Pride and Prejudice? Structural Evidence of Social Pressure from a Natural Field Experiment with Committees*

Tom Gole[†] and Simon Quinn[‡]

September 2016

Abstract

When members of a committee face social pressure to agree with each other, they overweight public information, generating status quo bias. We test this behavioural hypothesis by running a novel field experiment — a large debate tournament with random assignment of judges to committees. To analyse our experimental data, we build a structural model for estimating discrete Bayesian games with correlated unobservable signals and with dynamic updating of coordination preferences. Our estimates show that, in a committee context, social pressure can cause coordination on weaker candidates.

*We thank Phillipe Aghion, George-Marios Angeletos, Richard Blundell, Steve Bond, Pierre-André Chiappori, Ian Crawford, Vince Crawford, Elwyn Davies, Stefano DellaVigna, Markus Eberhardt, Chris Erskine, İrem Güçeri, Daniel Gutknecht, Oliver Hart, Cecilia Heyes, Richard Holden, Michael Koelle, Clare Leaver, Erin O'Brien, Rossa O'Keeffe-O'Donovan, Gilat Levy, Julien Labonne, Shengwu Li, Eric Maskin, Aureo de Paula, Matthew Rabin, Gautam Rao, Chris Roth, Bertel Schjerning, Matt Shum, Kathy Spier and Ken Wolpin for helpful comments and feedback. An earlier version of this work formed part of Tom Gole's Ph.D dissertation at Harvard University; we particularly thank the members of Tom's committee: Alberto Alesina, David Laibson and Andrei Shleifer. We also thank seminar audiences at the Australian National University, Harvard University, the Queensland University of Technology, the University of Copenhagen, the University of New South Wales, the University of Oxford and the University of Queensland.

[†]**Boston Consulting Group**.

[‡]**Department of Economics, University of Oxford**; simon.quinn@economics.ox.ac.uk.

1 Committees and social pressure

Social pressure can provide powerful behavioural incentives for pro-social behaviour — for example, by encouraging voting (DellaVigna, List, Malmendier, and Rao, 2015; Gerber, Green, and Larimer, 2008), charitable donations (DellaVigna, List, and Malmendier, 2012) and effort in the workplace (Mas and Moretti, 2009). But social pressure can also have a dark side — for example, by biasing individuals' expressed opinions towards the preferences of partisan observers (Asch, 1951; Garicano, Palacios-Huerta, and Prendergast, 2005; Parsons, Sulaeman, Yates, and Hamermesh, 2011), or by discouraging effort in order to conform to norms against high achievement (Bursztyn and Jensen, 2015).

In this paper, we highlight a new mechanism by which social pressure can hamper effective decision-making: by distorting the preferences of participants in committees. Our hypothesis is simple: *when members of a committee face social pressure to agree with each other, they over-weight public information in reaching their decisions*. In effect, a committee can operate like a Keynesian beauty contest, where participants worry not only about their own perceptions of the candidates, but also about the perceptions of others (Keynes, 1936). To test the magnitude of this distorting effect, we run a large randomized field experiment with a novel design, in which participants are repeatedly assigned to different three-member committees to assess competing teams in a tournament; this means that we randomize both (i) the formation of committees and (ii) the assignment of different committees to assess different competing teams. To analyse our experimental data, we build a structural model in which we treat committee decisions as the Markov Perfect Equilibrium of a series of discrete Bayesian games with correlated signals. Specifically, we model each debate as generating a set of stochastic, private but correlated signals — with teams' relative pre-tournament rankings as public information. In doing so, we follow a recent literature that combines natural field experiments with structural models to understand the role of behavioural preferences (see DellaVigna, List, Malmendier, and Rao (2016), DellaVigna, List, Malmendier, and Rao (2015), Rao (2015), Huck, Rasul, and Shephard (2015), DellaVigna, List, and Malmendier (2012) and Card, DellaVigna, and Malmendier (2011)).[1] As in DellaVigna, List, Malmendier, and Rao (2016), we pre-register predictions from our structural model; we check those predictions against a

[1] A closely related literature estimates behavioural parameters from structural models and non-experimental data: see DellaVigna, Lindner, and Schmieder (2015), Barseghyan, Molinari, O'Donoghue, and Teitelbaum (2013) and Laibson, Repetto, and Tobacman (2007).

subsequent iteration of our experiment as a form of out-of-sample model validation.

Using our structural model, we show that dissenting from peers causes committee members to have a stronger preference for future agreement. Markov Perfect Equilibrium allows us to analyse this with a dynamic structure, in which committee members optimally engage in status quo bias to limit the total number of dissents they anticipate over some future horizon. To understand the magnitude of our results, we propose a new measure of equivalent variation for discrete games, by exploiting predicted behaviour under what we term a 'Faustian option'. To do this, we imagine a situation in which committee members — like the famous Faust of German legend — are presented the option of 'selling' their opportunity to express their opinion, in exchange for avoiding the risk of dissenting. Our model makes specific predictions about the proportion of committee members who would take such an option; this proportion then serves as a useful measure for the strength of coordination preferences. This measure reveals large and significant preferences for coordination; some committee members would be willing to forgo between approximately 20% and 50% of their decisions in order to avoid the risk of dissent.

This shows that, in a committee context, public information can cause coordination on weaker candidates. Because signals are relatively noisy, however, we find that the practical effect of coordination preferences on committee outcomes is limited: for relatively close decisions, the preference for coordination changes the committee outcome in fewer than 10 percent of cases. In sum, we find that committees can serve as a mechanism by which social pressure generates status quo bias; however, we also show that the magnitude of such effects is likely to be limited in environments where participants face substantial uncertainty about their peers' signals.

In doing so, we contribute to several bodies of literature. Most directly, we contribute to empirical behavioural literature on the role of social pressure. DellaVigna, List, and Malmendier (2012) combine a randomized field experiment with a structural model to estimate the role of social pressure in encouraging charitable donations to door-to-door solicitation. DellaVigna, List, Malmendier, and Rao (2015) and Gerber, Green, and Larimer (2008) both use randomized field experiments (and, in the case of DellaVigna, List, Malmendier, and Rao (2015), a structural model) to estimate the importance of social image for encouraging voting. Similarly, several recent experiments highlight the role of social comparisons upon employee performance (Breza, Kaur, and Shamdasani, 2015; Cohn, Fehr, Herrmann, and Schneider, 2014; Bandiera, Barankay, and Rasul, 2010). We build upon these results by providing the first empirical evidence that, in a real-world setting, such social comparisons can drive bias in collective decision-making. This result provides empirical support for several theoretical assertions about committee behaviour. For example, Levy (2007) models the effect of transparency on committees, arguing that committee members' career concerns can encourage members to 'conform to preexisting biases'.[2] Our results emphasize that dissent aversion can be understood as a phenomenon that changes dynamically depending on circumstances and incentives, rather than being solely a psychological or 'knee-jerk' response to social context.

Second, our results complement the large empirical literature on statistical discrimination, by showing a new mechanism by which such discrimination can arise. It has long been recognized that exclusion of disadvantaged groups can be reinforced through discriminatory beliefs (Phelps, 1972; Arrow, 1973; Coate and Loury, 1993), a point highlighted in a large number of empirical studies (see, for example, Goldin and Rouse (2000), Bertrand and Mullainathan (2004), Lavy (2008), Bagues and Esteve-Volart (2010), Price and Wolfers (2010), Anwar, Bayer, and Hjalmarsson (2012) and Bartoš, Bauer, Chytilová, and Matějka (2016)). A recent behavioural literature highlights the role of social familiarity in helping to overcome such discrimination: see Rao (2015) and Burns, Corno, and La Ferrara (2016). We contribute to this literature by highlighting salient institutional features — namely, the structure of committee voting — that might exacerbate the role of social pressure in encouraging a decision-maker to place more emphasis on commonly-held beliefs about background.

Finally, we build on recent literature on the structural estimation of discrete games, by developing a new parametric framework for incorporating correlated signals. Previous work on structural estimation of Bayesian games (for example, Bajari, Hong, Krainer, and Nekipelov (2010) and de Paula and Tang (2012)) has generally relied upon the assumption that, conditional on covariates observable to the researcher, players' signals are independent. Signal correlation is central to our empirical context; each committee observes the same debate, and each committee member casts a vote based on her or his perception of what the committee observed together. However, as researchers, we can predict teams' relative performance using only teams' pre-tournament rankings. This is informative, but cannot capture any of the idiosyncracies of individual teams' relative performances in particular debates; for this

[2] Visser and Swank (2007) consider a model in which committee members prefer to conceal disagreement, in order to preserve their reputation.

reason, it would be unreasonable to model our experiment as involving conditionally independent signals. Recent theoretical work shows conditions under which identification under correlated signals may be achieved using a very rich support assumption: for example, Xu (2014) and Wan and Xu (2014) exploit an assumption that the researcher observes continuously distributed and excludable covariates.[3] However, in many empirical contexts — including ours — it is not possible to find continuously distributed and excludable covariates. We show how discrete shocks to player preferences — the kind of shocks observed in many experimental settings, including ours — can be exploited to point-identify the magnitude of players' preference for coordination in a stage game, even under correlated signals. We then show how such strategic committee voting can be nested in a finite dynamic framework. Our consequent structural estimates allow us to interpret our experimental results in terms of underlying preferences.

Five sections follow. Section 2 describes our experimental design. Section 3 develops a new structural model and shows conditions for identification. We discuss the results in section 4, and show robustness in section 5. We conclude in section 6.

2 A novel field experiment

In private corporations and in government agencies, most important decisions are taken by committees. The rules that govern committee voting can be critical for decisions as diverse as the hiring of job candidates (Goldin and Rouse, 2000), the setting of monetary policy (Hansen and McMahon, forthcoming; Hansen, McMahon, and Rivera, 2014; Riboni and Ruge-Murcia, 2014; Jung, 2011), and the determinations of courts of law (Iaryczower, Shi, and Shum, 2013; Iaryczower and Shum, 2012; Fischman, 2011; Blanes i Vidal and Leaver, 2013; Levy, 2005).

For causal inference on the role of committee coordination preferences, researchers ideally need an experimental context with several quite peculiar features. First, participants should be assigned to committees randomly, so that observed behavior cannot be attributed to endogenous committee formation. Second, participants should face random or quasi-random shocks to their preferences over coordination, and

[3] As Wan and Xu (2014, p.237) explain, this "requires that for each player there exists a special regressor which is continuously distributed and has unbounded support conditional on the rest of regressors. We require that the conditioning variables include not only the rest of the regressors of player j, but also the rival's regressors, which implies an exclusion restriction". See also Liu, Vuong, and Xu (2014).

these shocks should be asymmetric between different committee members; this allows researchers to measure the effect of coordination preferences, distinct from the effect of other information that a committee may receive. Third, committee voting procedures should be specific and standardized across different committees. Finally, researchers should ideally study a situation where payoffs are meaningful and where participants are very familiar with the context and the committee protocols — that is, a 'natural field experiment' (Harrison and List, 2004).[4] We run a novel experiment that has all of these features. To our knowledge, no previous research has considered committee voting in a randomized field context. In this way, the present experimental context provides a new method for testing committee interactions.

