracy.*—We incentivized factual beliefs and confidence measures with a binarized quadratic scoring rule that paid in lottery tickets. Depending on their report  $r \in [0, 100]$  and the objective binary answer  $R \in \{0, 1\}$ , subjects receive a lottery ticket that paid off a monetary prize of  $M$  with the following winning probability:

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

Our general instructions in the baseline survey used both the mathematical equation, a simple quantitative example, and an intuitive explanation that truthful reporting optimizes the likelihood of winning the monetary prize (see online Appendix D).<sup>5</sup>

At the offline tournaments the belief elicitation prize  $M$  was 30 euros. At the online tournaments, we varied incentives between a small prize of 5 euros and a large prize of 50 euros, randomized at the team-round level.<sup>6</sup> This variation in the accuracy bonus  $M$  allowed us to investigate whether a higher cost of self-persuasion reduces its prevalence. At the end of the debate, we randomly selected one report incentivized with price  $M$  to be paid out to subjects, i.e., one report in the offline tournament and two reports in the online tournament.

For the attitude elicitation, subjects allocated up to 10 euros 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 for this choice on the subjects' behalf.



















*Survey Overview.*—Table 1 summarizes the timing and collection of outcomes in each survey, highlighting slight differences in implementation between the offline and online tournaments. The baseline survey takes place on the first day of the tournament before the start of preliminary rounds. In each round, the pre-debate survey is collected between the end of preparation time and the start of the debate, and the



<sup>4</sup>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.

<sup>5</sup>In theory, this randomized quadratic scoring rule is 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. In the online tournaments, the formula and quantitative example were available upon clicking a box.

<sup>6</sup>The level of randomization was chosen in order to maximize the salience of differences in incentives. See online Appendix D for further detail.

TABLE 1. DEBATER SURVEYS: CONTENTS AND TIMING

	Baseline	Pre-debate	Post-debate
Background	 		
Factual beliefs	 	 	 
Confidence		 	
Revealed attitudes		 	 
Arguments for/against		 	
Timeline			
		Motion announced Prep (15 min)	Debate (1 hour) Rating announced
	Baseline survey	Pre-debate survey	Post-debate survey

Notes:  denotes offline and  online tournaments.

post-debate survey is collected right after the debate. Only after the post-debate survey is over do debaters receive the ranking from adjudicators.

From all tournaments we also collected the “ballot,” the official score sheet summarizing the deliberation of adjudicators. This includes two main performance measures: the ranking of teams and the debaters’ individual speaking performance. We also collected an adjudicator survey. At the offline tournaments this survey asked adjudicators to provide their own independent rating of each debater’s persuasiveness post-debate. At the online tournaments we instead surveyed adjudicators pre-debate and asked factual belief questions, their predictions of the average response on each side of the debate, and their prediction of the average allocation of charitable donations for our revealed attitude questions on each side of the debate.<sup>7</sup> We incentivized adjudicators’ responses at the online tournament using the same scoring rule as for debaters and randomly selected one question to be payoff relevant at the end of the tournament.

The content of all surveys is described in greater detail in online Appendix D, where Tables D.1 and D.2 provide all motions, factual statements and charities used for our elicitations.

*Preregistration.*—The first round of data collection was preregistered on the AEA RCT registry (AEARCTR-0003922) with a preanalysis plan. In a longer working paper, we execute the preanalysis plan and describe minor deviations (Schwardmann, Tripodi, and van der Weele 2019, pp. 86–90). Additional hypotheses to be tested in the second wave of data collection, targeted sample size, and the alignment of factual questions and charities were also preregistered as amendments to the plan.<sup>8</sup>

<sup>7</sup>The offline survey was designed to capture additional measures of persuasiveness. We only elicited this offline, as the post-debate survey interfered with adjudicators’ deliberation about ballot scores and was difficult to administer online.

<sup>8</sup>Amendments were submitted on November 27, 2020 (one day before the Amsterdam Open), and February 5, 2021 (two days before the LSE Open).



### B. Survey Versions and Administration Procedures

Before the tournament, we coordinated with the chief adjudicators to converge on a final set of motions for the debate. For each motion, we developed several factual questions and motion-related charities, and varied the order in which factual questions and charities were presented to random subgroups of debaters.<sup>9</sup> The use of multiple questions in different orders means that debaters are not asked the same factual question twice. This helps rule out that results are driven by a desire to provide consistent answers to repeated questions and reduces concerns about experimenter demand effects. It also implies that no result depends on the answer to a single question or the order in which questions were asked. Moreover, since baseline and pre-debate questions were different both within and across subgroups, participants could not be influenced through discussion of the answers with others.

We administered the baseline after registration and introductory remarks by the organizers and research team, and shortly before the announcement of the first round motion. The full survey took about 25 minutes and was the same for all participants, except for the factual questions that related directly to the in-round motions, which were randomized. In offline tournaments, all debaters completed the survey in a single hall under the supervision of several enumerators ready to answer clarification questions. In the online tournaments, we administered the baseline in the virtual debate rooms where each enumerator was in charge of supervising eight debaters to maintain a high level of control and supervision.

In each debating round, the motions were announced in the central meeting room, after which debaters made their way to the assigned debating room. In the online tournaments, the central announcements took place via the app Discord, while debates occurred on Google Meet (Amsterdam Open) or Zoom (LSE Open). After the preparation period, enumerators distributed the pre-debate survey in the separate debating rooms. Debaters were given up to five minutes to answer the survey and enumerators ensured that they did not use this time to prepare for the debate. At the beginning of the debate enumerators also distributed the adjudicator survey which was collected along with pre-debate surveys at the online tournaments and after the debate at the offline tournaments.

After the pre-debate survey, the adjudicators opened the debate. The debate lasted about an hour and was attended by the enumerators. Once the adjudicators declared the end of the debate, enumerators distributed the post-debate survey, which was to be answered by debaters within five minutes.

### C. Sample Characteristics and Balance

On average, our sample has spent more than two years in debating, has qualified for more than seven quarterfinals of an international tournament, is about 21.5 years old, and tends to hold a relatively liberal ideology. The share of debaters that identify as women is 34.8 percent. The cultural background of participating debaters is fairly diverse: 61 percent are from Europe (including Russia and

<sup>9</sup>See online Appendix D for a detailed description.

Turkey), 24 percent from Asia, 8 percent from North America, and 7 percent are either from Israel, Latin America, Africa, or Australia. Only 15 percent of participants are nationals of the country where the tournament is hosted. In online Appendix Table A.2 we show balance of individual characteristics and baseline alignment with the proposition across debaters with different persuasion goals.

## II. Main Results

### A. Overview

We start with an overview of the dynamics of our main outcome variables (factual beliefs, confidence, and revealed attitudes) across three points in time: at baseline, pre-debate and post-debate. The graph in Figure 1 displays the mean and 95 percent confidence intervals of each outcome for both proposition and opposition debaters. Histograms for the distribution of alignment for our three outcome variables pre- and post-debate are provided in online Appendix Figure A.1.

The first panel of Figure 1 shows the dynamics of factual beliefs, i.e., debaters' subjective probability that a state that favors the proposition is true. More information on which answers to factual statements, or states, are favorable to the proposition is provided in online Appendix Table D.1. In the baseline survey, before debaters know the motion or before they are assigned to a side of the debate, factual beliefs of proposition and opposition debaters are identical. This implies that the randomization was successful. The pre-debate survey, taken after debaters prepared their arguments for 15 minutes, delivers evidence for self-persuasion: a 7 percentage point gap in factual beliefs opens up between proposition and opposition debaters. Self-persuasion persists in the post-debate survey, although the gap narrows to about 5 percentage points.

The second panel of Figure 1 displays the dynamics of debaters' confidence in the strength of the proposition side of the debate. We see a clear gap of about 6 percentage points pre-debate, which widens to about 8 percentage points post-debate.

The third panel shows the results for revealed attitudes, measured by how much money the debater allocates to the charitable cause that is more aligned with the proposition. Recall that allocations were made along a concave budget constraint in nine discrete steps, so we use these steps as measurement units. We find a small pre-debate gap of about 5 percent of a discrete donation step. This gap then increases in the post-debate survey to about a quarter of a donation step.

