

Pre-analysis plan: Beliefs and attitudes in debating competitions

Peter Schwardmann* Egon Tripodi† Joël J. van der Weele‡

1. Introduction

This project consists of field experiments at two international debating competitions that take place in March 2019. The set of questions our field experiments seeks to answer as well as the data analysis can be split into two parts.

Part 1 features the first field test of the strategic self-deception hypothesis (Hippel and Trivers, 2011), which holds that people bias their beliefs to align with the position they want to convince others of because truly believing in one's position enhances persuasiveness. Debaters are randomly assigned to the position they have to argue. This allows us to test whether persuasion goals have a causal effect on private factual beliefs, attitudes, and confidence in one's own position. Our data will also allow us to ask whether debaters whose baseline beliefs are more aligned with their randomly assigned persuasion goal are more persuasive in the eyes of experienced judges.

Part 2 is concerned with the effect of debating on belief polarization in our sample of debaters. Much hope has been placed in public debate and the open exchange of ideas across the ideological divide to lead to consensus in democratic decision making. Moreover, lack of communication across the ideological divide is often blamed for increased polarization. By comparing beliefs and attitudes before and after a debate, our second set of hypotheses asks whether debating does in fact decrease polarization.

2. Experimental Setting and Design

We collect data from two international debating competitions that take place in March 2019. These tournaments exclusively feature impromptu debates, where motions are revealed only 20 minutes ahead of debates and two-player teams are randomly assigned to argue either in favor or against the given motion.

Debates take place in British Parliamentary format: four teams, with two teams representing each side of the motion (*proposition* and *opposition*), face each other in a debate. Debates have

*University of Munich, pschwardmann@gmail.com

†European University Institute and University of Bonn, egon.tripodi@eui.eu

‡University of Amsterdam, j.j.vanderweele@uva.nl

a precise structure in terms of order according to which each team speaks. While the order of teams speaking in each debate is also random, it is each team's choice to determine which team member speaks first. All speakers are given 8 minutes to present their arguments.

Both competitions feature 52 teams each and are structured in two phases: a preliminary phase and a knock-out phase. Our focus is on the preliminary phase, because it provides repeated observations of a set of outcomes without attrition. We track multiple rounds of debate (five in the first tournament and four in the second) of the preliminary phase in which all teams take part in. In each round, teams are randomly partitioned into 13 parallel debating sessions of 4 teams each, and accumulate qualifying points for the knock-out phase of the competition.¹

Our study surveys debaters (i) at the very beginning of the tournament (*baseline*), (ii) after the preparation time of each debating session, just before the debate begins (*pre-debate*), (iii) right after each debate (*post-debate*), and (iv) after all debates. The debaters' performance is evaluated by experienced judges. Enumerators, one of whom is present in each room, record additional information from watching the debate.

2.1. Surveys

2.1.1. Baseline survey

The 30 minute baseline survey features the elicitation of beliefs related to the motions of the preliminary phase of the competition. To obfuscate relation of elicitations and motions, which are meant to be secret, we ask about several topics. The baseline survey includes:

1. Questions about age, gender, country of residence, political Ideology 1-item scale, debating experience (time and achievements);
2. Fifteen incentivized belief questions;²
3. Multiple choice question on what makes a good debater.

¹The randomization of teams into different debating rooms is not in the hands of us researchers. A so called "tabbing" algorithm tries to (i) randomly match teams that have accumulated a similar score until that point of the competition, and (ii) let all teams experience all 4 speaking orders.

²Three of these questions (different across two groups of debaters) are used as control questions. In the endline survey we elicit from each participant answers to the control questions that were not asked at baseline.

2.1.2. Pre-debate survey

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

1. One belief question (every round);
2. Monetary allocation between a baseline charity (say, World Child Cancer) and a charity aligned with one of the sides represented in the debate;³
3. Prediction of share of winning teams on the Proposition side of the motion (every round);⁴
4. How many good arguments did you come up with in favor of *proposition*? What share of these arguments would you consider to be very strong? (unincentivized; every round);
5. How many good arguments did you come up with in favor of *opposition*? What share of these arguments would you consider to be very strong? (unincentivized; every round).⁵

2.1.3. Post-debate survey

This 5 minute survey after each debate includes:

1. Two belief questions (in each of the five rounds);
2. One monetary allocation between a fixed baseline charity (say, World Child Cancer) and a charity aligned with one of the sides represented in the debate;
3. Subjective ranking of team performance in the debate (unincentivized; every round);
4. Three control questions (only last round).