2.1 The World Schools Debating Championships as a natural field experiment

The World Schools Debating Championships are an annual debate tournament between high school students. Debaters are drawn from around the world to represent their countries; each nation is entitled to one team in the competition. The Championships are the premiere international debate tournament for school students. We study the Championships held in 2010 (in Doha, Qatar), in 2011 (in Dundee, Scotland) and in 2012 (in Cape Town, South Africa). A total of 66 countries competed at these three tournaments, of which 39 countries participated at all three.[5]

Each debate pitches one national team against another; teams are randomly assigned to argue either for or against a controversial idea.[6] The Championships comprise both Preliminary Rounds and Finals Rounds. In the Preliminary Rounds, each nation competes against eight randomly-drawn opponents. These eight debates occur across four days: Rounds 1 and 2 on the first day, Rounds 3 and 4 on the second day, and so on. The top 16 teams then progress to the Finals Rounds, a series of four

[4] Harrison and List (2004) describe a natural field experiment as a context "where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are in an experiment" (page 1014). In our context, participants were told that data would be collected about their decisions, and that this might be used for academic research in economics.

[5] Extensive information on the Championships — including on the rules and history of the tournament — is available at the official website: `http://www.schoolsdebate.com/`.

[6] For example, the 2010 Championships began with a debate on the proposition "That we should support military intervention in Somalia"; the same Championships ended with a Grand Final debate on the proposition "That governments should never bail out big companies".

knock-out debates culminating in the Grand Final.[7] Our analysis focuses exclusively on data from the Preliminary Rounds.[8]

The winner of each debate is determined by a committee of three judges. Together, this committee is required to decide which team has argued more persuasively.[9] Judges assess the debate separately. Each judge is required to complete a ballot, in which he or she records speaker points and decides the winner of the debate; judges may not award a tie. An example ballot is provided in Online Appendix B, along with explanation on the marking categories. The debate is won by whichever team wins two or three of the judges; committee outcomes can therefore be either 'unanimous' (3-0) or 'split' (2-1).

Critically, judges are not allowed to communicate with each other (or with the competitors) until *after* making their decisions.[10] Having made their decisions, the three judges then leave the room to confer; judges may not change their decisions after leaving the room. Having discussed the debate together, the committee returns to the room; one judge announces the committee's result, and gives a brief justification for the committee's decision. Teams and their coaches are then encouraged to speak separately with the judges; at this point, there is a strong emphasis on constructive feedback.

In this way, the Championships provide an ideal field experiment in which to study the consequences of committee voting for the expression of members' private information. The literature on committee voting mentioned tends to emphasize two distinct roles for committee processes: (i) aggregation of disparate information and (ii) communication/persuasion between committee members (see, for example, Austen-Smith and Feddersen (2009, 2006) and Feddersen and Pesendorfer (1996)). Commu-

7 For example, in 2010, Canada won all eight of its Preliminary Round debates, defeating (in order) Bangladesh, Botswana, Thailand, Argentina, Namibia, South Korea, Palestine and Pakistan. In the Finals Rounds, the team then defeated Ireland, New Zealand, Singapore and England (in order), to become the World Schools Debating Champions.

8 We limit our sample in this way because judge assignment for the Finals Rounds is not random.

9 That is, the committee does *not* decide whether it agrees or disagrees with the proposition being debated. Judges' personal views about the issue under debate are not considered to be relevant for the assessment of which team has better argued its case.

10 Judges are seated apart. There is no evidence of judges trying to 'cheat' by looking at each other's notes; indeed, there are strong norms at the Championships against such behavior. Judges are also discouraged from allowing their facial expressions or body language to indicate their views on the debate.

nication/persuasion is an important aspect of many real-world committees. However, the effect that we study is an effect on committee members' expression of private information; that is, an incentive that may discourage committee members from sharing their private perceptions in a completely informative way. It is critical for our field experiment that judges cannot communicate before they vote, because this allows us to isolate the effect of past dissent on each judge's *individual* decision.

2.2 Implementing the field experiment

We implemented random assignment of judges to committees at the World Schools Debating Championships.[11] Judges were assigned to committees randomly in each round (using a computer), and judges knew this.[12] This assignment was subject to several constraints, designed to improve the 'balance' of the randomization. Most importantly, we stratified the randomisation based on experience/competence: each committee comprised one 'class 1' judge (most experienced/competent), one 'class 2' judge and one 'class 3' judge (least experienced/competent). These classes were assigned subjectively by the tournament organizers, to ensure a balance of judging experience across different committees. Second, each committee included at least one man and one woman (58% of all judges being male).[13] Third, where possible, we sought to limit cases of judges seeing the same team more than once in the same tournament, and no judge was allowed to assess his or her country's team.[14] Fourth, because the Preliminary Rounds were usually divided between different venues (often high schools), we often needed to assign pairs of judges to committees together in two debates on the same day.

[11] Quite apart from research purposes, there are numerous advantages of using random assignment of judges to committees in the preliminary stages of the Championships; among them, (i) this ensures a fair and balanced assignment of judge quality to all teams in the tournament, (ii) this ensures a fair and balanced assignment of judges of differing abilities to debates of different standards, and (iii) though computationally demanding, this approach avoids a lengthy process by which organizers need manually to form judging committees.

[12] The computer code was written in Stata, and is available on request.

[13] We were required to relax this constraint four times: in 2010, we allowed one all-male committee, in 2011, we allowed two all-female committees, and in 2012, we allowed one all-male committee.

[14] Readers may nonetheless be concerned about the incentive for judges to make decisions that help the position of their national team in the overall standings. There are several reasons that we do not believe that this is a common phenomenon. First, the complexity of the tournament often makes it difficult for participants to know how particular results may or may not assist other competing teams. Further, this kind of strategy could easily backfire through substantial reputational harm both to the individual judge and to the tournament as a whole — and, as we note shortly, such reputation appears to matter for participating judges.

2.3 Pre-tournament rankings as a public signal

Teams are ranked before each tournament. On the basis of this ranking, a random draw determines each team's position in the draw.[15] Pre-tournament ranking is necessary so that each team is drawn against opponents of a range of different qualities; *i.e.* so that a team does not face a disproportionate number of very strong teams in its Preliminary Rounds, nor a disproportionate number of weaker teams.[16] Teams are ranked on the basis of their performance in the Preliminary Rounds of the three previous tournaments.[17] The pre-tournament rankings are therefore critical public information about recent results. In the analysis that follows, we use this information as a proxy for judges' *a priori* expectations about teams' quality; for example, we define the 'favorite' in any debate as being the team with the better pre-tournament ranking.

There are two complementary reasons that these rankings work well as a proxy for judges' expectations of teams' performance. First, teams' approximate position in the rankings is well known by almost all judges. Second, even if a judge is not directly aware of teams' rankings, almost all judges know about teams' performance in recent tournaments; that is, the judges are generally aware of the underlying information on which the rankings are based.

2.4 Description of the main variables

Judges come from around the world to participate in the Championships; across the three tournaments studied, a total of 49 nations were represented on various judging committees. Judges are volunteers, and most are required to pay their own travel and accommodation expenses to participate.[18] Most judges are young and highly educated.

Following the 2012 Championships, we ran an online survey of the judges who participated in the 2010, 2011 and 2012 Champsionships. Of the 222 judges who par-

[15] This random draw is filmed and made available online; draw videos for several WSDC tournaments are currently available on YouTube.

[16] Many tournaments use a similar approach for seeding a random draw — including, for example, the FIFA World Cup.

[17] Ranking documents are available at `http://www.schoolsdebate.com/`.

[18] Some nations subsidize their judges' expenses. Additionally, in 2010, the host organization (QatarDebate) paid the travel and expenses of 37 experienced judges, in order to ensure that sufficient judges were able to participate in the tournament.

ticipated in the three tournaments we study, 174 answered the online survey (*i.e.* about 78%).[19] That survey showed that, in 2012, the median age of the judges being studied was 27. All judges had completed secondary school; 70% had completed an undergraduate degree, and 40% had completed a postgraduate degree (primarily in social sciences and humanities — for example, in politics, English, law, economics or history).

In total, we study 603 committees: 220 in the 2010 tournament, 192 in the 2011 tournament and 191 in the 2012 tournament.[20] In Table 1, we describe the key variables that will form the basis of our subsequent analysis: we show the number of committees in which each class of judge votes for the favourite, and the consequent probability of the favorite winning the committee's vote. Of these 603 committees, 456 decided for the favorite (76%); of these 336 voted unanimously, and 120 were split. Of the remaining 147 committees that decided against the favourite (24%), 71 voted unanimously and 76 were split.

< Table 1 here. >

As one would expect, the difference between the rankings of two opposing teams is a significant predictor of teams' performance; as the ranking difference between two opponents narrows, the probability of the favorite winning decreases and the probability of judge disagreement increases. To test this, we ran two probit models with 'ranking difference' as the sole explanatory variable. In the first descriptive probit, the outcome was whether the favorite wins; the estimated average marginal effect of ranking difference was about 1.5 percentage points. In the second descriptive probit, the outcome was whether the judging committee was split; the estimated average marginal effect was just under -1 percentage point. In both cases, 'ranking difference' was significant with $p < 0.001$. In Figure 1, we show the predicted values for both of these estimations; we also superimpose the density of ranking differences.

[19] We do not use this survey data for any substantive analysis, but we feel that it provides a reasonable description of judge characteristics.

[20] In the 2010 tournament, we had 28 debates in each round. However, one of the authors (Quinn) was one of the two Chief Adjudicators of that tournament, and was required to judge on four committees (in rounds 3, 4, 5 and 6); we dropped those committees from the dataset before performing any analysis. In 2010, we needed four extra debates — in a notional 'Round 0' — in order to ensure that the draw was balanced between the 57 competing teams. We have dropped these committees. In the 2011 and 2012 tournaments, we had 24 debates in each round. However, one judge fell ill during one debate in 2012, and was required to withdraw from the committee decision; we have also dropped that committee.

The figure establishes three important stylized facts. First, as the first descriptive probit indicated, the favorite's probability of winning increases from approximately 0.5 (for teams that are almost equally ranked) to 1 (for teams that are ranked far apart). Second, as the second descriptive probit indicated, the probability of a split committee decreases from approximate 0.45 (for teams that are almost equally ranked) to approximately 0.1 (for teams that are ranked far apart). Finally, as the histogram shows, we have substantial variation in this ranking diffference in our dataset; this is very useful, because it allows valuable exogenous variation that will assist in identifying coordination preferences.