Overall, Figure 1 shows clear evidence for self-persuasion both pre- and post-debate. The figure also hints at an unanticipated pattern: average pre-debate and post-debate outcomes are tilted toward the proposition. This does not affect the interpretation of the main results because identification relies on between-subject comparisons of proposition alignment, conditional on the exact question or motion. However, one may wonder whether the asymmetry is due to a stronger self-persuasion effect among proposition debaters, or is driven by the characteristics of the motions or questions. To disentangle this, we elicit factual beliefs of a "control group"—the adjudicators of the online debate who knew the motions and answered the same questions as the debaters, but were not invested in the outcome. Using their answers as a benchmark, we find that proposition and opposition debaters self-persuade to

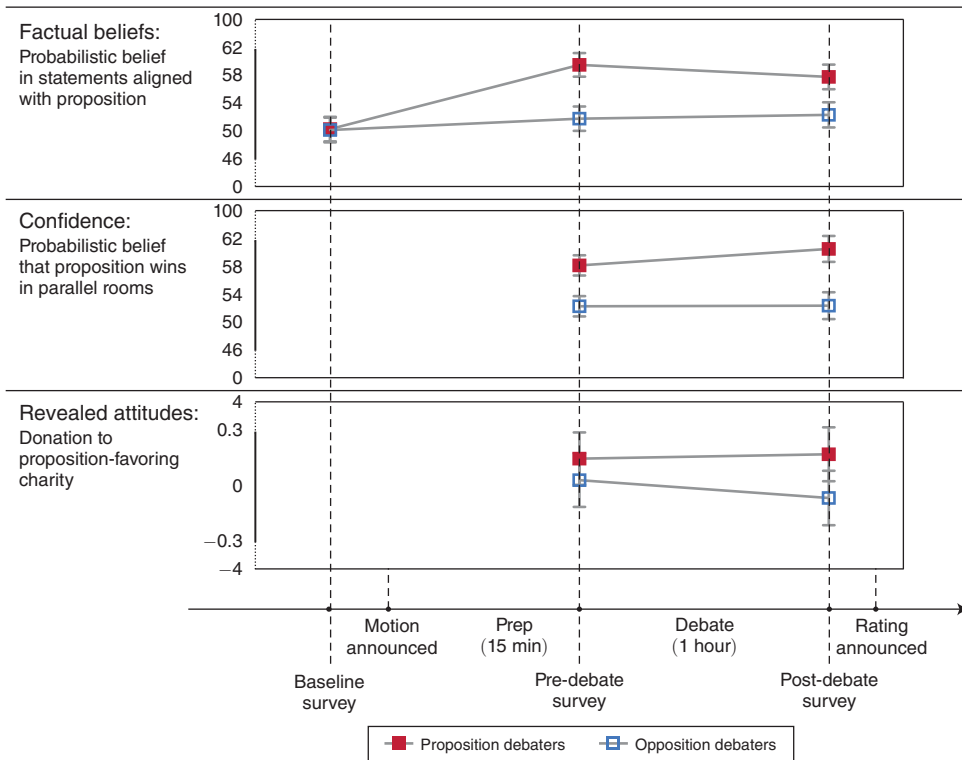


FIGURE 1. THE DYNAMICS OF FACTUAL BELIEFS, CONFIDENCE AND REVEALED ATTITUDES

*Notes:* The figure shows the dynamics of average alignment with the proposition for both sides of the debate, going from baseline, to pre-debate, to post-debate on our three main outcomes. For all three outcomes, higher values denote greater alignment with the proposition. The support of factual beliefs and confidence includes integers between 0 and 100, while revealed attitudes includes integers between  $-4$  and  $4$ . Each squared dot corresponds to the average of the alignment outcome at each point in time on each side of the debate and is placed between 95 percent confidence intervals. Dotted on the y-axis are segments of the support that are not plotted in the chart.

an equal extent.<sup>10</sup> Furthermore, we find that the correct answer is aligned with the proposition 47.3 percent of the time at pre-debate and 47.7 percent of the time at post-debate, so there is no strong imbalance. These findings suggest that debaters' beliefs simply happen to be more likely to favor the proposition.

### B. Pre-debate Self-Persuasion

We now turn to the statistical analysis of self-persuasion in the pre-debate elicitations. These effects reflect the cognitive processes taking place in the 15 minute

<sup>10</sup>In the adjudicator's survey that we conducted in the online tournaments, we ask adjudicators to predict debaters' beliefs and to state their own factual beliefs. Average factual beliefs of adjudicators are 54.6, which means they are slightly biased toward the proposition. These factual beliefs lie in the middle of proposition debaters' and opposition debaters' factual beliefs of 51.5 and 60.5 respectively.

preparation period after persuasion goals are assigned, but before the debate begins. We estimate self-persuasion effects in the following regression model:

$$(1) \quad y_{i,q} = \alpha + \beta \mathbf{1}(\text{proposition}_{i,q}) + \delta_q + d_i + \varepsilon_{i,q},$$

where  $y$  is the outcome variable of interest with value  $y_{i,q}$  for debater  $i$  answering question  $q$ ,  $\alpha$  is a constant,  $\mathbf{1}(\text{proposition}_{i,q})$  is an indicator variable for being assigned to the proposition,  $\delta_q$  is a question fixed effect,  $d_i$  is a debater random effect that is assumed to be orthogonal to the randomly assigned  $\text{proposition}_{i,q}$ , and the error term  $\varepsilon_{i,q}$  is clustered within each team of debaters.<sup>11</sup> For factual beliefs and confidence  $y$  is an integer between 0 and 100, while for revealed attitudes  $y$  is an integer between  $-4$  and  $4$ .

Panel A of Table 2 shows the regressions for the pre-debate treatment effect, confirming the visual evidence. In addition, we present separate results for the two offline tournaments (panel B, columns 1–3) and the online tournaments (panel B, columns 4–6). These data differ in the time of collection, as the offline tournaments preceded the COVID-19 pandemic and the online data were collected during it, and in some details of the debating and survey procedures. All effects are robust to the omission of question fixed effects (see online Appendix Table A.4).

In panel A, column 1 shows that proposition debaters are significantly more likely to believe that factual statements favoring the proposition are true and that statements favoring the opposition are false. Column 2 shows that proposition debaters are also significantly more confident that a majority of proposition teams will win the debates in the parallel rooms. The effects of factual beliefs and confidence are robust across tournaments, but are about 50 percent higher in the online than in the offline format, although the difference is not statistically significant.

For revealed attitudes, we do not see a statistically significant self-persuasion effect (panel A, column 3). In this case, there is a large difference in the online and offline format, with the effect being positive (and statistically significant) offline and negative and not statistically significant online. This appears to be partly driven by a survey design issue in the online format. Since the attitude elicitation was both the last and the most complicated survey item, time pressure and limited supervision may have reduced the attention of debaters to this question.<sup>12</sup> This interpretation is in line with the fact that the self-persuasion effect in attitudes is present offline, where supervision was stricter, and post-debate, where time pressure was lower.

*Effect Size.*—How should we think about the size of these self-persuasion effects? We provide two benchmarks to answer this question. First, we can compare the standardized effects to those of the rather sizable laboratory literature on this topic.

<sup>11</sup> We cluster the error term at the team level because this is the level at which all randomization takes place. However, the error term might be also correlated at the room level, especially post-debate. Our main results are robust when clustering at the room level (see online Appendix Table A.3).

<sup>12</sup> We see that the debaters online are much more likely to favor the neutral charity relative to the motion-related charity, where the latter happened to be aligned with the proposition slightly more often. In exploratory analyses we exclude 73 debaters who did not answer all surveys and hence might not have taken it seriously. In the equivalent of column 3, we find a substantially larger self-persuasion effect (0.138,  $p = 0.187$ ).

TABLE 2—PRE-DEBATE SELF-PERSUASION

	All tournaments					
	(1)	(2)	(3)			
<i>Panel A. Full sample</i>	Factual beliefs	Confidence	Revealed attitudes			
Proposition alignment in:						
Assigned to proposition	7.153 (1.058)	5.920 (0.974)	0.097 (0.097)			
Debaters	473	473	473			
Observations	2,217	2,213	2,212			
R <sup>2</sup>	0.216	0.110	0.194			
	Munich and Rotterdam (offline)			Amsterdam and LSE (online)		
<i>Panel B. Offline versus online</i>	(1)	(2)	(3)	(4)	(5)	(6)
Proposition alignment in:	Factual beliefs	Confidence	Revealed attitudes	Factual beliefs	Confidence	Revealed attitudes
Assigned to proposition	6.192 (1.802)	4.389 (1.492)	0.277 (0.140)	7.821 (1.286)	6.943 (1.282)	-0.020 (0.130)
Debaters	196	196	196	277	277	277
Observations	884	883	883	1,333	1,330	1,329
R <sup>2</sup>	0.140	0.034	0.136	0.268	0.157	0.223

*Notes:* Random effects linear regression model with standard errors (in parentheses) clustered at the team level. All specifications include question fixed effects. Each round, debaters are randomly assigned to argue either as proposition or opposition. The outcome is our measure of pre-debate alignment with the proposition in either factual beliefs, confidence, or revealed attitudes. For all three outcomes, higher values denote greater alignment with the proposition. The support of factual beliefs and confidence includes integers between 0 and 100, while revealed attitudes includes integers between -4 and 4. The number of observations is determined by valid responses from debaters over five (four in Rotterdam) rounds of debate.