³We randomize the baseline charity to be one of two charities unrelated to the motions for debate.

⁴Framed as binary event: *How likely do you think it is that at least 6 teams in Proposition rank 1st in all the other 12 parallel sessions of this debate?*

⁵Most debaters also think about arguments of the other side to choose their own strategically.

2.1.4. Endline survey

A 20 minute survey at the end of the preliminary phase of the competition includes:

1. Six questions to elicit whether the 4 statements and 2 attitude elicitation related to a given motion map into a particular position (incentivized).
2. Elicitation of belief that pre-debate belief alignment helps to be more persuasive (incentivized).

2.1.5. Judges' survey

By default, judges collectively (i) score the pure quality of arguments of each speaker and (ii) provide a ranking of the four teams in the debate (see Figure 2). In addition, we ask judges to provide individual scores of each debater's overall persuasiveness before filling out the shared score sheet with other judges.

2.1.6. Enumerators' survey

During the debate, enumerators answer a survey, which includes:

1. How many times each debater stands up to challenge the speaker with a Point of Information or with a Point of Order (every round).⁶
2. A subjective rating of how heated each debaters' argumentation is coming across.
3. Attractiveness score of each debater.⁷
4. Whether statements of the survey were mentioned in the debate.

⁶This is an objective measure that can give a sense of how heated the debate is and over the debates tell us something about debaters' types. To not confuse with Point of Clarification, which is not confrontational. Definitions on <http://howtodebate.blogspot.com/2014/11/debating-glossary.html>

⁷We will have 5 measurements of debater's attractiveness by different enumerators – which we will average.

2.2. Experimental procedures

2.2.1. Motion-relevant belief elicitation

Any given question may not have a tight enough link to the motion in debaters' minds or give rise to a high degree of certainty in debaters' beliefs and may therefore be ill-suited to pick up a treatment effect. To diversify this risk, we come up with 4 questions (A, B, C, D) for each motion and administer them as illustrated in the table below. This approach also ensures that (i) no debater is asked the same question twice, and (ii) we protect the baseline and pre-debate belief elicitation from any potential information spillovers.⁸

Timing:	Beginning of Day 1	Day 1 or Day 2	
	Baseline	Pre-debate	Post-debate
Subgroup 1	A	D	B, C
Subgroup 2	B	C	A, D

Whenever we provide incentives for beliefs over a factual question, we frame the question in terms of a binary event and employ an incentive compatible elicitation procedure, i.e. a Quadratic Scoring Rule (QSR) that awards lottery tickets for predictions that are closer to the objective realized state, where the probability of winning a lottery prize coincides with the score of the QSR (Harrison et al., 2014).

We explain to subjects that only one incentivized belief question will be randomly drawn at the end of the tournament to be payoff relevant. Subjects report the chance $r_a \in [0, 100]$ that a binary statement is true. The answer on the payoff-relevant question will determine their probability of winning a 30 euro prize according to the following quadratic scoring rule:

$$\text{Probability of winning 30 euro prize}(r_a) = \begin{cases} 100 - 100 \times \left(1 - \frac{r_a}{100}\right)^2 & \text{if } E_i = A \\ 100 - 100 \times \left(0 - \frac{r_a}{100}\right)^2 & \text{if } E_i = B \end{cases}$$

2.2.2. Alignment of motion related belief questions with the motion proposition

To identify the relationship between the direction of the bias in beliefs and alignment with the motion we use the last set of questions in the Endline survey. In these questions, we ask debaters to tell us how they think people see the relationship between alignment with the proposition and facts. In particular, for each motion belief question we ask them to choose one

⁸Point (i) is important to limit the scope for both experimenter demand effects that could magnify our effects, and desire to be consistent that could attenuate our self-deception effects on pre-debate beliefs.

of three answers that identify the relationship of the motion related belief question with the proposition:

1. Most people aligned with the proposition would believe this statement to be *true*
2. Most people aligned with the proposition would believe this statement to be *false*
3. There is no systematic correlation between what people believe about this statement and alignment with the proposition

We incentivize response to this question using a procedure proposed by Krupka and Weber (2013) to elicit social norms in an incentive compatible manner.