< Figure 1 here. >

2.5 Descriptive evidence of social pressure and status quo bias

Judges at WSDC face two primary incentives. First, every judge wants to make the 'correct' decision, by voting for the team that is more deserving of a win. There are strong norms in the international debate community — and a large degree of professional respect — for being a competent judge who accurately recognizes effective debating. Many judges also pay substantial sums of money to attend the tournament, and generally take pride in participating in a high-quality tournament that is judged fairly. Second, many judges may prefer to avoid dissenting from their peers. For some judges, at least, there are strong norms that dissent is embarrassing: dissent can be seen as a strong indicator of having made the 'wrong' decision. In our online survey conducted after the 2012 tournament, we found that a substantial share of judges acknowledged these kind of attitudes: 22% of respondent judges agreed that "in general, better judges are less likely to dissent", and 37% agreed that "I am more likely to worry that I have made a bad decision when I have dissented than when the result is unanimous".

In Figure 2, we provide suggestive descriptive evidence linking past dissent stock to the probability of voting for the favorite. In that figure, we show the sample probability of voting for the favorite, conditional on (i) the tournament round and (ii) the judge's stock of dissents entering that round. (In parentheses, we show the number of observations for each combination of round and dissent stock.) We see suggestive evidence that judges with a higher dissent stock are more likely to vote for the favorite — something that would be consistent with a mechanism by which dissenting increases judges' social pressure to conform, and therefore distorts votes in favour of the pre-tournament favorite. To be sure, these differences are not statistically significant in this simple descriptive — our experiment is not powered to detect such differences without placing more structure on the problem — but this captures the intuition of the key hypothesis that we want to test.

< Figure 2 here. >

These two primary incentives — to balance the desire to make the 'correct' decision and the social pressure not to dissent — are not limited to judges at WSDC. One would expect similar incentives for members of a typical hiring committee — where, for example, each member may have a preference over which candidate is hired, and an incentive to agree with other committee members. Similarly, a rich literature in the field of law and economics shows that many judges on courts of law have a preference for judicial consensus, in addition to a preference about the relative merits of the parties' arguments. For example, Fischman (2011) estimates a strong 'cost of dissent' among judges on the Ninth Circuit Court of Appeals considering asylum cases (see also Fischman (2013), Lindquist (2006), Revesz (1997) and Cross and Tiller (1998)). Fischman and other authors suggest several complementary reasons that judges may hold such preferences; for example, Posner (2008) describes norms of 'dissent aversion' and of collegiality (see also Ginsburg (1992) and Seitz (1991)), and Epstein, Landes, and Posner (2011) note that dissenting judges may need to spend more time to justify their decisions (see also Posner (1993), in discussing 'going-along' judicial voting). Sunstein (2000) relates such behavior to a 'reputational externality' accompanying the expression of individual opinions; this links to a large experimental literature in social psychology, studying the effect of 'hidden profile' information (Asch, 1951; Stasser and Titus, 1985; Christensen, Larson, Abbott, Ardolino, Franz, and Pfeiffer, 2000; Lu, Yuan, and McLeod, 2012).

In this way, our experiment asks empirical questions that are relevant for a wide range of contexts — all of which, like WSDC, involve a trade-off for committee members between following their personal judgement and avoiding the risk of dissent. *How much weight do committee members place upon their individual perceptions?* Conversely, *how much do committee members value agreement with peers? How much — if at* *all — does this distort the decision-making process?*

To answer these questions, we build and estimate a dynamic structural model. We take this structural approach for three related reasons. First, the questions we pose concern 'deep' preferences for coordination; we want to be able to interpret our experimental results directly in terms of those preferences. Second, we want to be able to predict behaviour under counterfactuals not observed in the dataset (in particular, to calculate the equivalent variation, using our 'Faustian option') — an objective for which structural models are particularly well suited.[21] Finally — and more pragmatically — the phenomenon we describe is highly non-linear in observable covariates (in particular, in the various dissent stocks and the ranking difference); our model improves efficiency by imposing structure on the problem.

3 A structural model of committee behavior under social pressure

Structural models are increasingly relevant for the analysis of randomized field experiments, particularly in contexts where, as here, researchers are concerned to compare the magnitude of underlying preferences, and to understand likely behavior under a counter-factual (see, for example, DellaVigna, List, Malmendier, and Rao (2016), DellaVigna, List, Malmendier, and Rao (2015), Rao (2015), Huck, Rasul, and Shephard (2015), Dupas (2014), DellaVigna, List, and Malmendier (2012), Duflo, Hanna, and Ryan (2012), Attanasio, Meghir, and Santiago (2012), Todd and Wolpin (2006) and Shearer (2004)). At the heart of our structural model is a Bayesian game: a game of incomplete information, in which each player receives a signal and then acts in anticipation of the other players' choices.

In this paper, we use a Bayesian game to formalize our intuition that committee members can engage in status quo bias to increase the chances of agreement; that is, debate teams' pre-tournament rankings can, like the statements of a central bank in a currency crisis, act as public information that plays a coordination role. Specifically, we model each player as taking a binary decision, and we model the distribution

[21] Duflo, Hanna, and Ryan (2012, page 1265) make a similar point in justifying the use of structural models in experimental analysis, and one that applies directly to our problem: "A primary benefit of estimating a structural model of behavior is the ability to calculate outcomes under economic environments not observed in the data."

of players' signals as trivariate normal. The trivariate normal provides an elegant structure for allowing correlation between players' unobservable signals. In some empirical contexts, it may be entirely reasonable to assume that, conditional on variables observable to the researcher, players' signals are independent: see, for example, de Paula and Tang (2012) (who study programming of radio commercial breaks) and Bajari, Hong, Krainer, and Nekipelov (2010) (on stock recommendations by equity analysts). But, in many contexts, it is unreasonable to assume that the common elements to players' signals are observed by the researcher. The present empirical context provides one illustration: each committee watches the same debate, so receives correlated signals (in the form of speakers' presentations), and those signals cannot be captured fully by any variables observed by the researcher.[22] The trivariate normal implies a very convenient form for each player's best response function, as well as a calculable log-likelihood — and does so while allowing for correlated player signals. To avoid resting our identification solely upon a distributional assumption, we exploit past dissent as an excludable shock to player payoffs.[23]

3.1 Specifying and solving the stage game

Figure 3 shows the evolution of the dissent stock for different judges across different tournament rounds, where the numbers on the nodes represent the number of judges. This two-dimensional space will, in due course, form the state space for the dynamic structural model; ultimately, our interest is in estimating a dynamic repeated game, in which players gain utility from their stock of dissents at the end of a tournament, and then reason backwards to engage strategically in status quo bias. However, it is not possible to solve this model — nor to show identification — without first considering a one-shot static game. This static game forms the stage game for our dynamic model.

[22] Our structural estimates, reported shortly, support this claim; we strongly reject a null hypothesis that signals are conditionally independent. In section 4.5, we show that we obtain substantially different estimates of the preference for coordination if we use the misspecified model with independent signals.

[23] In this sense, we take a broadly similar approach to Grieco (2014), who uses excludable covariates and a bivariate normal distribution to identify a binary choice game between two players. However, our paper differs from Grieco (2014) in the form of the information structure: Grieco uses 'public shocks', observed by the players but not the researcher, and two conditionally independent 'private shocks', each observable only to its respective player. This would be similar to our approach in a two-player game, but not for more than two players – by allowing correlated signals, we allow different players to perceive each other's signals with different precisions. We also differ from Grieco (2014) in treating the discrete stage game as part of a dynamic process.

< Figure 3 here. >

Stage game model setup: We model each separate committee as Bayesian game between three players. For each committee, we denote judge class by $i \in \{1, 2, 3\}$. Each judge i receives a signal, x_i, of the relative performance in the debate of the *ex ante* favorite (where $x_i > 0$ implies that the judge thought the favorite performed better in the debate, and $x_i < 0$ implies the underdog performed better), and then chooses whether to vote for the favorite ($a_i = 1$) or against ($a_i = 0$). Judge i receives utility from two mechanisms: (i) from voting for the team that (s)he prefers (where the strength of that preference is determined by the signal x_i), and (ii) from agreeing with judge j and/or with judge k. We treat these mechanisms as additively separable.[24]

Assumption 1 (Player utility in the stage game)

$$U_i(a_i; a_j, a_k, x_i) = \begin{cases} x_i + \delta_i & \textit{if } a_i = 1, a_j = 1, a_k = 1; \\ x_i + \delta_{ij} & \textit{if } a_i = 1, a_j = 1, a_k = 0; \\ x_i + \delta_{ik} & \textit{if } a_i = 1, a_j = 0, a_k = 1; \\ x_i & \textit{if } a_i = 1, a_j = 0, a_k = 0; \\ 0 & \textit{if } a_i = 0, a_j = 1, a_k = 1; \\ \delta_{ik} & \textit{if } a_i = 0, a_j = 1, a_k = 0; \\ \delta_{ij} & \textit{if } a_i = 0, a_j = 0, a_k = 1; \\ \delta_i & \textit{if } a_i = 0, a_j = 0, a_k = 0. \end{cases} \tag{1}$$

In this model, the δ parameters are therefore the key behavioral parameters of interest. Note that, for example, δ_{ij} measures the utility gain for judge i from voting with judge j, and that δ_i measures the gain from agreeing with *both* judges k and j. We assume that each judge weakly prefers agreement over dissent, and weakly prefers agreeing with both peers over agreeing with just one: $\delta_i \geq \delta_{ij} \geq 0; \delta_i \geq \delta_{ik} \geq 0$.

In taking this approach, we treat voting in this context as 'expressive' rather than 'instrumental' (see Hamlin and Jennings (2011) and Schuessler (2000)). This follows a common approach in empirical voting literature, where voters are assumed to choose the candidate who is closest to them in an ideological space (see, for example, Downs

[24] This kind of additively separable 'reduced form' specification is common for structural models of incomplete information: see, for example, Grieco (2014), de Paula and Tang (2012), Bajari, Hong, Krainer, and Nekipelov (2010) and Aguirregabiria and Mira (2007). It is also useful in many applications for structural models of games with complete information: see, for example, Tamer (2003) and Ciliberto and Tamer (2009).