Table 3 provides an overview of laboratory work on self-persuasion.<sup>13</sup> By comparing the standardized effect sizes in the final column, we see that our field estimates are smaller than those of most other studies, although they are larger than estimates in Soldà et al. (2019) and comparable to Schwardmann and van der Weele (2019). The relatively smaller effect size in our field setting mirrors the results of Della Vigna and Linos (2020), who show that the effects of “nudge unit” interventions in the field are about 25 percent of those obtained in comparable academic studies. They attribute this to publication bias in the academic literature and differences in the details of the intervention. In addition, we conjecture that the laboratory studies may have been designed and piloted to contain sufficient scope for self-persuasion. In Section IIIC, we show that there is a large heterogeneity in self-persuasion between motions, making topic selection consequential.

A second way to benchmark effect sizes is to compare them to the degree of political polarization in the outcome variables. Political ideology typically coincides with persuasion goals in the field, but can be separated here due to the orthogonal

<sup>13</sup>We focus on self-persuasion in situations where subjects are (i) incentivized to persuade or negotiate with others, and (ii) face incentivized belief measurements, as Bullock et al. (2015) shows that accuracy incentives in surveys reduce polarization effects. While we are not aware of other studies that fit these criteria, we do not guarantee this is an exhaustive list. We also include Festinger and Carlsmith (1959) as a seminal reference point, although we cannot compute standardized effect sizes as no details of the sample distributions are reported, an issue that also plagues other early studies in psychology, like O’Neill and Levings (1979).

TABLE 3—LITERATURE REVIEW EFFECT SIZE

Paper	Context	Persuasion objective	Treatment	Control	Outcome	Sample	TE	TE/ $\sigma_y$
1. Festinger and Carlsmith (1959)	Boring effort task	Convince others that the task is enjoyable	Incentive 20 USD	Incentive 0 USD	Self-reported interest in the task (Likert scale -5 to 5); unincentivized	Laboratory subjects; observations = 40	0.4	NA
			Incentive 1 USD	Incentive 0 USD	Self-reported interest in the task (Likert scale -5 to 5); unincentivized	Laboratory subjects; observations = 40	1.8	NA
2. Thompson and Loewenstein (1992)	Fictitious wage bargaining	Negotiate favorable settlement	Submit wage offer as union	Submit wage offer as manager	Fair wage (in USD); unincentivized	Laboratory subjects observations = 40	0.15	0.780
3. Loewenstein et al. (1993)	Fictitious trial	Negotiate favorable settlement	Argue prosecutor side	Argue defendant side	Fair settlement (in USD); unincentivized	Laboratory subjects; observations = 160	17,710	1.086
					Judge prediction (in USD); unincentivized	Laboratory subjects; observations = 160	14,527	0.834
4. Babcock, Issacharoff, and Camerer (1995)	Fictitious trial	Negotiate favorable settlement	Don't know persuasion objective before reading materials	Know persuasion objective before reading material	Within pair difference in fair settlement (in USD); incentivized	Laboratory subjects; observations = 94	26,031	1.087
					Within pair difference in prediction of judge (in USD); incentivized	Laboratory subjects; observations = 94	25,491	0.932
5. Chen and Gesche (2017)	Financial advice game	Make advisee buy asset A over alternatives	Commission to recommend asset A	No commission to recommend asset A	Own choice of whether to buy asset A (binary); incentivized	Laboratory subjects; observations = 99	0.173	1.150
6. Gneezy et al. (2020)	Financial advice game	Make advisee buy asset A over alternative	Review asset before learning commission for asset A	Review asset after learning commission for asset A	Belief that advisee prefers asset A to B (binary); unincentivized	Amazon MTurk workers; observations = 900	0.338	1.519
7. Schwardmann and van der Weele (2019)	Verbal persuasion task	Persuade others verbally of high test performance	Know about persuasion task	Does not know about persuasion task	Belief about own IQ test score (probability); incentivized	Laboratory subjects; observations = 688	0.060	0.309
8. Soldà et al. (2019)	Written persuasion task	Persuade others in writing of high test performance	Already completed persuasion task	No awareness of persuasion task	Belief number of correct answers (0-31); incentivized	Amazon MTurk workers; observations = 600	0.650	0.110
9. This paper	High profile debating competition	Win debate	Argue for motion	Argue against motion	Belief that facts in favor of the motion are true (probability); incentivized	Expert debaters; observations = 2,217	0.078	0.264
					Confidence that teams arguing in favor of the motion win debates (probability); incentivized	Expert debaters; observations = 2,213	0.058	0.227
					Donations towards charitable organizations supporting causes in favor of the motion (rank 0-9); incentivized	Expert debaters; observations = 2,212	0.116	0.048

*Notes:* This table presents treatment effects (TE) and standardized treatment effects (TE/ $\sigma_y$ ) from related experimental paradigms. The only study that we know has undergone exact replication is Babcock, Issacharoff, and Camerer (1995). The average standardized effect in the replication of Hippel and Hoepfner (2019) is about one-half the original effect (in the original study, effect sizes on outcomes 1 and 2 are 1.087 and 0.932, respectively; in replication effect sizes on outcomes 1 and 2 are 0.646 and 0.390, respectively), but still sizable. The unweighted average of the standardized effect size for studies 2-8 is 0.867. The average effect size across the three main outcomes of the present study is 0.180, which is 20.8 percent of the unweighted average of the rest of the literature.



assignment of persuasion goals. To investigate the degree of political polarization, we construct a dummy variable that takes a value of one if debater's political leanings are aligned with the political leanings of a motion's proposition. We elicited debaters' political leanings on a left-right scale from zero (very left) to ten (very right). We classify a debater as left leaning if their reply falls into the range of zero to four, and as right leaning otherwise. According to this classification 26.4 percent of debaters are right leaning.

We classify a motion as left leaning if and only if, at baseline, left-leaning debaters are more likely to believe in the factual statements that support the proposition. Basing our classification on the revealed viewpoints of our debaters at the start of the tournament has the advantage of incorporating the political perceptions of the people whose partisan attitudes we are investigating. However, all our results are robust to an alternative political classification of motions, based on the ratings of an independent sample of 23 debaters (see online Appendix Table A.5 for a detailed overview).

We say that a debater is politically aligned with the proposition if both debater and proposition are left leaning or right leaning. To illustrate, a right-leaning debater is clearly aligned with the proposition "This House would suspend trade union powers and significantly relax labor protection laws in times of economic crisis." We can now estimate political polarization with regressions that are analogous to those for self-persuasion in equation (1); we just replace the indicator for *arguing for* the proposition side of the motion with an indicator for being *politically aligned with* the proposition side.

The results of this exercise are presented in the first three columns of Table 4. Comparing the effect sizes of this exercise with the self-persuasion effects in Table 2, we find that for factual beliefs, political polarization is about two-thirds of the self-persuasion effect. Turning to revealed attitudes, where the effect of self-persuasion is small, we find that the political effect is almost ten times as large as the self-persuasion effect across all tournaments and slightly larger than the self-persuasion in revealed attitudes we see in the offline tournaments. Finally, we do not see partisan polarization on confidence, where the point estimate is negative and statistically insignificant.

Thus, in two out of three outcome variables, we observe a larger effect of self-persuasion than of political polarization, indicating that the self-persuasion effect is a quantitatively important driver of polarization in this setting. Note, however, that political polarization in our setting may be less pronounced than in the general population, as ideological heterogeneity is relatively small, and not all motions evoke clear ideological difference between left and right.