We will have about 20 answers to identify the relationship of each motion related belief question with the proposition.

For each such belief question

- If we find that at least 51 % of the respondents chooses 1., we define alignment of beliefs as an upward bias in the prediction;
- If we find that at least 51 % of the respondents chooses 2., we define alignment of beliefs as a downward bias in the prediction;
- In all other cases, we acknowledge that we didn't do a good job at picking this specific question and we exclude responses to this question from the analysis of Hypotheses 1-8.

2.2.3. Motion-relevant attitude elicitation

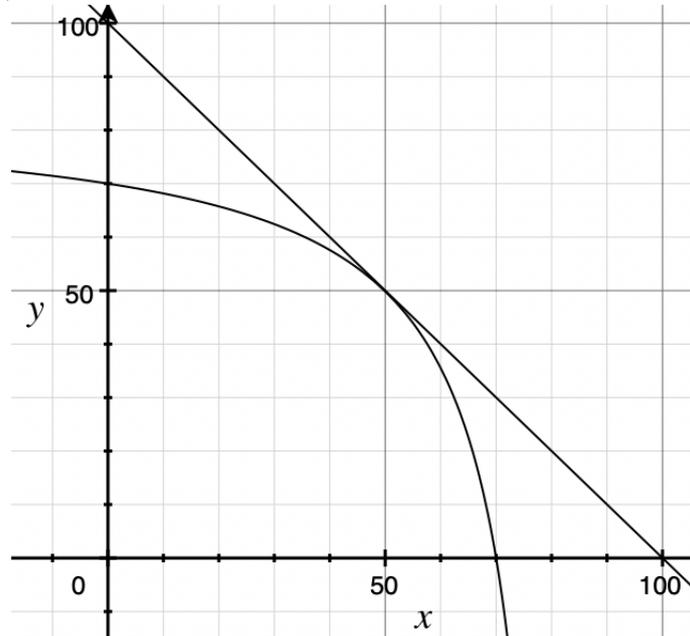
We elicit attitudes toward the motion by means of an allocation of charity donations between baseline charities, constant throughout the tournament, and a motion relevant charity.

For each elicitation, donations to the two charities are most likely close to being perfect substitutes. This in turn implies that a linear budget constraint likely leads to boundary allocations. These are informative about the preference ranking of the two charities, but give no information about the extent to which charity A is preferred to charity B.

By making the budget constraint concave (see Figure 1), we will reduce the occurrence of boundary allocations and obtain richer information about the relative weights that each debater assigns to each of the two charities.

Figure 1: Concave Charity Budget Allocation

Choose one option	<input type="checkbox"/>									
You want to give	0.0	1.3	2.5	3.7	5.0	5.9	6.4	6.7	7.0	euro to World Child Cancer
and	7.0	6.7	6.4	5.9	5.0	3.7	2.5	1.3	0.0	euro to Motion Charity
A total of	7.0	8.0	8.9	9.6	10.0	9.6	8.9	8.0	7.0	euro goes to charity



Note: Concave budget in elicitation based on the inner line of this chart.

In the endline survey, we measure the alignment of motion related attitudes with the motion's proposition in a way similar to the belief questions and exclude attitude elicitation that do not align with either side of the motion from the data analysis.

2.2.4. Confidence in own position.

To identify how confident debaters are in the quality of the arguments that are available on their side of the debate, we ask them to make incentivized predictions, in the pre-debate survey, on whether more or less than 6 teams on the Proposition side of the debate will rank first in the debates from each of the other 12 parallel sessions.

3. Hypotheses

3.1. Hypotheses about strategic self-deception

Our first set of hypotheses is based on the theory that people deceive themselves in order to better deceive others. We first ask whether private beliefs and attitudes move in the direction of randomly assigned persuasion goals and then ask whether private beliefs at baseline that are more aligned with the randomly assigned persuasion goal make debaters more persuasive.

A strong test of strategic self-deception asks whether persuasion goals distort private beliefs about facts in the direction that makes the persuasion goal easier to defend.

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

We also ask whether persuasion goals affect attitudes.