(1957), Black (1958), Hinich and Munger (1997), Degan and Merlo (2009) and Merlo and de Paula (forthcoming)). Indeed, our utility function can be viewed as a simple application of the 'spatial theory of voting', in which x_i measures the difference between a judge's distance to the favourite and the judge's distance to the underdog.[25] We take this approach for two reasons. First, and most importantly, this accords strongly with the norms in our empirical context — in which judges are encouraged to vote sincerely to reflect their perception of a debate, and in which judges cannot engage in strategic communication prior to voting (see Austen-Smith and Feddersen (2005) and Gerardi and Yariv (2007)). Second, as the following section shows, the assumption allows for an additive random utility framework, which makes our model tractable to empirical analysis.

Assumption 2 (Distribution of signals) *For each committee, the distribution of signals is trivariate normal (where we assume non-negative correlations, $\rho_{12}, \rho_{13}, \rho_{23} \geq 0$):*

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \right). \tag{2}$$

The signal x_i is composed of of a publicly observable mean (μ_i), which reflects judge i's ex-ante view on the difference in quality between the two teams, and a privately observed but correlated term, which reflects their view on the outcome of the actual debate.[26] x_i therefore plays a dual role (see Aradillas-Lopez and Tamer (2008)): it directly affects the relative utility of voting for the favorite, and it determines the conditional expectation of the other judges' signals:

$$\begin{pmatrix} x_j \\ x_k \end{pmatrix} \Bigg| \; x_i \sim \mathcal{N}\left[\begin{pmatrix} \mu_j + \rho_{ij} \cdot (x_i - \mu_i) \\ \mu_k + \rho_{ik} \cdot (x_i - \mu_i) \end{pmatrix}, \begin{pmatrix} 1 - \rho_{ij}^2 & \rho_{jk} - \rho_{ij} \cdot \rho_{ik} \\ \rho_{jk} - \rho_{ij} \cdot \rho_{ik} & 1 - \rho_{ik}^2 \end{pmatrix} \right]. \tag{3}$$

[25] This approach differs from much of the the extensive theoretical literature on voting in political economy — such as Austen-Smith and Banks (1996), Austen-Smith and Feddersen (2006), Feddersen and Pesendorfer (1996), Feddersen and Pesendorfer (1997) and Feddersen and Pesendorfer (1998) — where voter utility depends upon the outcome of the vote.

[26] Note that, alternatively, it would be possible to express a model in which judges vote based on inferences about a common underlying fundamental (plus dissent aversion). This comes, however, at the cost of increasing the model complexity, and so we have opted on the side of parsimony. One argument for such a 'common underlying fundamental' approach is that it suggests an alternative channel for our results — namely, that when judges disagree, they revise their beliefs about the precision of their signals, and so begin to weight the common fundamental more heavily. We test for this alternative 'tuning' hypothesis in the context of our empirical results in Section 5.3, and find no support for it.

Judge i must choose a best response $a_i^*(x_i)$. We limit attention to cutoff strategies: $a_i^*(x_i) = \mathbf{1}(x_i \geq x_i^*)$, where x_i^* denotes the cutoff and $\mathbf{1}(\cdot)$ the indicator function.[27] At $x_i = x_i^*$, judge i must be indifferent between $a_i = 0$ and $a_i = 1$; that is,

$$\begin{aligned} x_i^* = & \left[\Pr\left(a_j = 0, a_k = 0 \mid x_i = x_i^*\right) - \Pr\left(a_j = 1, a_k = 1 \mid x_i = x_i^*\right)\right] \cdot \delta_i + \\ & \left[\Pr\left(a_j = 0, a_k = 1 \mid x_i = x_i^*\right) - \Pr\left(a_j = 1, a_k = 0 \mid x_i = x_i^*\right)\right] \cdot \left(\delta_{ij} - \delta_{ik}\right). \end{aligned} \tag{4}$$

Stage game equilibrium: In the static game, an equilibrium is a vector of cutoffs $\left(x_i^*, x_j^*, x_k^*\right)$ such that each player is indifferent between $a_i = 0$ and $a_i = 1$ given the expected play of the other players. With the information structure described above, x_i^* is therefore defined by:

$$\begin{aligned} 0 = x_i^* + & \left\{\Phi_2\left[-\alpha_j(x_i^*), -\alpha_k(x_i^*), \omega_{jk}\right] - \Phi_2\left[\alpha_j(x_i^*), \alpha_k(x_i^*), \omega_{jk}\right]\right\} \cdot \delta_i \\ + & \left\{\Phi_2\left[-\alpha_j(x_i^*), \alpha_k(x_i^*), -\omega_{jk}\right] - \Phi_2\left[\alpha_j(x_i^*), -\alpha_k(x_i^*), -\omega_{jk}\right]\right\} \cdot (\delta_{ij} - \delta_{ik}), \end{aligned} \tag{5}$$

where Φ_2 is the *cdf* of the bivariate normal, $\alpha_j(x_i^*) = \frac{x_j^* - \mu_j - \rho_{ij}(x_i^* - \mu_i)}{\sqrt{1-\rho_{ij}^2}}$, $\alpha_k(x_i^*) = \frac{x_k^* - \mu_k - \rho_{ik}(x_i^* - \mu_i)}{\sqrt{1-\rho_{ik}^2}}$ and $\omega_{jk} = \frac{\rho_{jk} - \rho_{ij} \cdot \rho_{ik}}{\sqrt{\left(1-\rho_{ij}^2\right) \cdot \left(1-\rho_{ik}^2\right)}}$. x_j^* and x_k^* are defined analogously.

We can now characterise the circumstances in which conditional state monotonicity and equilibrium uniqueness arise, in Propositions 1 and 2 respectively.

Proposition 1 (Conditional state monotonicity) *Holding other judges' cutoff strategies fixed, for committee member i, the difference in utility between $a_i = 1$ and $a_i = 0$ is monotonically increasing in x_i, and therefore each judge has a unique cutoff x_i^*.*

Proof: Proofs are in the appendix.

It is worth noting that, while it is sufficient for Proposition 1 that no judges have a preference for 'discoordination' with another judge ($\delta_i \geq \delta_{ij}, \delta_{ik}$), it is not necessary; further details are available in the proof.

[27] Focusing attention on cutoff strategies is common in this literature (see, for example, Morris and Shin (2003)). Iterated elimination of dominated strategies implies that where a unique equilibrium exists in these games, it will be monotone.

Proposition 2 (Sufficient condition for unique equilibrium) *If, for each judge i,*

$$\delta_i < \sqrt{\frac{\pi}{2(1-\omega_{jk}^2)}} \cdot \left(\sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}} + \sqrt{\frac{1-\rho_{ik}}{1+\rho_{ik}}} \right)^{-1} \quad (6)$$

then there is a unique strategy profile that survives iterated deletion of strictly dominated strategies.

Proposition 2 is an extension to three players of the equilibrium uniqueness results for coordination games in Morris and Shin (2006), with one main modification. In our formulation, we fix the noise of agents' signals to unity and allow the returns to coordination to vary, while Morris and Shin (2006) fix the returns to coordination and allow the noise of agents' signals to vary. Put in this way, Proposition 2 generates an insight into uniqueness conditions for co-ordination games of this type that is not emphasized in the literature: *it is sufficient for uniqueness that players do not care too much about coordination*, relative to the correlation of their signals.[28]

Status quo bias emerges naturally as an equilibrium property of this stage game:

Proposition 3 (Status quo bias) *If $\mu_m - \rho_{mn}\mu_n > 0 \ \forall \ m, n \in i, j, k$, then $x_i^* \leq 0 \ \forall \ i$, with strict inequality if $\delta_i > 0$.*

Consider a 'no-coordination' benchmark by setting each of the δ terms equal to zero. Then each judge selects $x_i^* = 0$, and votes for the favorite if and only if (s)he believes that the favorite genuinely deserved to win. However, provided that all three judges broadly agree on who the expected winner of the debate is and by how much (as described by the condition in Proposition 3) each judge also knows that at $x_i = 0$, each of their colleagues is more likely than not to vote for the favorite. If this is true, then any increase in the preference for coordination will generate status quo bias - at $x_i = 0$, judge i's expected utility from voting for the favorite will be greater than from voting for the underdog, and so i will shift her cutoff slightly to the left; that is, $\delta_i > 0$ implies $x_i^* < 0$. Judge i will have a range of signals ($x_i \in [x_i^*, 0)$) where

[28] It is also worth noting that this proof has bounds that are tighter than the equivalent restrictions for two players, even when i's signal is completely uninformative of some other player's signal (*i.e.* $\rho_{ij} = 0$ or $\rho_{ik} = 0$). This is because the proof of uniqueness relies on translations of monotone strategies, and the same shift in i's signal with three players generates a greater potential change in payoffs than it would if i had only one other player with whom to coordinate.

(s)he would privately prefer to vote for the underdog, but instead votes for the favorite out of a desire to increase the probability of coordinating with the other judges.

This result is important for two reasons. First, it shows the existence of the status quo bias effect we have described earlier in the paper. Second, it describes the circumstances in which it will occur, and in particular shows that it is not in general sufficient merely that all judges agree on who the favorite is ($\mu_i, \mu_j, \mu_k > 0$). Instead, they must also have a requisite level of agreement of the degree of difference between the two candidate teams. If they do not, then one of the judges can have a signal distribution with a mean so far away from the other judges — and, therefore, a much higher *ex ante* probability of voting for the favorite — that increasing the rewards to coordination leads that judge to become *less* likely to vote for the favorite, in the hopes of coordinating with her more moderate colleagues. Note, however, that even in that case incentives to coordinate still generate statistical discrimination in the other judges, with the force even greater than that would be in the absence of the extremist judge.

In this sense, Proposition 3 enables us to state the circumstances in which an increase in preferences for coordination leads to a status quo bias, and the circumstances in which it generates a moderating effect (for at least one of the judges, but not for all three). We turn now to identification and estimation of this game, but return to the implications of this result in the conclusion. In particular, a theoretical extension of this model would be able to generate insights into 'optimal committee design' and the circumstances in which conformity and incentives to coordinate improve or weaken committee decision-making processes.

3.2 Identification of the stage game

In this section, we show identification of the stage game. That is, we imagine (for now) a large cross-section of observations of the stage game.[29] This will be essential for showing identification in the dynamic context, in which we allow players' preferences for coordination to evolve optimally.

Assumption 3 (Parameterization of the model for estimation) *To take the model to data, we assume common values for ρ_{12}, ρ_{13} and ρ_{23} across all committees. For each committee*

[29] Alternatively, imagine a stationary repeated stage game, in which players' preferences for coordination do not change as the tournament proceeds.

c, we denote the difference in pre-tournament rankings by $R^c > 0$. We allow this ranking difference to shift judges' signal means; for simplicity, we adopt a linear specification, and allow a different relationship for each judge class:[30] $\mu_1^c = \beta_1 \cdot R^c$; $\mu_2^c = \beta_2 \cdot R^c$; $\mu_3 = \beta_3 \cdot R^c$.