### *C. Post-Debate Self-Persuasion and Convergence*

We now turn to the post-debate survey. As we discussed in the introduction, there are two ex-ante plausible hypotheses about the dynamics of self-persuasion over the course of the debate. On the one hand, the pooling of arguments from both sides should lead impassionate reasoners to reach similar conclusions and reduce polarization. On the other hand, the very act of debating may reinforce the effect of persuasion goals and increase polarization at the end of the debate.

TABLE 4—POLITICAL POLARIZATION

Time of elicitation:	Pre-debate			Post-debate		
	Factual beliefs	Confidence	Revealed attitudes	Factual beliefs	Confidence	Revealed attitudes
Proposition alignment in:	(1)	(2)	(3)	(4)	(5)	(6)
Politically aligned with proposition	4.606 (1.435)	-1.367 (1.043)	0.379 (0.115)	1.392 (1.616)	-0.700 (1.173)	0.081 (0.121)
Debaters	463	463	463	462	271	462
Observations	2,178	2,174	2,173	2,141	1,277	2,139
$R^2$	0.201	0.087	0.207	0.236	0.121	0.226

*Notes:* Random effects linear regression model with standard errors (in parentheses) clustered at the team level. All specifications include question fixed effects. Each round, debaters are randomly assigned to argue either as proposition or opposition. The outcome in columns 1–3 is our measure of pre-debate alignment with the proposition in either factual beliefs, confidence, or revealed attitudes. For all three outcomes, higher values denote greater alignment with the proposition. We call debaters right leaning if they report political views on the zero-to-ten political scale above four. For each round, we regress baseline factual belief alignment on being a right leaning debater and categorize the motion of that round to be right leaning if the regression coefficient is positive (in online Appendix Table A.5 we conduct a different categorization based on a small follow-up survey and find the results to be qualitatively robust). “Politically aligned with proposition” equals one if both the motion and the debater are left/right leaning, and zero otherwise. The outcome in columns 4–6 is the post-debate alignment analog. The number of observations is determined by valid responses from debaters over five (four in Rotterdam) rounds of debate.

The visual evidence in Figure 1 suggests that the answer lies in the middle, with little overall evidence of either convergence or divergence. The statistical results are presented in Table 5. Columns 1–3 show the treatment effects of our three outcome variables post-debate for all tournaments combined, analogous to the first three columns of Table 2. We now see a sizable and statistically significant self-persuasion effect for all three variables. This includes the coefficient for revealed attitudes, which was not statistically significant pre-debate.

To estimate the size of convergence or divergence, the regression models reported in columns 4–6 include a dummy for being in the proposition, a dummy for the post-debate survey, and the interaction between the two. The coefficients for the latter term give the size of the difference-in-differences between the pre- and post-debate treatment effects. We do not find evidence of consistent or statistically significant convergence or divergence across the three outcome variables.<sup>14</sup>

*Information Transmission during the Debates.*—Given the lack of movement of our main variables, one may wonder if the debates resulted in any information transmission at all. To investigate this question, we use some additional measurements. First, we asked each debater for the number of separate arguments for either side of the motion that they could think of.<sup>15</sup> We find that the total number of arguments cited by debaters increases by 20 percent over the course of the debate, from an average of 6.4 at pre-debate to an average of 7.7 post-debate, a difference that is statistically significant ( $t$ -test,  $p < 0.001$ ). These numbers are not due to debaters’

<sup>14</sup>Our study has 79 percent power to detect post-debate convergence of the size of 50 percent of the pre-debate gap in our preregistered primary outcome, i.e., factual beliefs.

<sup>15</sup>At offline tournaments, we only measure this pre-debate. At online tournaments, we measure this both pre- and post-debate. Here, we only report the numbers for the online tournament, so we can focus on changes.

TABLE 5—POST-DEBATE SELF-PERSUASION AND CONVERGENCE

Proposition alignment in:	Post-debate			Difference-in-differences		
	Factual beliefs (1)	Confidence (2)	Revealed attitudes (3)	Factual beliefs (4)	Confidence (5)	Revealed attitudes (6)
Assigned to proposition	5.055 (1.264)	7.940 (1.295)	0.200 (0.095)	7.139 (1.173)	6.922 (1.261)	0.077 (0.103)
Post-debate				0.455 (1.418)	-2.181 (0.792)	-0.093 (0.093)
Assigned to proposition × post-debate				-2.245 (1.991)	1.098 (1.203)	0.132 (0.132)
Debaters	470	274	470	473	277	473
Observations	2,171	1,286	2,169	4,388	2,616	4,381
$R^2$	0.236	0.159	0.224	0.098	0.151	0.152

*Notes:* Random effects linear regression model with standard errors (in parentheses) clustered at the team level. All specifications include question fixed effects. Each round, debaters are randomly assigned to argue either as proposition or opposition. The outcome of columns 1–3 is our measure of post-debate alignment with the proposition in either factual beliefs, confidence, or revealed attitudes. The outcome of columns 4–6 is our measure of alignment with the proposition in either factual beliefs, confidence, or revealed attitudes—either at pre-debate or post-debate. For all outcomes, higher values denote greater alignment with the proposition. The number of observations is determined by valid responses from debaters over five (four in Rotterdam) rounds of debate. Post-debate confidence was collected only at online tournaments.

exclusively generating more arguments for their side: the number of arguments for the *opposite* side increases over the course of the debate from 2.8 to 3.5, or 25 percent (*t*-test,  $p < 0.001$ ). This shows that debaters learned new arguments during the debate. In Section IVA, we investigate the role of the number of arguments in the development of self-persuasion in more detail. One might be concerned that post-debate self-persuasion stems from a desire to not “admit defeat” in front of the experimenter. Inconsistent with that interpretation, we observe that the reported share of arguments in favor of their own position shifts from 57.0 pre-debate to a more balanced 54.5 post-debate (*t*-test,  $p < 0.001$ ).

Another way to investigate the impact of the debates is to look at their effect on political polarization. To this end, we compare the degree of political polarization pre-debate and post-debate. As can be seen in columns 4 and 6 of Table 4, post-debate factual beliefs and revealed attitudes reflect less political polarization than their pre-debate counterparts. Thus, in contrast to the polarization induced by self-persuasion, debates did have a clear mitigating effect on political polarization. Note that in contrast to the randomized persuasion goals, the dimension of political partisanship was not reinforced during the debate, where many subjects argued against their own political leanings. This may explain why political polarization, but not self-persuasion, declines during the debate, and suggests that persistence of disagreement requires the reinforcement of persuasion goals during the debate.

*Persistence of Post-debate Effects.*—How long does the effect of self-persuasion persist? Our ability to answer this question is limited by the two days’ length of our tournaments and our inability to contact debaters afterwards. Nevertheless, we addressed this point in the final survey of the online tournaments, where we asked debaters again about their factual beliefs related to all five motions in the

qualifying rounds. The factual questions were the exact same that we asked them at baseline, allowing us to see if factual beliefs shifted between the beginning of the tournament and the end of the qualifying rounds. This is a strong test of persistence, since a concern to be consistent may reduce the difference between the two elicitation.

In online appendix Table A.6, we investigate the self-persuasion effect on day two—for factual beliefs—with the same regression model as our main analysis. We find that this effect remains sizable at 80 percent the post-debate effect size and statistically significant ( $p = 0.003$ ). We show that this effect is not driven by the fifth round of debate, which took place on day two of each competition right before the final survey. Thus, we conclude that the self-persuasion effect persists until the next day, despite the intervening engagement in at least one unrelated debate.

### III. Heterogeneity

In this section we look at heterogeneity of the pre-debate self-persuasion effect across debater experience, incentives for accuracy, and the topics of motions.

#### A. *Experience and Past Success*

Is experience or past success associated with less self-persuasion? Experience may allow people to learn and reduce behavioral biases, as has been documented in the case of the endowment effect (List 2003). High-profile debate tournaments are uniquely suited to study this question: While all participants have some degree of experience, there is still substantial heterogeneity in the number of years debaters have been debating as well as in their past successes, measured by how many times they previously made it out of the preliminary rounds into to the semifinals at big tournaments.

In Table 6, we present regressions for factual beliefs and confidence, the two outcome variables where we find significant pre-debate self-persuasion. In columns 1 and 2, we interact the treatment (being assigned to the proposition) with an indicator for having more than the median years of debating experience. In columns 3 and 4, we interact the treatment with an indicator for more than median number of semi-final attainments. These two binary indicators capture related aspects of experience (correlation  $\rho = 0.486$ ), and the interaction terms with the treatment show how experience correlates with self-persuasion.