Hypothesis 2. (Secondary) *Debaters pre-debate attitudes are biased in the direction of their persuasion goal.*

A shift in attitudes may be the direct result of a shift in factual beliefs. Alternatively, it may be driven by the fact that debaters will attempt to generate many (more) arguments in favor of their side. If a debater then (intentionally or unintentionally) fails to correct for the endogenous generation of evidence when she evaluates positions before the debate, then her attitudes may shift toward her persuasion goal. We will account for the role played by this mechanism in our test of hypothesis 2, by controlling for the number of (high-quality) arguments in favor and against her position a debater generated.⁹

The strategic self-deception hypothesis suggests that belief distortions in the direction of persuasion goals occur to enhance the effectiveness of communication by credibly showing confidence in a position. We use an incentivized belief of how well other debaters in the same position will do in parallel debating sessions to provide evidence of such overconfidence.

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

The causal effect of the alignment of private beliefs with the persuasion goal on persuasiveness is why self-deception is valuable in the first place. To check whether outcomes in the debate are consistent with this account, we ask whether being randomly assigned to a more aligned persuasion goal results in greater persuasiveness.

⁹Note that debaters are uniquely suited to answer the challenging question of how many good arguments they found for their position.

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

3.2. Hypotheses about the effect of debating on belief convergence

A separate set of hypotheses is concerned with the effect of debating on consensus in our sample of debaters.

Hypothesis 5. *Post-debate attitudes are less dispersed than pre-debate attitudes.*

This hypothesis is in line with a classical Bayesian learning account and the popular idea that the open exchange of ideas leads to consensus.¹⁰

We also study the evolution of factual beliefs as we move from baseline (or pre-debate) beliefs and to post-debate beliefs.

Hypothesis 6. (Secondary) *Post-debate factual beliefs are less dispersed than pre-debate and baseline beliefs.*

In many real-world settings, individuals end up arguing the side of the debate that lines up with their baseline beliefs. This is not the case if persuasion goals are assigned at random. We also test for convergence focussing on only those subjects whose baseline beliefs are aligned with the position they are randomly assigned to.

Hypothesis 7. (Secondary) *Post-debate factual beliefs are less dispersed than pre-debate and baseline beliefs, looking at only those debaters who got to argue their baseline position.*

According to early social identity theories from social psychology (Tajfel et al., 1979), the open exchange of ideas can lead to the surprising outcome that views become more polarized after the debate than they were before if discourse is especially heated.

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

¹⁰Although perhaps less likely, the reverse could be true, if the very act of debating leads to debaters convincing themselves of their own position and the attitudes that are aligned with it.

3.3. Exploratory hypotheses

We will investigate whether post-debate beliefs and attitudes remain distorted in the direction of persuasion goals.

We will also study whether more experienced debaters are more or less likely to exhibit pre and post debate beliefs/attitudes that are aligned with their persuasion goal. This can provide some indication regarding whether self-deception is a persuasion strategy that individuals learn with experience.

We will investigate whether gender correlates with the extent to which pre and post debate beliefs/attitudes are aligned with the persuasion goal, as well as persuasiveness scores.

Both directly and through greater self-confidence, attractiveness is often thought to affect individual effectiveness in communication. We ask whether more attractive debaters are more likely to be persuasive in the eyes of judges.

4. Empirical Analysis

Testing Hypotheses 1 and 2

To test these hypotheses we use all elicitations of debater's pre-debate beliefs $pre_{b,m}$ (with $m \in \{1, \dots, M\}$) and attitudes $pre_{a,m}$.¹¹

For ease of comparability across elicitations, we conduct Normal standardizations of each belief and attitude, and impose the sign of the standardized measure $y_{i,o,m}$ to be positive (negative) if aligned with the proposition (opposition) of the debate.¹² We use the standardized outcomes $y_{i,o,m}$ for parametric tests of the hypothesis. In particular, we estimate a fixed effect regression model of the form

$$y_{i,o,m} = \alpha_i + \beta Proposition_{i,m} + \gamma X_i + \delta_m + \epsilon_{i,o,m} \quad (4.1)$$

where X_i is a vector of control variables that includes observable characteristics, δ_m are motion fixed effects, and $\epsilon_{i,o,m}$ is the error term allowing for a team component.

¹¹ M is 5 in Munich and 4 in Rotterdam.