Further, denote D_i^c to represent some measure of the dissent history for judge i on committee c. (For example, D_i^c could represent the total stock of dissents, or could be a dummy for whether the judge has dissented in the previous tournament round.) We allow this variable to shift each judge's preference for agreement; for the purposes of identification, we require only that, for some committees, we observe $D_1^c = D_2^c = D_3^c = 0$. We allow different classes of judges to be differentially affected by previous dissent, and we impose that each judge is indifferent between agreeing with one peer and agreeing with two: $\delta_1^c = \delta_{12}^c = \delta_{13}^c = \delta_1 \cdot D_1^c$; $\delta_2^c = \delta_{21}^c = \delta_{23}^c = \delta_2 \cdot D_2^c$; $\delta_3^c = \delta_{31}^c = \delta_{32}^c = \delta_3 \cdot D_3^c$.

This reflects the central hypothesis being tested: that past dissent increases a judge's preference for coordination. The equations also show two important limitations of this estimation method. First, the current experimental context does not allow us to identify δ_i, δ_{ij} and δ_{ik} separately; this is because the exogenous variation (past dissent) operates at the level of the individual judge.[31] Second, we use the structural model to identify the preference for coordination driven by past dissent, but we do not seek to identify the preference for coordination generally. That is, if $D_i^c = 0$, we impose $\delta_i^c = 0$.

We constrain the model so that $\rho_{12}, \rho_{13}, \rho_{23} \in [0, 0.99]$, and so that the covariance matrix is positive definite. We impose $\delta_1, \delta_2, \delta_3 \geq 0$. Together, these constraints ensure conditional state monotonicity (Proposition 1). Additionally, we assume that each stage game has a unique equilibrium; we check this assumption in due course.[32]

30 These specifications imply that, in the hypothetical case that two teams were equally matched ($R^c = 0$), then $\mu_1^c = \mu_2^c = \mu_3^c = 0$. This is exactly as we would expect and require.

31 We *could* use the present structural methodology to separately identify δ_i, δ_{ij} and δ_{ik}, if we observed some exogenous shock operating at the level of the *relationship* between judges *i* and *j*. This might be more realistic in an applied IO context; if this were a coordination game between three firms, for example, one might imagine an exogenous regulatory shock changing the coordination incentive between just two of those firms. If such variation were observed, our global identification result could readily be extended to identify $(\delta_i, \delta_{ij}, \delta_{ik})$ generally.

32 Clearly, this restriction could be relaxed for a different empirical context; for example, by using an equilibrium selection rule, or by allowing equilibria to co-exist according to a discrete finite mixture distribution. See de Paula and Tang (2012) and Su (2014).

Proposition 4 (Identification of the stage game) *The stage game is identified.*

The intuition for identification is straightforward. The covariance parameters (ρ_{12}, ρ_{13} and ρ_{23}) and the 'slope' parameters (β_1, β_2 and β_3) are identified from those committees in which no committee member has a preference for coordination; this is simply a trivariate probit, whose properties (and identification) are well understood. To identify the coordination preference of class 1 judges (*i.e.* δ_1), we can compare these no-coordination committees to committees in which $D_1^c > 0$ and $D_2^c = D_3^c = 0$. We can think intuitively of these committees as revealing a different value of x_1^{c*} for each different value of D_1. Holding fixed D_2 and D_3, we have a one-to-one mapping between δ_1 and x_1^{c*} (see Proposition 3), so δ_1 is identified. We can identify the coordination preference of class 2 and class 3 judges by repeating the logic. Of course, in practice, we estimate all parameters jointly — but this thought experiment captures intuitively the way that exogenous variation in coordination preferences allows identification of the structural parameters.

3.3 Pre-registration of stage game predictions in the cross-section

This stage game lies at the centre of our structural model; it provides the central building block for the Markov Perfect Equilibrium that follows. Having formulated the stage game and showed identification — and while in the process of developing the dynamic implementation of the model — we used the stage game to make testable predictions of voting behavior in the 2013 WSDC championships. We registered these predictions with the J-PAL Hypothesis Registry.[33] Such registration is increasingly important for randomized controlled trials, where registration is seen as a valuable method for committing to specifications and hypotheses before experiments are run (see, for example, Casey, Glennerster, and Miguel (2012) and Olken (2015)). Pre-registration also provides a novel method of assessing goodness-of-fit for structural models of field experiments; for example, DellaVigna, List, Malmendier, and Rao (2016) registered a structural model along with an experimental design for their study of gift exchange in the workplace.[34] We return to our pre-registered predictions shortly, when assessing the model's goodness-of-fit.

[33] We registered this in January 2013.

[34] DellaVigna, List, Malmendier, and Rao (2016) registered this model in November 2014.

3.4 Solving a full dynamic model

We now nest the stage game in a dynamic structure, by allowing judge i to choose his or her value for δ_i before playing each stage game.[35] Denote each tournament round by t, with the total number of rounds being $T = 8$. For simplicity, we now denote x_{it} as the signal received by judge i in round t, and denote a_{it} as the action of judge i in round t (denoted a_i in the stage game). Denote s_{it} as the total stock of dissents that judge i has recorded *after* round t. We assume that, over the course of a tournament, a judge receives a 'super-utility' ($\tilde{U}$), that combines (i) the utility a judge gains from expressing his or her preferences across multiple rounds, and (ii) the disutility from the total number of dissents that a judge has recorded by the end of a tournament: $\widetilde{U} = \left(\sum_{t=1}^{T} x_{it} \cdot a_{it}\right) - \kappa(s_{iT})$.

The appropriate solution concept is Markov Perfect Equilibrium — in which each player chooses a_{it} as a function of (i) his or her own stock of previous dissents ($s_{i,t-1}$) and (ii) the distribution of all other players' stocks of previous dissents (which we denote using the vector $\boldsymbol{s}_{-i,t-1}$). Markov Perfect Equilibrium has proved a very useful framework for a large number of applications in empirical industrial organisation (Maskin and Tirole, 1988; Doraszelski and Pakes, 2007; Aguirregabiria and Mira, 2007), as well as for the analysis of strategic voting in political economy (Acemoglu and Robinson, 2008; Burgess, Jedwab, Miguel, Morjari, and Padró i Miquel, 2015). It is appropriate for our context given the assumption that each player ultimately cares about his or her dissent history only through the terminal dissent stock, s_{iT}.[36] Player i in period t therefore faces the following problem:

$$\max_{a_{it} \in \{0,1\}} \mathbb{E}\left[\widetilde{U}(x_{i1}, \ldots, x_{iT}, a_{i1}, \ldots, a_{iT}, s_{iT}) \;\middle|\; s_{i,t-1}, \boldsymbol{s}_{-i,t-1}\right]. \tag{7}$$

This requires solving every game on the equilibrium path — including a large number of games that are never actually played. Specifically, we assume that (i) each judge acts after knowing his or her committee peers for period t, (ii) assuming that peer assignment is random in future periods, and (iii) treating the pool of potential

[35] In the stage game, we used i to index judges within a given committee: $i \in \{1, 2, 3\}$. For simplicity, we now make a minor change in notation: for the dynamic model, we use i to index judges across the dataset.

[36] Note that the dissent stock can never decrease over time: this implies that our structure can also be considered a 'dynamic directional game' (Iskhakov, Rust, and Schjerning, 2015). However, given our result on multiple equilibria (discussed in section 5), we do not need to exploit a recursive lexicographical search method in this application.

peers as 'large'.[37]

We can therefore solve using backward induction.[38] Let V_t be the expected utility from all future play, entering round t. Notice that (without loss of generality) we can treat the disutility from dissent as accruing at the conclusion of the final tournament round; for notational convenience, we denote this as $V_{T+1}(s_{iT}) = -\kappa(s_{iT})$. To begin the backward induction process, consider the final round, T. Specifically, consider the point after which judge i has been randomly assigned to a committee with peers j and k, observing ranking difference r and judge i's signal, x_{iT}. Then the value of voting for the favorite is: $x_{iT} + \left[1 - \Pr\left(a_j = a_k = 0 \,|\, x_{iT}\right)\right] \cdot V_{T+1}(s_{i,t-1}) + \Pr\left(a_j = a_k = 0 \,|\, x_{iT}\right) \cdot V_{T+1}(s_{i,t-1}+1)$. The value of voting against the favorite is: $\left[1 - \Pr\left(a_j = a_k = 1 \,|\, x_{iT}\right)\right] \cdot V_{T+1}(s_{i,t-1}) + \Pr\left(a_j = a_k = 1 \,|\, x_{iT}\right) \cdot V_{T+1}(s_{i,t-1}+1)$. Indifference between the two pins down x^*_{iT}. Trivially, note that the solution is isomorphic to the solution of the stage game (equation 5), where $\delta_i = \delta_{ij} = \delta_{ik} = V_{T+1}(s_{i,t-1}) - V_{T+1}(s_{i,t-1}+1)$. We can calculate the value of entering a committee by integrating over the possible values of the signal x_{iT}.[39]

This expected utility is then conditional on the ranking difference, and on the dissent stock of the judge's peers. We can therefore sum over these variables, using the empirical distributions observed in the data.[40] Denote $x^*_{1t}(r, s_i, s_2, s_3)$ as the cutoff for judge i of class 1, having dissent stock s_i on a committee with peers having dissent stock s_2 (class 2) and s_3 (class 3), where the ranking difference is r. For ease of notation, denote by $x^*_{2t}(r, s_i, s_2, s_3)$ and $x^*_{3t}(r, s_i, s_2, s_3)$ respectively the cutoffs for the class 2 and class 3 judges in the same debate; denote by R_t the set of all ranking differences in round t, and by S_{2t} and S_{3t} respectively the sets of all dissent stocks in round t for judges of classes 2 and 3. Generically, we can then write the following Bellman equation (where ϕ denotes the *pdf* of the normal distribution, Φ_3 denotes

[37] That is, we treat as equivalent the concepts of 'the distribution of all players' stocks of previous dissents' and 'the distribution of all *other* players' stocks of previous dissents'. Given the size of our tournaments, there is almost no difference between these distributions. However, there is a fundamental difference in computational burden: the former concept leads to a computationally tractable log-likelihood, whereas the latter does not.

[38] This is in contrast to many dynamic discrete games, where the horizon is infinite and researchers assume a stationary Markov Perfect Equilibrium. See, for example, Aguirregabiria and Mira (2007).