In column 1, the experienced group shows about one-half the treatment effect on the factual beliefs measure, while for confidence in column 2, we do not see an attenuating effect from experience. We obtain comparable results in column 3 and 4: high achievers have about one-half the self-persuasion effect on factual beliefs, but there is not much difference in self-persuasion on confidence and the sign of the effect is reversed. Note that the self-persuasion effect on confidence for experienced debaters cannot be justified by their superior performance, as debaters were predicting the outcome of other, simultaneous debates, not their own. Thus, debaters of all levels of experience are subject to self-persuasion, although experienced and successful debaters show a smaller effect on the factual beliefs measure.

TABLE 6—HETEROGENEITY AND STAKES

Proposition alignment in:	Experience and achievements				Stakes of elicitation	
	Factual beliefs (1)	Confidence (2)	Factual beliefs (3)	Confidence (4)	Factual beliefs (5)	Confidence (6)
Assigned to proposition	9.968 (1.484)	5.791 (1.258)	9.612 (1.520)	5.369 (1.241)	7.538 (2.106)	9.044 (1.729)
Experienced	3.569 (1.698)	0.242 (1.793)				
Assigned to proposition $\times$ experienced	-5.912 (2.109)	0.445 (1.994)				
High achiever			2.587 (1.617)	0.974 (1.790)		
Assigned to proposition $\times$ high achiever			-5.368 (2.076)	1.406 (1.842)		
High incentive					1.196 (2.112)	2.219 (1.907)
Assigned to proposition $\times$ high incentive					0.598 (3.041)	-4.085 (2.589)
Debaters	465	465	463	463	277	277
Observations	2,187	2,183	2,177	2,173	1,333	1,330
$R^2$	0.218	0.110	0.227	0.110	0.268	0.160

*Notes:* Random effects regression model with standard errors (in parentheses) clustered at the team level. Each round, debaters are randomly assigned to argue either as proposition or opposition. The outcome is our measure of pre-debate alignment with the proposition in either factual beliefs or confidence. For both outcomes, higher values denote greater alignment with the proposition. In columns 1 and 2 we interact the treatment with experienced, a binary indicator for above median years of experience in debating. In columns 3 and 4 we interact the treatment with high achiever, a binary indicator for above median number of international tournaments in which the debater reached the kick-out phase. In columns 5 and 6 we interact the treatment with high incentive, a binary indicator for randomly assigned incentive for the question. The number of observations is determined by valid responses from debaters over five (four in Rotterdam) rounds of debate. In particular, remember that experimental variation in the stakes of elicitation was introduced only at online tournaments.

### B. Incentives

Economic theories of motivated cognition predict that beliefs are sensitive to the costs of misperceptions (Bénabou and Tirole 2016). The previous literature has yielded some evidence for this prediction (Zimmermann 2020), but there are also several null results (Mayraz 2011; Coutts 2019). In our setting, the cost of misperceptions is given by the monetary incentives for accuracy: the higher the deviation from the true answer, the lower the chance of winning the prize.

To test whether these incentives influence self-persuasion, we implemented exogenous variation in the prize that could be won. In the low incentive condition, debaters could win 5 euros with a correct answer, while in the high incentive condition, we increased the incentive tenfold to 50 euros. These conditions were implemented within-subject and in the online tournaments only, with factual belief and confidence questions randomly assigned to either condition. We informed debaters by displaying a “5 euro” or “50 euro” signal in front of the relevant questions.

Columns 5 and 6 of Table 6 show the regressions of pre-debate factual beliefs and confidence on a dummy for being in the proposition, a dummy for being in the

high incentive condition, and the interaction between the two. The coefficient for the interaction term shows that self-persuasion on confidence is mitigated somewhat by high incentives, although the effect is not precisely estimated and not statistically significant. Similarly inconclusive results obtain from analogous regressions for the post-debate outcomes. Thus, we find no clear evidence for an effect of incentives.

A related question is how costly self-persuasion is to the debaters, in terms of foregone earnings from accurate answers. To this end, we compare the performance of the online debaters at pre-debate with a control group that did not have a persuasion goal. Like above, we use the adjudicators, who answered the same questions as the debaters in the online pre-debate survey. We find that adjudicators have a probability of winning the prize of 0.693, whereas debaters have a probability of winning the prize of 0.656 in the low stakes condition and 0.653 in the high stakes condition. This comparison implies that the cost of self-deception for debaters is 0.19 euros in the low stakes condition (with a 5 euro monetary prize) and 2 euros in the high stakes condition (with a 50 euro monetary prize).

### C. *Heterogeneity across Topics and the Predictability of Self-Persuasion*

The debates generate self-persuasion across multiple policy motions, which allows us to investigate the consistency of the effect. The three panels of Figure 2 show standardized self-persuasion effects for factual beliefs, confidence and revealed attitudes, aggregated at the motion level.<sup>16</sup> There are two main takeaways from this figure. First, there is a lot of heterogeneity in effect sizes across motions for a given outcome variable.<sup>17</sup> For example, debaters arguing for and against engaging private military companies to combat terrorism exhibited self-persuasion on factual beliefs that was more than three times larger than the average effect. Second, self-persuasion on one outcome does generally not predict self-persuasion on other outcomes.<sup>18</sup> In fact, it is hard to find a single motion for which we see (sizable) self-persuasion on all outcomes. Together these results suggest that studies should be careful in generalizing findings of polarization from any single topic or outcome.

The variability in the treatment effect across outcome variables may be explained by the variation in questions and charities selected by the researchers. However, we also see substantial heterogeneity in the treatment effect on confidence, where the elicitation question is always the same. This suggests the presence of a second source of heterogeneity that is related to the topic of the motions. This finding is in line with Tappin (2020), who identifies large variation in partisan political polarization across issues in US politics.

<sup>16</sup>Figure 2 shows only 15 out of 19 motions due to some of the motions' sensitive nature. Our analysis is based on all 19 motions.

<sup>17</sup>A test for heterogeneous treatment effects (Cochran, 1954) rejects the hypothesis that the treatment effect is homogeneous across rounds for factual beliefs ( $Q$ -test,  $p < 0.001$ ), confidence ( $Q$ -test,  $p < 0.043$ ) and revealed attitudes ( $Q$ -test,  $p < 0.098$ ). To interpret the heterogeneity in treatment effects across rounds we estimate the  $I^2$  (Higgins and Thompson 2002), which measures the proportion of total variation across rounds that is due to heterogeneity rather than within-round sampling error. This measure is 62.2 percent for factual beliefs, 38.8 percent for confidence and 31.0 percent for revealed attitudes.

<sup>18</sup>We cannot reject the null hypothesis that the correlation of standardized effects by motion across outcomes is zero. It is  $\rho = 0.154$  ( $p = 0.293$ ) between factual beliefs and revealed attitudes,  $\rho = 0.036$  ( $p = 0.882$ ) between factual beliefs and confidence, and  $\rho = 0.048$  ( $p = 0.844$ ) between revealed attitudes and confidence.



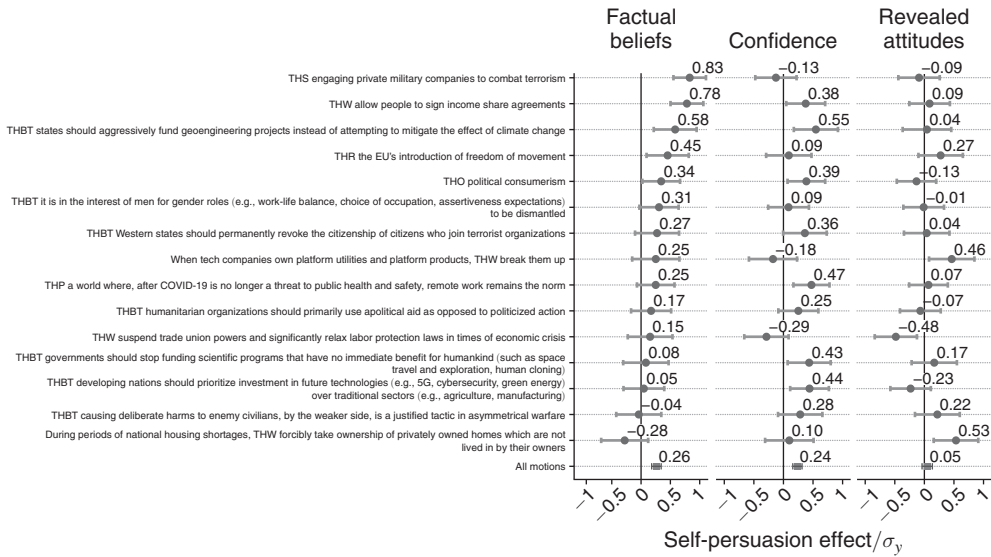


FIGURE 2. SELF-PERSUASION BY MOTION, ORDERED BY EFFECT SIZE IN FACTUAL BELIEFS

Notes: The figure shows self-persuasion effects  $\beta$  from regression model equation (1), estimated separately by motion. The estimated effect is divided by the standard deviation in the outcome variable  $\sigma_y$ . Capped ranges are 95 percent confidence intervals. Acronyms: THS = This House supports, THW = This House would, THR = This House regrets, THBT = This House believes that, THO = This House opposes, THP = This House proposes.