¹²

$$y_{i,o,m} = \begin{cases} \left(\frac{pre_{i,o,m} - \mu(pre_{o,m})}{\sigma(pre_{o,m})} \right) & \text{if alignment with the motion is expected to bias debaters' beliefs/attitudes upwards} \\ (-1) \left(\frac{pre_{i,o,m} - \mu(pre_{o,m})}{\sigma(pre_{o,m})} \right) & \text{if alignment with the motion is expected to bias debaters' beliefs/attitudes downwards} \end{cases}$$

For both outcomes $o \in \{a, b\}$ separately, we estimate β and perform a one-sided t-test to assess the hypothesis that $\beta > 0$.

We take an extra step to understand the behavioral channels through which persuasion goals might affect beliefs. Anecdotally, the speaking order of teams in the debate should affect the share of pro and contra arguments that debaters develop before the debate. We exploit random variation in the team's speaking order as an instrumental variable for the share of arguments pro and contra the debater's persuasion goal. This allows us to assess the causal effect of the share of pro arguments considered just before the elicitation on pre-debate beliefs and attitudes—under the identifying assumption that the team's speaking order does not affect beliefs and attitudes directly. For both beliefs and attitudes separately we estimate the following 2SLS regression model:

$$y_{i,o,m} = \alpha_i + \beta_o \widehat{SharePro}_{i,m} + \gamma_1 Proposition_{i,m} + \gamma_2 X_i + \delta_m + \epsilon_{i,o,m} \quad (4.2)$$

where $\widehat{SharePro}_{i,m}$ are the predicted values of $SharePro_{i,m}$ (the share of arguments in favor of the proposition) from the first stage, we include $Proposition_{i,m}$ to control for the side of the motion the debater has to defend, and the error term $\epsilon_{i,o,m}$ allows for a team component. We conduct a one-sided t-test to assess the hypothesis that $\beta_a > 0$ and $\beta_b = 0$.

Testing Hypothesis 3

We ask how the prediction that at least 6 teams of debaters on the proposition side of the motion win in the other 12 parallel sessions is affected by the debater herself defending the proposition side of the motion. We do so in a fixed effect framework like the following:

$$AtLeastSixinPropWin_{i,m} = \alpha_i + \beta Proposition_{i,m} + \gamma X_i + \delta_m + \epsilon_{i,m}$$

where for each motion and debater, $Proposition_{i,m}$ denotes whether the individual debater i is on the Proposition side of the debate, X_i is a vector of controls, and $\epsilon_{i,m}$ is the error term allowing for a team component. We conduct a one-sided t-test to assess our confidence hypothesis that $\beta > 0$.

Testing Hypothesis 4

We take two steps to evaluate this hypothesis.

First, we provide correlational evidence that pre-debate alignment of beliefs with the persuasion goal predicts the average of debater's persuasiveness score as provided by three judges independently. We do this in the fixed effects regression framework

$$Persuasiveness_{i,m} = \alpha_i + \beta (\mathbb{1}_{y_{i,b,m} \geq 0} \mathbb{1}_{Proposition_{i,m}} + \mathbb{1}_{y_{i,b,m} < 0} \mathbb{1}_{Opposition_{i,m}}) + \gamma X_i + \delta_m + \epsilon_{i,m}$$

where X_i includes a score of pure quality of arguments (evaluated by judges), and standard errors clustered at the team level. We use a one-sided t-test to assess the hypothesis that $\beta > 0$.

In a second step, we rerun this analysis on replacing $y_{i,b,m}$ with its equivalent for baseline beliefs $y_{i,b,m}^{base}$:

$$Persuasiveness_{i,m} = \alpha_i + \beta_{base} (\mathbb{1}_{y_{i,b,m}^{base} \geq 0} \mathbb{1}_{Proposition_{i,m}} + \mathbb{1}_{y_{i,b,m}^{base} < 0} \mathbb{1}_{Opposition_{i,m}}) + \gamma X_i + \delta_m + \epsilon_{i,m}.$$

We use a one-sided t-test to assess our hypothesis that $\beta > \beta_{base}$.

Testing Hypotheses 5, 6, 7 and 8

For each attitude and belief question, we construct an outcome $d_{i,o,m,q,s}$ (with q being a question unique identifier and $s \in \{Baseline, Pre - debate, Post - debate\}$) that captures the absolute distance of individual response from the median response.