[39] This integration is trivial. Note that, for any variable $X \sim \mathcal{N}(\mu, 1)$, $\mathbb{E}(X \,|\, X > z) = \mu + \phi(z-\mu) \cdot [1 - \Phi(z-\mu)]^{-1}$. Imagine, then, a judge who will use cutoff x^*_{iT} when facing a signal $x_{iT} \sim \mathcal{N}(\mu_{it}, 1)$; it follows that $\mathbb{E}\left(x_{iT} \,|\, x_{iT} > x^*_{iT}\right) \cdot \Pr(x_{iT} > x^*_{iT}) = \phi\left[x^*_{iT} - \mu_{iT}\right] + \left\{1 - \Phi\left[x^*_{iT} - \mu_{iT}\right]\right\} \cdot \mu_{it}$.

[40] That is, we imagine each committee member as summing over the possible distribution of the ranking difference and the dissent stock of his or her potential peers; we use the observed empirical distribution of these variables as a non-parametric density estimator.

the *cdf* of the trivariate normal, and $\mathbf{\Omega}$ denotes the signal covariance matrix):

$$
\begin{aligned}
V_t(s_{i,t-1}) = V_{t+1}(s_{i,t-1}) + \sum_{r \in R_t} \Pr(R = r) \cdot \sum_{s_2 \in S_{2t}} \Pr(S_2 = s_2) \cdot \sum_{s_3 \in S_{3t}} \Pr(S_3 = s_3) \times \\
\Bigg\{ \phi \left[x^*_{1t}(r, s_{i,t-1}, s_2, s_3) - \beta_1 \cdot r \right] + \left\{ 1 - \Phi \left[x^*_{it}(r, s_{i,t-1}, s_2, s_3) - \beta_1 \cdot r \right] \right\} \cdot \beta_1 \cdot r \\
+ \Bigg(\Phi_3 \left\{ x^*_{it}(r, s_{i,t-1}, s_2, s_3) - \beta_1 \cdot r, - \left[x^*_{2t}(r, s_{i,t-1}, s_2, s_3) - \beta_2 \cdot r \right], - \left[x^*_{3t}(r, s_{i,t-1}, s_2, s_3) - \beta_3 \cdot r \right], \mathbf{\Omega} \right\} \\
+ \Phi_3 \left\{ - \left[x^*_{it}(r, s_{i,t-1}, s_2, s_3) - \beta_1 \cdot r \right], x^*_{2t}(r, s_{i,t-1}, s_2, s_3) - \beta_2 \cdot r, x^*_{3t}(r, s_{i,t-1}, s_2, s_3) - \beta_3 \cdot r, \mathbf{\Omega} \right\} \Bigg) \\
\times \left[V_{t+1}(s_{i,t-1} + 1) - V_{t+1}(s_{i,t-1}) \right] \Bigg\}. \qquad (8)
\end{aligned}
$$

Symmetrically, it is straightforward to modify the equation for judges of classes 2 and 3.

3.5 Identification and estimation of the dynamic model

Given Proposition 4, identification for the dynamic model is trivial. In an infinitely large dataset, we could identify $\kappa(s_{iT})$ non-parametrically simply from observing the stage game in the final round. We would need only to assume that, for some value of the dissent stock s, we have $\kappa(s) = \kappa(s+1)$. This restriction is necessary to ensure that, with some positive probability, we observe some judges having no preference for coordination; recall that this was an important aspect of the proof for identification in the stage game.

However, we have a finite sample — indeed, a sample in which we never observe $s_{iT} > 4$. To estimate, we therefore need to assume a functional form for $\kappa(s_{iT})$. We take a very simple iterative structure, in which $\kappa(s_{iT})$ is zero for a dissent stock of zero or one, and then linear for higher values of s_{iT}.

Assumption 4 (Disutility of the accumulated dissent stock)

$$
\kappa(s_{iT}) = \begin{cases} 0 & \textit{if } s_{iT} = 0 \textit{ or } s_{iT} = 1; \\ \gamma_i + \kappa(s_{iT} - 1) & \textit{otherwise.} \end{cases} \qquad (9)
$$

It is worth emphasising what the assumption $\kappa(0) = \kappa(1) = 0$ means in practice. Recall that κ is the aggregate disutility from dissent (which, without loss of generality, we treat as being imposed at the conclusion of the final round). Therefore, this assumption simply imposes that a judge who enters the *final* round without yet having dissented will vote completely sincerely; *i.e.* without fearing any disutility if she or he dissents for the first time in that final round. (Of the 1809 judge-committee observations in our dataset, this applies to 116 observations; *i.e.* about 6% of the votes we observe.) We view this as a reasonable assumption in this context, and not just because it is important for identification. A judge who has a 'clean sheet' of no dissents after seven separate votes can be assumed to feel very comfortable in his or her performance — to the point that we are willing to impose that the judge is indifferent between whether she or he dissents in that final debate. Crucially, note that this assumption certainly does *not* imply that any judge in any earlier round is spared the fear of dissenting. Because Markov Perfect Equilibrium requires rational updating over future outcomes, *all* judges entering *all* preceding rounds still face a risk of concluding round 8 with more than one dissent in total. As our subsequent estimates show, this implies that all votes in all rounds — aside from the 6% of votes that are cast by judges in round 8 with no previous dissents — are affected to some extent by social pressure.

Estimation is then straightforward, if computationally demanding. We use Maximum Likelihood with a nested optimization algorithm. Denote the stacked parameter vector as $\boldsymbol{\theta} = (\beta_1, \beta_2, \beta_3, \rho_{12}, \rho_{13}, \rho_{23}, \gamma_1, \gamma_2, \gamma_3)'$. Then, for some candidate value for $\boldsymbol{\theta}$, the inner loop solves the dynamic game by backward induction. We then calculate the log-likelihood $\ell(\boldsymbol{\theta})$ using a standard triprobit structure (where we approximate the *cdf* of the trivariate normal using the method of Genz (2004)). The outer loop updates using a Sequential Quadratic Program. We calculate p-values for our parameter estimates using Likelihood Ratio tests.[41] For the present empirical

[41] Note that several important hypotheses lie on the boundary of the admissible parameter set. ρ_{12}, ρ_{13}, ρ_{23}, γ_1, γ_2 and γ_3 are each constrained to be non-negative; for each, we test a null hypothesis that the parameter is zero by using a Likelihood Ratio statistic compared to a $\bar{\chi}^2(0, 1)$ distribution. Further, we test a joint hypothesis that all covariance terms are zero; we use the union-intersection principle to test against the alternative hypothesis that all three covariance terms are positive: Andrews (1999); Silvapulle and Sen (2011). Parameters β_1, β_2 and β_3 are unconstrained; we therefore use Likelihood Ratio tests (each compared to the $\chi^2(1)$ distribution). We also test $H_0 : \rho_{12} = \rho_{13} = \rho_{23}$; this is a standard Likelihood Ratio test, in which we compare to a $\chi^2(2)$.

application, a single evaluation of the log-likelihood requires solving 8590 games.[42] We discuss in more detail the process of solving and estimating the model in Online Appendix C.

4 Results

4.1 Structural results

Table 2 reports our results. Consider each set of parameters in turn. First, we estimate significant variation in signal mean, driven by variation in the rankings differences: $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are all positive.[43] Second, we estimate large and highly significant correlation in signals. For each test of H_0: $\rho_{12} = 0$, H_0: $\rho_{13} = 0$ and H_0: $\rho_{23} = 0$, we find an LR statistic over 50.[44] Further, we find that our signal correlation *cannot* be reduced to a single 'public signal': a common signal of this kind would imply $\rho_{12} = \rho_{13} = \rho_{23}$, which we strongly reject ($LR = 12.461$). Together, these two tests provide strong empirical support for one of the key methodological innovations in this paper: the modelling of correlated signals using an unrestricted covariance matrix. We discuss this in more detail shortly.

< Table 2 here. >

Most importantly, we estimate a large and significant preference for coordination: particularly among judges of class 1 ($\hat{\gamma}_1 = 1.106$) and of class 3 ($\hat{\gamma}_3 = 0.547$). (Note, however, that we do not reject a null hypothesis that the three classes share the same coordination preference: when we test $H_0 : \gamma_1 = \gamma_2 = \gamma_3$, we obtain $p = 0.281$.) To understand the magnitude of these estimates, we back out the estimated coordination preference within each committee (*i.e.* δ) as a function of tournament round and dissent stock. Figure 4 illustrates (using data from the 2012 tournament); the

[42] We solve these games using a parallel architecture with 12 threads; this substantially improves computation time. For numerical feasibility, we also discretize the ranking difference: $R \in \{1, 2, 3, 4, 5, 8, 16, 25, 35, 44\}$.

[43] The magnitudes of these three estimates are as we might expect. For example, the estimates imply that a class 1 judge with no preference for coordination has a 70% probability of voting for a favorite who is ranked 20 places ahead of its opponent, and an 85% probability of doing so when the ranking difference is 40.

[44] We therefore strongly reject each hypothesis separately. By the union-intersection principle, we also reject the joint hypothesis H_0: $\rho_{12} = \rho_{13} = \rho_{23} = 0$.

top surface represents class 1 judges, the second surface represents class 3, and the lowest surface, class 2. The figure is useful for understanding the intuition of the dynamic model. First, consider the profile of the functions for tournament round 8: this is the function $\kappa(s_{iT}+1)-\kappa(s_{iT})$, for each class of judge; note the identifying assumption $\kappa(1)-\kappa(0)=0$. Second, note that, as we would expect, the function increases monotonically in dissent stock — but, holding the stock constant, the function decreases monotonically in tournament round.

< Figure 4 here. >

4.2 Goodness-of-fit and the counterfactual

We assess the goodness-of-fit in two main ways. First, we return to the pre-registration of stage game predictions. In Online Appendix A, we show the in-sample goodness-of-fit for our static model in the cross section (that is, the fit for the tournaments held in 2010, 2011 and 2012). We also show the out-of-sample fit for the moments we predicted for the 2013 tournament. We do this by comparing a set of real moments with equivalent moments from a simulation of our model in the cross-section. This static model performed well in matching cross-sectional moments, both in-sample and out-of-sample. Of the 72 moments checked for in-sample prediction, nine lie outside the 90% confidence interval from the simulated data; of the 24 moment predictions made for the 2013 tournament, three lie outside the same interval. It is important to stress that this exercise is hardly a check of goodness-of-fit for the whole model — rather, we show that check in the following section, after developing the dynamic structure — but we nonetheless found these results very reassuring, and a useful check as we moved from static to dynamic implementation.

Second, we map the dynamic evolution of the dissent stock. Figure 3 showed the evolution of the dissent stock for different judges across different tournament rounds. To assess our model's goodness of fit, we compare this figure to the predicted evolution of the dissent stock, which we show in Figure 5.[45] The figure shows that the model fits the data well.

< Figure 5 here. >

[45] We construct this figure by running a large number of simulations of our dynamic model, then rounding each transition moment to its nearest integer.