*Predicting Motion Effects.*—Is polarization predictable across motions and questions? The answer to this question will help us understand where and when disagreement arises and can potentially be avoided. As a first step toward answering this question, we investigate whether self-persuasion can be predicted by a group of experts. Such predictions also provide an additional benchmark for our effect sizes and help understand the awareness of the effect in the debating community (DellaVigna and Pope 2018; DellaVigna, Otis, and Vivaldi 2020).

We asked the adjudicators in the online tournament to predict the treatment effect. Adjudicators are arguably the best placed group to predict the effect of self-persuasion and of motion variation. They are experts in this particular context, as they have intimate knowledge of the debating environment and are experienced at debating as well as evaluating other debaters. In a pre-debate survey for each motion, we provided adjudicators with the factual questions and attitude elicitation related to the motion, and asked them to estimate the average responses for both proposition and opposition debaters. We incentivized their answers with the same scoring rule we used for the debaters, with a potential prize of 15 euros.

Table 7 compares the predictions with the actual (pre-debate) effect sizes, and shows that adjudicators do reasonably well in predicting pre-debate self-persuasion in the online tournaments. They overestimate the effect sizes for factual beliefs by about 34 percent and are strikingly accurate for confidence. For revealed attitudes the adjudicators overestimate the mark substantially, as there was no effect in the online competitions, although their estimates are close to the actual effect in the offline competitions. Table 7 also shows the correlations of the predictions and

TABLE 7—ADJUDICATOR PREDICTIONS VERSUS ACTUAL EFFECT SIZES (ONLINE ONLY)

	Factual beliefs	Confidence	Revealed attitudes
Actual effect size	7.81	7.22	-0.02
Predicted effect size	10.48	7.74	0.35
Correlation actual-predicted (motion level)	0.36	0.43	-0.07
Correlation actual-predicted (question level)	0.34	0.43	-0.04
Motions	10	10	10
Questions	30	10	20

*Notes:* Adjudicators' predictions are only available for online tournaments. Actual effect size at the motion level is calculated as the average proposition alignment among proposition debaters minus average proposition alignment among opposition debaters at pre-debate. Predicted effect size at the motion level is calculated as the average proposition alignment predicted by adjudicators minus average proposition alignment predicted by adjudicators among opposition debaters.

actual effects on the motion level. An overview of the predictions, organized by motion, is given in online Appendix Figure A.2.

The individual questions are another source of heterogeneity. In particular, for each motion, we have three different questions to elicit factual beliefs, and two different charities to elicit revealed attitudes. Some of these questions or charities may be more salient or have a stronger connection to the core arguments in the debate, and hence generate more self-persuasion. In the bottom row of Table 7, we show the correlations of the adjudicators' predictions and actual self-persuasion on the question level. Adjudicators have similar performance as for motions, showing that at least for factual beliefs, they are able to predict self-persuasion to some degree. Overall, adjudicators do a reasonable job at not just predicting the existence of an overall effect, but at predicting its heterogeneity over motions and questions.

These results are of interest for several reasons. First, it is striking that the debaters in our prestigious tournament succumb to the self-persuasion effect, while their (experienced) peers predict it. This suggests that self-persuasion works despite an awareness of its existence, as suggested in Saccardo and Serra-Garcia (2020). It also shows that people anticipate the biases of others, complementing results for the case of present-bias (Fedyk 2018), and more conflicting evidence on overconfidence (Ludwig and Nafziger 2011). Note that we asked adjudicators explicitly about their beliefs for both the opposition and opposition debaters, which may have raised the salience of this split. While it is hard to avoid such measurement effects, the anticipation of nonsalient biases remains an open question. Second, the adjudicators seem able to use some of the content of the motions and questions in their predictions of the self-persuasion effect. This raises questions about the exact contextual features that generate self-persuasion. Given the limited number of motions/questions in our sample and the large number of potentially relevant dimensions, we leave this as a challenge for future research.

#### IV. The Why and How of Self-Persuasion

We now turn to the psychological mechanisms underlying self-persuasion and the potential benefits of self-persuasion. We first investigate a plausible mediator of the self-persuasion effect, namely the biased generation of arguments. We then discuss

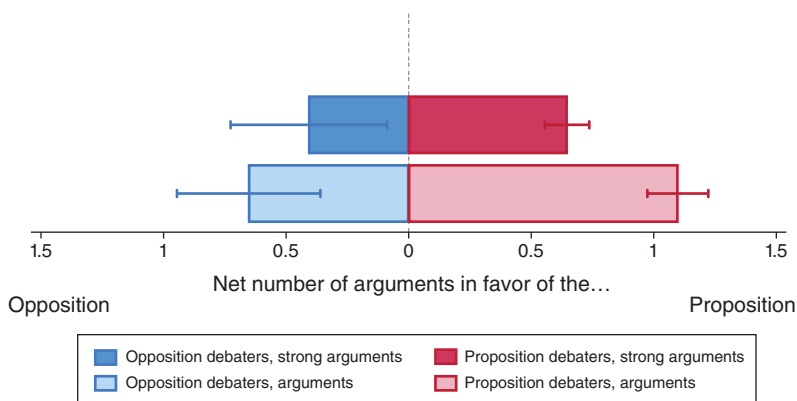


FIGURE 3. DIFFERENCES IN THE NUMBER OF ARGUMENTS

*Notes:* The figure shows how debaters on each side disproportionately enumerate more arguments for their side at pre-debate. Bars with low intensity fill are based on all arguments debaters enumerate, and full bars are based on the number of arguments for each side that debaters deem as very strong. Capped ranges are 95 percent confidence intervals.

whether self-persuasion has instrumental value in helping debaters win the debate. We conclude this section by ruling out experimenter demand effects as a potential confound.

### A. *The Biased Generation of Arguments*

A number of theories point to the biased generation of the number of arguments as a mediator of self-persuasion. For instance, according to “persuasive argument theory” (Vinokur and Burstein 1974), the number of new arguments that a side brings to the table is a key driver of persuasion. Mercier and Sperber (2011) theorize that our reasoning abilities have developed in order to persuade others through the biased generation of arguments, which produce self-persuasion as a by-product. Bénabou, Falk, and Tirole (2019) argue that persuasion and justification in moral dilemmas occur as the result of a selective search for “narratives.” On the empirical side, Thompson and Loewenstein (1992) show that people asymmetrically recall facts in a bargaining situation, while other papers find people engage in biased search of evidence to align with their persuasion goals (Smith, Trivers, and von Hippel 2017; Soldà et al. 2019).

To understand the role of biased argument generation in our debate setting, we asked debaters in the pre-debate 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 3 shows the average net number of arguments debaters came up with on both sides, split by treatment. As is clear from the graph, debaters engage in asymmetric selection of arguments. On average, they come up with almost one additional argument and one-half of a “strong” argument in favor of their own side.

To quantify the impact of this asymmetry for self-persuasion, we conduct a parametric causal mediation analysis (Imai, Keele, and Yamamoto 2010). We define  $s_i$ , the number of arguments aligned with the persuasion goal as a fraction of total arguments considered by debater  $i$  during preparation time, and investigate how this mediates self-persuasion on our three main outcome variable. The analysis decomposes the average treatment effect into the average direct effect and the average causal effect mediated by  $s_i$ . Overall,  $s_i$  drives 14 percent of the self-persuasion effect in factual beliefs, 44 percent for confidence and 58 percent for revealed attitudes. In online Appendix C we provide further details of the mediation effects both offline and online.