We then pool data from all belief elicitations and estimate the following fixed effect regression

$$d_{i,o,m,q,s} = \alpha_i + \beta_1 Baseline_{i,m,q} + \beta_2 Predebate_{i,m,q} + \gamma X_i + \delta_q + \delta_m + \epsilon_{i,o,m,q,s}$$

where $Baseline_{i,m,q}$ indicates that the outcome is measured at baseline, $Predebate_{i,m,q}$ indicates that the outcome is measured pre-debate, and the residual category is that the outcome is measured post-debate. δ_q are question fixed effects, δ_m are motion fixed effects, and $\epsilon_{i,o,m,q,s}$ is the error term allowing for a team component.

For Hypotheses 5, about convergence in attitudes, we estimate the above regression without baseline data (not collected). We then use a one-sided t-test to assess our hypothesis that $\beta_2 > 0$.

For Hypothesis 6, about convergence in beliefs, we conduct a one-sided t-test to assess the hypothesis that $\beta_1 > 0$ and a one-sided t-test to assess the hypothesis that $\beta_2 > 0$.

For Hypothesis 7, we run the same tests after estimating the regression model only for debaters whose baseline beliefs align with their persuasion goal.

To test Hypothesis 8, we use our enumerator-measured proxy of conflict in the debate. We label debates as *Heated* (*Calm*) if they score above (below) the median level of our measure of conflict. We reproduce the analysis conducted to test Hypothesis 5 and 6, including the interaction of our measure of *heat* in the debate with the indicator variables for baseline and

pre-debate.¹³ We conduct one-sided t-tests to assess our hypothesis of negative interactions between *heat* and convergence.

References

- Harrison, Glenn W, Jimmy Martínez-Correa, and J Todd Swarthout**, “Eliciting subjective probabilities with binary lotteries,” *Journal of Economic Behavior & Organization*, 2014, 101, 128–140.
- Krupka, Erin L. and Roberto A. Weber**, “Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?,” *Journal of the European Economic Association*, June 2013, 11 (3), 495–524.
- Tajfel, Henri, John C Turner, William G Austin, and Stephen Worchel**, “An integrative theory of intergroup conflict,” *Organizational identity: A reader*, 1979, pp. 56–65.
- von Hippel, William and Robert Trivers**, “The evolution and psychology of self-deception,” *Behavioral and Brain Sciences*, February 2011, 34 (1), 1–16.

¹³Our main measure of conflict is based on the objective number of times debaters stand up to challenge the speaker. Then for robustness, we combine this objective score of conflict with the subjective heat score assigned by enumerators (using principal component analysis).

A. Additional Figures and Tables

Figure 2: Judges Questionnaire

**North American Debating Championships
British Parliamentary Ballot**

Round #: _____

Room #: _____

Judges: _____

First Proposition Team					First Opposition Team				
Team Name:		Score			Team Name:		Score		
First Speaker:		<input style="width: 40px; height: 20px;" type="text"/>			First Speaker:		<input style="width: 40px; height: 20px;" type="text"/>		
Second Speaker:		<input style="width: 40px; height: 20px;" type="text"/>			Second Speaker:		<input style="width: 40px; height: 20px;" type="text"/>		
		Team Total:					Team Total:		
		<input style="width: 40px; height: 20px;" type="text"/>					<input style="width: 40px; height: 20px;" type="text"/>		
First Prop. ranked: 1st 2nd 3rd 4th <small>(Please circle.)</small>					First Opp. ranked: 1st 2nd 3rd 4th <small>(Please circle.)</small>				

Second Proposition Team					Second Opposition Team				
Team Name:		Score			Team Name:		Score		
First Speaker:		<input style="width: 40px; height: 20px;" type="text"/>			First Speaker:		<input style="width: 40px; height: 20px;" type="text"/>		
Second Speaker:		<input style="width: 40px; height: 20px;" type="text"/>			Second Speaker:		<input style="width: 40px; height: 20px;" type="text"/>		
		Team Total:					Team Total:		
		<input style="width: 40px; height: 20px;" type="text"/>					<input style="width: 40px; height: 20px;" type="text"/>		
Second Prop. ranked: 1st 2nd 3rd 4th <small>(Please circle.)</small>					Second Opp. ranked: 1st 2nd 3rd 4th <small>(Please circle.)</small>				