Figure 6 shows the evolution of the stock of dissents under a counterfactual in which no judge has a preference for coordination (*i.e.* $\gamma_1 = \gamma_2 = \gamma_3 = 0$). Relative to Figure 5, Figure 6 shifts mass to higher values of the dissent stock. In this way, Figure 6 illustrates the central mechanism being studied in this paper: coordination preferences lead committee members to engage in status quo bias, which then reduces dissent.

< Figure 6 here. >

We extend the analysis by showing the counterfactual: the proportion of debates in which, if (s)he had no preference for coordination, a judge would shift from voting for the favorite to voting against. Figure 7 shows the difference between the probability of voting for the favorite with a coordination preference and the probability of voting for the favorite with no coordination preference; we show this by each judge class separately, and then for the overall committee outcome. The figure highlights an important implication of the model: though the preference for coordination is relatively large (as we quantify in the next section), the practical effect of this preference on committee outcomes is relatively limited. This is because the coordination mechanism — a shift of cutoffs in a context where signals are quite noisy — still provides substantial uncertainty as to the committee outcome. For this reason, Figure 7 shows that the largest distortion through coordination preferences is felt in debates between opponents that are not closely matched (namely, debates whose discretized ranking differences are 16, 25 and 35). For decisions that are relatively evenly matched — for example, with a ranking difference of eight or less — we find that the preference for coordination increases the favorite's prospects by less than 10 percentage points.

< Figure 7 here. >

4.3 Equivalent variation: Introducing a 'Faustian option'

What do the estimated coordination preferences mean in terms of trade-offs? To answer this question, we construct a new measure of equivalent variation. Typically, equivalent variation measures the amount of money that a consumer would pay to avoid an increase in price. In this context, we have no prices: the relevant change is an increase in the dissent stock. Similarly, we have no money: in an expressive voting model, the relevant 'currency' is the ability to express an opinion by voting. We therefore construct a new measure of equivalent variation, appropriate to the context of committee voting: *the proportion of decisions that a committee member would give up to avoid dissent*.

To do this, we introduce a hypothetical 'Faustian option': in which each judge can 'sell' his or her opportunity to express an opinion, in exchange for avoiding the possibility of disagreeing with his or her peers. Specifically, in addition to the possibility of voting for or against the favorite, we imagine that each judge can refuse to make a decision, and thus avoid both (i) the utility of expressing an opinion and (ii) the risk of dissenting. In the stage game, the expected utility of voting for the favorite is $x_i + \delta_i \cdot \left[1 - \Pr\left(a_j = 0, a_k = 0 \mid x_i\right)\right]$, and that the expected utility of voting against the favorite is $\delta_i \cdot \left[1 - \Pr\left(a_j = 1, a_k = 1 \mid x_i\right)\right]$. Then the utility from a hypothetical Faustian option is simply $0.5x_i + \delta_i$: under such an option, the judge 'splits the difference', and is guaranteed the payoff to coordination.[46]

Under this Faustian option, judge i now votes for the favorite if $x_i \geq \hat{x}_i^+$, where $\hat{x}_i^+ \equiv 2\delta_i \cdot \Pr\left(a_j = 0, a_k = 0 \mid x_i = \hat{x}_i^+\right)$. Symmetrically, judge i votes against the favorite if $x_i \leq \hat{x}_i^-$, where $\hat{x}_i^- = -2\delta_i \cdot \Pr\left(a_j = 1, a_k = 1 \mid x_i = \hat{x}_i^-\right)$. For $x_i \in \left(\hat{x}_i^-, \hat{x}_i^+\right)$, the judge exercises the Faustian option. It is straightforward to solve $\hat{x}_i^+$ and $\hat{x}_i^-$ numerically for each judge on each committee: in our dynamic model, $\delta_i \equiv V_{T+1}(s_i) - V_{T+1}(s_i + 1)$. We then measure the equivalent variation by the mass $\Phi\left(\hat{x}_i^+ - \mu\right) - \Phi\left(\hat{x}_i^- - \mu\right)$: for each judge on each committee, this mass measures the proportion of decisions that the judge would abandon in order to avoid dissent.[47] Note that the equivalent variation increases with δ, and is zero for $\delta_i = 0$.

Figure 8 shows the equivalent variation for each observed game, as a function of judge class and ranking difference. The figure shows starkly the magnitude of the estimated preferences for coordination: in many committees, judges in class 1 and class 3 would be willing to forgo between approximately 20% and 50% of their decisions to avoid the risk of dissent.

[46] It is important that the difference be halved — rather than split by some other proportion — in order to retain the symmetry of voting for the favorite and against.

[47] Note, trivially, that $\hat{x}_i^+$ and $\hat{x}_i^-$ can always be solved uniquely for a given judge on a given committee, and that $\hat{x}_i^+ \geq \hat{x}_i^-$.

< Figure 8 here. >

4.4 What do we learn from allowing dynamic updating?

In this paper, we have showed how a model of strategic committee voting may be nested in a tractable dynamic framework — rather than having to assume a repeated single-shot framework. This substantially increases the computational burden of the estimation; does it also change the conclusions of the model?

To answer this question, we compare our estimates to results from a 'static' version of the model — that is, a version in which we estimate the stage game in a pooled cross-section, without allowing for a time dimension.[48] Table 3 shows the results. Two key conclusions emerge. First, the dynamic model fits the data better, but not significantly better. The dynamic model generates a log-likelihood of -848.872, rather than the log-likelihood of -851.339 for the static model; however, when we run a Vuong (1989) test for comparing non-nested models, we obtain $p = 0.479$. Second, by imposing more structure on the problem, the dynamic model nonetheless provides much tighter estimates. Specifically, in contrast to the dynamic model, the pooled static model reveals no preference for coordination among either class 1 or class 2 judges — though the static model would not reject that such coordination preferences are large for all three judge classes.[49]

In sum, it is theoretically preferable to allow for full dynamic updating — in the sense that a static model implicitly assumes that all judges are completely myopic — and the dynamic model provides both a better overall fit and tighter estimates on the preferences for coordination. Of course, our results on this point are hardly conclusive — but illustrate nonetheless that dissent aversion need not be considered solely as a 'knee-jerk' response to a previous decision, nor as a phenomenon driven purely by institutional design. Rather, our results suggest that dissent aversion can vary over time within the same set of institutions, based on agents trading off the merits of voting for their preferred outcome and the risk of dissent.

< Table 3 here. >

[48] We use the parameterization originally set out in section 3.2.

[49] We can check this by testing in the static model the null hypothesis that the static coordination parameters are of the same magnitude as the coordination parameters in the dynamic model for a judge having dissented once; that is, we test the joint hypothesis $H_0 : \delta_1 = 1.106; \delta_2 = 0.189; \delta_3 = 0.549$. We obtain $\ell_r = -853.270$, implying $LR = 3.862$, and $p = 0.261$.

4.5 What do we learn from allowing correlated signals?

Our main estimates show significant correlation in signals. Such correlation is clearly an important aspect of the strategic environment, and one that should be modeled explicitly. But what would it matter if we assumed conditional signal independence? What conclusions would we draw about committee members' preferences?

To explore this, we estimate the dynamic model imposing conditional signal independence. Table 4 shows the results — which are strongly supportive of our decision to allow for correlated signals. Imposing conditional signal independence not only substantially worsens the model fit — it also clearly misleads on the magnitude and significance of coordination preferences. Specifically, we estimate a similar magnitude of coordination preference for judges in class 2 and class 3 — but we substantially underestimate the coordination preference for class 1 judges (indeed, we now cannot reject a null hypothesis that class 1 judges have no coordination preference). Conditional signal independence is clearly an inappropriate assumption in this context — and, we posit, may be just as problematic as an assumption in a wide variety of other empirical applications.

< Table 4 here. >

5 Robustness

5.1 Robustness to multiple equilibria

In section 3, we assumed that each stage game has a unique equilibrium — and, therefore, that there is a unique Markov Perfect Equilibrium for each tournament.[50] We now test that assumption, by checking for multiple equilibria in every game on the equilibrium path (including the 603 games observed).

We do this for each game by exploiting numerically our earlier theoretical restriction that equilibria must survive the iterated elimination of dominated strategies. Consider any given stage game. Start by supposing that each player believes with certainly that his or her two colleagues will vote for the favorite. In that case, player i should vote for the favorite if and only if $x_i \geq -\delta_i$. Denote $\underline{x}_i^n$ as the lower bound

[50] This assumption could be relaxed in other applications of our framework (see, for example, Iskhakov, Rust, and Schjerning (2015)).

cutoff for player i in iteration n of this algorithm; thus, the previous supposition is equivalent to $\underline{x}_i^1 = -\delta_i$ (and, symmetrically, $\underline{x}_j^1 = -\delta_j$ and $\underline{x}_k^1 = -\delta_k$).[51] By the logic of iterated elimination of dominated strategies, we can tighten the lower bound by imposing that, in iteration $n > 1$, each player best-responds to colleagues' lower bound cutoffs established in iteration $n - 1$:

$$0 = \underline{x}_i^n + \left\{ \Phi_2 \left[-\alpha_j(\underline{x}_i^n; \underline{x}_j^{n-1}), -\alpha_k(\underline{x}_i^n; \underline{x}_k^{n-1}), \omega_{jk} \right] - \Phi_2 \left[\alpha_j(\underline{x}_i^n; \underline{x}_j^{n-1}), \alpha_k(\underline{x}_i^n; \underline{x}_k^{n-1}), \omega_{jk} \right] \right\} \cdot \delta_i. \tag{10}$$

Symmetric reasoning allows us to generate upper bounds, iterating from $(\overline{x}_i^1, \overline{x}_j^1, \overline{x}_k^1) = (\delta_i, \delta_j, \delta_k)$. This algorithm provides a tool for checking our assumption of equilibrium uniqueness; for a given stage game, it is a sufficient condition for uniqueness that $\lim_{n\to\infty} \overline{x}_i^n = \lim_{n\to\infty} \underline{x}_i^n$, $\lim_{n\to\infty} \overline{x}_j^n = \lim_{n\to\infty} \underline{x}_j^n$ and $\lim_{n\to\infty} \overline{x}_k^n = \lim_{n\to\infty} \underline{x}_k^n$. This condition can be exhaustively checked through numerical brute force.[52]

We implement this algorithm for the 8590 stage games that form the equilibrium path at our maximum likelihood parameter estimates. We find that the unique-equilibrium condition holds in every single one of these games; since our algorithm provides a sufficient condition for uniqueness, we conclude that it was reasonable, in this context, to estimate a unique Markov Perfect Equilibrium.[53]

5.2 Robustness to latent signal heterogeneity

To this point, our analysis has assumed that, conditional on dissent stock and ranking difference, each judge's signals are independent across tournament rounds. However, one might be concerned that this assumption ignores important latent signal heterogeneity. It may be, for example, that different committee members have different underlying tendencies to vote for or against the favorite; some judges might be generally sympathetic to favorites, whereas others may be inherently more contrarian. If that were the case, our model would be misspecified — because other

[51] We can think of iteration 0 as representing completely uninformative bounds; *i.e.* a lower bound cutoff at $-\infty$.