These results indicate a substantial, although heterogeneous and incomplete role for biased argument generation. To explain the effect of biased argument generation on beliefs, it must be the case that debaters fail to correct for this bias when they assess their position. This failure may either result from a lack of sophistication, in line with a literature on selection neglect (Juslin, Winman, and Hansson 2007; Barron, Huck, and Jehiel 2019), or may itself be motivated by the wish to align beliefs and attitudes with the persuasion goal.

### B. Alignment and Persuasiveness

Is self-persuasion “useful” to win a debate? The answer to this question speaks to theories about the social origins of self-persuasion. For instance, Von Hippel and Trivers (2011) theorize that self-persuasion is a strategic action aimed at increasing persuasiveness through the reduction of nervous tics, giveaway tells or other manifestations of cognitive dissonance that arise from a gap between beliefs and persuasion goals. This theory has received support in recent laboratory studies (Smith, Trivers, and von Hippel 2017; Schwardmann and van der Weele 2019; Soldà et al. 2019).

While our experiment cannot distinguish this theory from a self-persuasion-as-byproduct account discussed above, we can test the prediction that self-persuasion has benefits for persuasion. To this end, we investigate the effect of having aligned beliefs on the “ballot score,” a rating between 60 and 100 given by adjudicators to each individual debater at the end of the debate. The team with the highest ratings is declared the winner of the debate. To understand the usefulness of self-persuasion, we regress these scores on the alignment of factual beliefs and confidence with the own persuasion goal, which are the variables that show substantial self-persuasion. Note that factual beliefs and confidence at pre-debate are potentially endogenous as they depend partially on the individual degree of self-persuasion. Therefore, we also look at the alignment of factual beliefs in the baseline survey, which is exogenous due to the random assignment of persuasion goals.

Table 8 shows the results of these exercises. We find a positive correlation between alignment of beliefs and the ballot score. For factual beliefs, the correlation is similar for both baseline beliefs (column 1) and pre-debate beliefs (column 2). For confidence, where we don’t have a baseline elicitation, we find a positive correlation with pre-debate beliefs (column 3). In all cases, the coefficients are marginally significant with  $p < 0.100$ , so this evidence is only indicative, and in need

TABLE 8—DOES ALIGNMENT HELP TO WIN THE DEBATE?

	Individual speaker score		
	(1)	(2)	(3)
Belief alignment with own side at baseline (standardized)	0.028 (0.017)	0.028 (0.017)	0.026 (0.017)
Belief alignment with own side pre-debate (standardized)		0.032 (0.019)	
Confidence in own side pre-debate (standardized)			0.028 (0.016)
Debaters	459	459	459
Observations	2,179	2,151	2,148
$R^2$	0.029	0.033	0.031

*Notes:* Fixed effects linear regression model with standard errors (in parentheses) clustered at the team level. All specifications include motion and debater fixed effects. The number of observations is determined by valid responses from debaters over five (four in Rotterdam) rounds of debate. The outcome is a metric of individual performance adjudicated in the ballot. Speaker scores for a handful of debaters could not be matched to our dataset, as they did not agree to this information becoming publicly available.

of confirmation by future research. In summary, we find some evidence in line with the idea that self-persuasion has a beneficial effect on the adjudicators' evaluations, which may explain its persistence in our sample of experienced debaters.

### C. Ruling Out Experimenter Demand

Tappin, Pennycock, and Rand (2020) point out a common flaw in experiments that randomly assign persuasion goals to study politically motivated reasoning. Subjects may believe that the experimenter asked them to argue a particular viewpoint because of its empirical or logical validity, which aligns their beliefs even without any self-persuasion. The experimenter can avoid this by explicitly announcing the random nature of the assignment. However, this may lead the subject to second-guess the goal of the study, possibly introducing an experimenter demand effect.

Debating tournaments avoid these pitfalls, due to the nature of the randomization. Because it is public and explicit, debaters know not to infer anything from the assignment about the merits of their case. At the same time, it is a familiar and inconspicuous part of the competition and is therefore unlikely to direct participants' attention to our research question. To confirm this last claim, we asked subjects in the last survey to guess the aim of our research. We find that 19 percent of subjects made a guess that resembled our main hypotheses. If these subjects are driving our results, the effect should get smaller when we exclude them from the analysis. Online Appendix Table B.2 shows that this is not the case, indicating that experimenter demand is not a main factor in this setting.

## V. Conclusions

Our results show that the self-persuasion effects previously found in the laboratory are relevant in the field. We find that debaters distort factual beliefs and confidence in the direction of a position they are randomly assigned to argue. Self-persuasion occurs despite incentives for accuracy and persists after an intense

exposure to opposing views. These results obtain in prestigious tournaments, in a sample of experienced debaters that regularly supplies future elites and politicians.

Our result may contain insights for other applications, that we enumerate here. This extrapolation involves a degree of speculation, as there are alternative explanations that could be disentangled by future research. First, self-persuasion is likely to drive belief formation in political contexts, where convincing others is of central importance. This may explain why greater engagement with the political process causes greater and persistent polarization (Mullainathan and Washington 2009) and why polarization is more severe in the US congress than it is in the American public (Fiorina and Abrams 2008). It also suggests additional motives for political behavior such as canvassing and proselytizing, which may be important not just to convert others, but also for deepening the convictions of those doing the canvassing (Gal and Rucker 2010).

Relatedly, self-persuasion may be at work in markets with asymmetric information. It predicts that sellers in economic transactions “drink the Kool-Aid” and become overly optimistic about their product. This may explain why financial advisors privately invest in the underperforming funds for which they receive sales commissions (Linnainmaa, Melzer, and Previtro 2018). It may also be a driving force behind the development of asset market bubbles, for instance during the financial crisis of 2007–2008, where private real-estate portfolios of agents working in sales departments of mortgage providers underperformed those of other agents as well as nonspecialists (Cheng et al. 2014). Self-persuasion may also be involved in the spectacular rise and fall of start-up companies like Theranos, as entrepreneurs trying to lure investors become overconfident and miscalibrated.

Finally, one may wonder if there is a connection between our findings and polarization among “regular” people. Unlike politicians, lawyers and entrepreneurs, most people do not earn money for persuading others. Yet, as evidenced by heated discussions on social media, at dinner tables, and at football games, many people are intrinsically motivated to convince others of what they believe to be true or what aligns with their identity. In our setting, we cannot disentangle the relative importance of intrinsic and extrinsic motives to be persuasive, as tournament debaters are likely driven by both. On the one hand, they are engaged in a quest for status and visibility. On the other hand, they are unpaid enthusiasts who enjoy the act of persuasion. We conjecture that both types of motivation can induce self-persuasion including among nonprofessionals, but testing this conjecture remains a task for future research.

Our results leave open some other interesting questions. For instance, measuring the impact of persistent or long-run persuasion goals, like those arising from group membership or party affiliation, may help understand the formation of personal identity. Another question concerns the design of institutions that revolve around debating. We show that debates do not necessarily resolve conflicts of opinion and can actually make them worse. At the same time, debating tournaments are an extremely competitive context, and our results may not extend to settings where parties aim to reach consensus (Felton et al. 2015). Thus, an important question is how to design debating contexts to promote a shared understanding of facts and mitigate disagreement.



## REFERENCES

- Andrews, Isaiah, and Maximilian Kasy.** 2019. "Identification of and Correction for Publication Bias." *American Economic Review* 109 (8): 2766–94.
- Babcock, Linda, Samuel Issacharoff, and Colin Camerer.** 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review* 85 (5): 1337–43.
- Babcock, Linda, and George Loewenstein.** 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives* 11 (1): 109–26.
- 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.** 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences* 115 (37): 9216–21.
- Barron, Kai, Steffen Huck, and Philippe Jehiel.** 2019. "Everyday Econometricians: Selection Neglect and Overoptimism when Learning from Others." Wissenschaftszentrum Berlin für Sozialforschung Discussion Paper SP II 2019-301.
- Bénabou, Roland, Armin Falk, and Jean Tirole.** 2019. "Narratives, Imperatives and Moral Reasoning." Unpublished.
- Bénabou, Roland, and Jean Tirole.** 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30 (3): 141–64.
- Bullock, John G., Alan S. Gerber, Seth J. Hill, Gregory A. Huber.** 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10 (4): 519–78.
- Chen, Zhuoqiong Charlie, and Tobias Gesche.** 2017. "Persistent Bias in Advice-Giving." University of Zurich Department of Economics Working Paper 228.
- Cheng, Ing-Haw, Sahil Raina, and Wei Xiong.** 2014. "Wall Street and the Housing Bubble." *American Economic Review* 104 (9): 2797–2829.
- Cochran, William G.** 1954. "The Combination of Estimates from Different Experiments." *Biometrics* 10 (1): 101–29.
- Coutts, Alexander.** 2019. "Testing Models of Belief Bias: An Experiment." *Games and Economic Behavior* 113: 549–65.
- Della Vigna, Stefano, and Elizabeth Linos.** 2020. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." NBER Working Paper 27594.
- Della Vigna, Stefano, Nicholas Otis, and Eva Vivalt.** 2020. "Forecasting the Results of Experiments: Piloting an Elicitation Strategy." *AEA Papers and Proceedings* 110: 75–79.
- Della Vigna, Stefano, and Devin Pope.** 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy* 126 (6): 2410–56.
- Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrodsky.** 2007. "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters." *Quarterly Journal of Economics* 122 (1): 209–41.
- Elliot, Andrew J., and Patricia G. Devine.** 1994. "On the Motivational Nature of Cognitive Dissonance: Dissonance as Psychological Discomfort." *Journal of Personality and Social Psychology* 67 (3): 382–94.
- Engel, Christoph, and Andreas Glöckner.** 2013. "Role-Induced Bias in Court: An Experimental Analysis." *Journal of Behavioral Decision Making* 26 (3): 272–84.
- Fedyk, Anastassia.** 2018. "Asymmetric Naivete: Beliefs about Self-Control." SSRN 2727499.
- Felton, Mark, Amanda Crowell, and Tina Liu.** 2015. "Arguing to Agree: Mitigating My-Side Bias through Consensus-Seeking Dialogue." *Written Communication* 32 (3): 317–31.
- Festinger, Leon, and James M. Carlsmith.** 1959. "Cognitive Consequences of Forced Compliance." *Journal of Abnormal Psychology* 58 (2): 203–10.
- Fiorina, Morris P., and Samuel J. Abrams.** 2008. "Political Polarization in the American Public." *Annual Review of Political Science* 11: 563–88.
- Gal, David, and Derek D. Rucker.** 2010. "When in Doubt, Shout! Paradoxical Influences of Doubt on Proselytizing." *Psychological Science* 21 (11): 1701–07.
- Ganguly, Ananda, and Joshua Tasoff.** 2016. "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future." *Management Science* 63 (12): 4037–60.
- Gentzkow, Matthew.** 2016. "Polarization in 2016." Unpublished.
- Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen.** 2020. "Bribing the Self." *Games and Economic Behavior* 120: 311–24.
- Higgins, Julian P. T., and Simon G. Thompson.** 2002. "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine* 21 (11): 1539–58.

- Hippel, Svenja, and Sven Hoepfner. 2019. "Biased Judgements of Fairness in Bargaining: A Replication in the Laboratory." *International Review of Law and Economics* 58: 63–74.
- Hossain, Tanjim, and Ryo Okui. 2013. "The Binarized Scoring Rule." *Review of Economic Studies* 80 (3): 984–1001.
- Huffman, David, Collin Raymond, and Julia Shvets. 2019. "Persistent Overconfidence and Biased Memory: Evidence from Managers." Unpublished.
- Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25 (1): 51–71.
- Iyengar, Shanto, Yphtach Leikes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22: 129–46.
- Janis, Irving L., and Bert T. King. 1954. "The Influence of Role Playing on Opinion Change." *Journal of Abnormal and Social Psychology* 49 (2): 211–18.
- Juslin, Peter, Anders Winman, and Patrik Hansson. 2007. "The Naive Intuitive Statistician: A Naive Sampling Model of Intuitive Confidence Intervals." *Psychological Review* 114 (3): 678–703.
- Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21 (2): 153–74.
- Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review* 111 (3): 831–70.
- Linnainmaa, Juhani T., Brian Melzer, and Alessandro Previtro. 2018. "The Misguided Beliefs of Financial Advisors." Kelley School of Business Research Paper 18-9.
- List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118 (1): 41–71.
- List, John A. 2006. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." *Journal of Political Economy* 114 (1): 1–37.
- Loewenstein, George, Samuel Issacharoff, Colin Camerer, and Linda Babcock. 1993. "Self-Serving Assessments of Fairness and Pretrial Bargaining." *Journal of Legal Studies* 22 (1): 135–59.
- Lord, Charles G., Mark R. Lepper, and Elizabeth Presto. 1984. "Considering the Opposite: A Corrective Strategy for Social Judgment." *Journal of Personality and Social Psychology* 47 (6): 1231–43.
- Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37 (11): 2098–2109.
- Ludwig, Sandra, and Julia Nafziger. 2011. "Beliefs about Overconfidence." *Theory and Decision* 70 (4): 475–500.
- Mayraz, Guy. 2011. "Wishful Thinking." CEP Discussion Paper 1092.
- Mercier, Hugo. 2011. "Self-Deception: Adaptation or By-Product?" *Behavioral and Brain Sciences* 34 (1): 35.
- Mill, John Stuart. 1840. "Coleridge." In *Collected Works of John Stuart Mill*, Vol. 10, 117–63. London: Routledge, 1991.
- Mercier, Hugo, and Dan Sperber. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory." *Behavioral and brain sciences* 34 (2): 57–74.
- Mullainathan, Sendhil, and Ebonya Washington. 2009. "Sticking with Your Vote: Cognitive Dissonance and Political Attitudes." *American Economic Journal: Applied Economics* 1 (1): 86–111.
- O'Neill, Patrick, and Diane E. Leving. 1979. "Inducing Biased Scanning in a Group Setting to Change Attitudes toward Bilingualism and Capital Punishment." *Journal of Personality and Social Psychology* 37 (8): 1432–38.
- Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review* 103 (2): 804–30.
- Perkins, David. 2019. "Learning to Reason: The Influence of Instruction, Prompts and Scaffolding, Metacognitive Knowledge, and General Intelligence on Informal Reasoning about Everyday Social and Political Issues." *Judgment and Decision Making* 14 (6): 624–43.
- Saccardo, Silvia, and Marta Serra-Garcia. 2020. "Cognitive Flexibility or Moral Commitment? Evidence of Anticipated Belief Distortion." Unpublished.
- Schlag, Karl H., and Joël van der Weele. 2013. "Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming Risk Neutrality." *Theoretical Economics Letters* 3 (1): 38–42.
- Schwardmann, Peter, Egon Tripodi, and Joël van der Weele. 2019. "Self-Persuasion: Evidence from Field Experiments at Two International Debating Competitions." Unpublished.
- Schwardmann, Peter, Egon Tripodi, and Joël J. van der Weele. 2022. "Replication Data for: Self-Persuasion: Evidence from Field Experiments at International Debating Competitions." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E148242V1>.

- Schwardmann, Peter, and Joël van der Weele.** 2019. "Deception and Self-Deception." *Nature Human Behavior* 3: 1055–61.
- Smith, Megan K., Robert Trivers, and William von Hippel.** 2017. "Self-Deception Facilitates Interpersonal Persuasion." *Journal of Economic Psychology* 63: 93–101.
- Soldà, Alice, Changxia Ke, Lionel Page, and William von Hippel.** 2019. "Strategically Delusional." *Experimental Economics* 23 (3): 604–31.
- Taber, Charles S., and Milton Lodge.** 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 755–69.
- Tappin, Ben M.** 2020. "Estimating the Between-Issue Variation in Party Elite Cue Effects." Unpublished.
- Tappin, Ben M., Gordon Pennycook, and David G. Rand.** 2020. "Thinking Clearly about Causal Inferences of Politically Motivated Reasoning: Why Paradigmatic Study Designs Often Prevent Causal Inference." *Current Opinion in Behavioral Sciences* 34: 81–7.
- Thompson, Leigh, and George Loewenstein.** 1992. "Egocentric Interpretations of Fairness and Interpersonal Conflict." *Organizational Behavior and Human Decision Processes* 51 (2): 176–97.
- Vinokur, Amiram, and Eugene Burstein.** 1974. "Effects of Partially Shared Persuasive Arguments on Group-Induced Shifts: A Group-Problem-Solving Approach." *Journal of Personality and Social Psychology* 29 (3): 305–15.
- von Hippel, William, and Robert Trivers.** 2011. "The Evolution and Psychology of Self-Deception." *Behavioral and Brain Sciences* 34 (1): 1–16.
- Zimmermann, Florian.** 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–61.