[52] We terminate the iteration when $\overline{x}_i^n$ is sufficiently close to $\overline{x}_i^{n-1}$ and $\underline{x}_i^n$ is sufficiently close to $\underline{x}_i^{n-1}$.

[53] For each game, we also test whether the sufficient condition in Proposition 2 holds. Of the 8590 games on the equilibrium path, 7097 (82.6%) violate Proposition 2; for the 603 games observed in the dataset, the proportion is 66.8% (400 games). This is also useful for understanding, in practical terms, the role of the condition in Proposition 2: the proposition is useful for showing that multiple equilibria are driven by large preferences for coordination — but, as a basis for empirical measurement of multiple equilibria, the condition is not directly relevant.

committee members could coordinate by focusing on the individual identity of their peers, rather than by engaging in status quo bias.

This assumption is testable through a random-effects framework. Suppose that, instead of the signal x_{it} (as specified earlier), each committee member actually receives a signal $x_{it} + \xi_i$, where $\xi_i \sim \mathcal{N}(0, \psi^2)$. We do not seek to build a full model under this more demanding error structure. However, it is straightforward to test $H_0 : \psi = 0$ — the restriction implied by our dynamic structural model. To do this, we estimate three random-effects probits — one for each class of judge — in which we treat the cutoffs from the dynamic structural model as fixed and test $H_0 : \psi = 0$ using a $\overline{\chi}^2(0,1)$ statistic.[54] In effect, this considers the 'generalized residuals' from the dynamic structural model, testing whether there remains any latent judge-level persistence. We report p-values from the $\overline{\chi}^2$ statistic, along with the estimated proportion of variation attributable to judge-level heterogeneity: $\hat{\psi}^2 \cdot (1 + \hat{\psi}^2)^{-1}$.

Table 5 shows the results. The null hypothesis passes at the 10% significance level for each judge class. Moreover, all of the estimates are small. We estimate latent heterogeneity being strongest for class 3 judges; even then, the proportion of signal variation explained at the individual judge level is only about 6%. We conclude that our dynamic estimates are robust to concerns of latent signal heterogeneity — both as a formal statistical test and as a practical matter of magnitude.

< Table 5 here. >

5.3 Robustness to alternative mechanisms: Learning, concentrating and tuning

This paper has focused on one possible mechanism by which dissent may lead to status quo bias. There are three primary alternative hypotheses, which we term 'learning', 'concentrating' and 'tuning'. To incorporate any of these mechanisms into our dynamic structural model would be infeasible, given both theoretical and computational limitations. However, we nonetheless run basic model specification tests, to show that our structural approach is robust to each alternative hypothesis.

[54] This statistic is appropriate because $H_0 : \psi = 0$ lies at the boundary of the admissible parameter set for ψ.

First, we consider the hypothesis that past dissent leads committee members to learn about the quality of their signal, and thereby somehow update their signal in the direction of voting for the favorite. To test this, we exploit the data on marking distance between judges; we posit that, if judges are learning about the quality of their signal, they should respond to the total distance between their marks and their peers' marks, rather than responding to the fact of having dissented. We therefore estimate a simple trivariate probit specification, in which we allow the latent index for each judge to comprise (i) an intercept term, (ii) the ranking difference between the teams (R^c), and (iii) the lagged mark difference between each judge and his or her 'senior' peer.[55] We use a Vuong (1989) test to compare between this specification and our structural model. We find that the structural model fits the data better than the reduced-form alternative (after adjusting for the different number of parameters), though not significantly so ($p = 0.77$).

Second, we test the hypothesis that past dissent leads committee members to concentrate more closely on the debate, to deliver a more precise signal (see Bartoš, Bauer, Chytilová, and Matějka (2016)). To test this, we fix our structural estimates for all parameters except ρ_{12}, ρ_{13} and ρ_{23}, and then allow each of those covariance terms to vary with the dissent stock. Specifically, we allow — for example — that the signal correlation between judge i (a class 1 judge) and judge j (a class 2 judge) to be $\rho_{12,i,j} = \eta^0_{12} + \eta^i_{12} \cdot s_{it} + \eta^j_{12} \cdot s_{jt}$, where s_{it} is the dissent stock of judge i in period t. Symmetrically, we use the equivalent specification for ρ_{13} and ρ_{23}. We then test $H_0 : \eta^1_{12} = \eta^2_{12} = \eta^1_{13} = \eta^3_{13} = \eta^2_{23} = \eta^3_{23} = 0$. If concentrating is an important mechanism, we should reject this null hypothesis: judges with a larger dissent stock should, by following the debate more closely, enjoy a higher signal covariance with their peers. However, this is emphatically not what we find: we obtain $p = 0.932$, implying that this substantial additional flexibility barely improves the model fit.

Third, we test the related hypothesis that, over the course of the tournament, judges simply become more attuned to what constitutes a good debate performance. To test this, we again fix our structural estimates, but allow the covariance terms to vary; for this third test, we allow the covariance terms to vary with tournament round, by specifying $\rho_{12} = \overline{\rho}_{12} + \tau_{12} \cdot t$, where t is the tournament round. Symmetrically, we use the equivalent specification for ρ_{13} and ρ_{23}. We then test $H_0 : \tau_{12} = \tau_{13} = \tau_{23} = 0$. If judges simply become more attuned over time to what constitutes a good debate

[55] That is, we use the distance to the class 1 judge for judges of classes 2 and 3, and the distance to the class 2 judge for the class 1 judge.

performance, we should reject this null hypothesis in favor of $\tau_{12}, \tau_{13}, \tau_{23} > 0$; that is, judges' signals should increase in correlation over the course of the tournament (conditional on having controlled for our dynamic structural model). This is not what we find. Rather, we estimate $\hat{\tau}_{12} = -0.0047$, $\hat{\tau}_{13} = 0.0258$ and $\hat{\tau}_{23} = -0.0194$; that is, the point estimates are relatively small, and two estimates are negative. Most importantly, we obtain $p = 0.335$ on the null hypothesis that the correlation structure does not change over the course of the tournament. We conclude that our estimates are robust to the alternative explanations of 'learning', 'concentrating' and 'tuning'.

6 Conclusion

In this paper, we report results from a large randomized field experiment on strategic voting, developing a structural model to estimate the magnitude of preferences for coordination. We introduced a trivariate normal structure to allow for correlated player signals; this is central to our empirical application, because committee members base their decisions on information that they can observe together, but which cannot be proxied well by any covariate available to us as researchers. We then introduced a Markov Perfect Equilibrium structure, to allow players to update rationally their concerns about dissenting over the course of the debate tournament. Our empirical framework could be applied in a wide variety of empirical contexts in which players make strategic binary decisions in a repeated structure — for example, judges voting repeatedly on legal cases, firms entering and exiting markets, or players interacting repeatedly in laboratory games.

Our results show a connection between dissent aversion and status quo bias, implying that social pressure and public information can lead to coordination on weaker candidates in contexts of committee voting. This effect is relatively large in what it means to the participants: using our 'Faustian measure' of equivalent variation, we estimate that some committee members would be willing to abandon between approximately 20% and 50% of their decisions to avoid the risk of dissent. However, because signals are quite noisy in this context, the increase in the favorite's probability of victory is generally less than 10 percentage points for most committee decisions. This approach to evaluating committee decision-making could be extended to a wide range of group settings, provided appropriate ways to identify parameters (such as the random allocation of judges in this paper) are in place.

The implication of this paper is the social pressures should be taken into account in the design of committee or group decision-making processes. In some contexts, this has involved a trend to anonymizing candidates from their assessors (see Goldin and Rouse (2000)); the results here suggest that, in certain circumstances, assessors should also be anonymized from each other. We are hopeful that future work will build on these results to identify the optimal level of this trade-off, and help ensure the design of committees that make both accurate and fair decisions.

References

Acemoglu, D., and J. A. Robinson (2008): "Persistence of Power, Elites, and Institutions," *American Economic Review*, 98(1), 267–293.

Aguirregabiria, V., and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1–53.

Andrews, D. W. (1999): "Estimation when a Parameter is on a Boundary," *Econometrica*, 67(6), 1341–1383.

Anwar, S., P. Bayer, and R. Hjalmarsson (2012): "The Impact of Jury Race in Criminal Trials," *The Quarterly Journal of Economics*, 127(2), 1017–1055.

Aradillas-Lopez, A., and E. Tamer (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business and Economic Statistics*, 26(3), 261–283.

Arrow, K. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter, and A. Rees, pp. 3–33.

Asch, S. E. (1951): "Effects off Group Pressure Upon the Modification and Distrotion of Judgments," in *Groups, Leadership, and Men*, ed. by H. S. Guetzkow, pp. 177–190. Carnegie Press.

Ashford, J., and R. Sowden (1970): "Multi-variate Probit Analysis," *Biometrics*, pp. 535–546.

Attanasio, O., C. Meghir, and A. Santiago (2012): "Education Choices in Mexico: Using a Structural model and a Randomized Experiment to Evaluate PROGRESA," *The Review of Economic Studies*, 79(1), 37–66.

Austen-Smith, D., and J. S. Banks (1996): "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *American Political Science Review*, pp. 34–45.

Austen-Smith, D., and T. Feddersen (2005): "Deliberation and Voting Rules," *Social Choice and Strategic Decisions*, pp. 269–316.

Austen-Smith, D., and T. Feddersen (2006): "Deliberation, Preference Uncertainty, and Voting Rules," *American Political Science Review*, 100(02), 209–217.

——— (2009): "Information Aggregation and Communication in Committees," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518), 763–769.

Bagues, M., and B. Esteve-Volart (2010): "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *Review of Economic Studies*, 77(4), 1301–1328.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): "Estimating Static Models of Strategic Interactions," *Journal of Business and Economic Statistics*, 28(4), 469–482.

Bandiera, O., I. Barankay, and I. Rasul (2010): "Social incentives in the workplace," *The Review of Economic Studies*, 77(2), 417–458.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103(6), 2499–2529.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *The American Economic Review*, 106(6), 1437–1475.

Bertrand, M., and S. Mullainathan (2004): "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review*, 94(4), 991–1013.

Black, D. (1958): *The Theory of Committees and Elections*. Cambridge University Press.

Blanes i Vidal, J., and C. Leaver (2013): "Social Interactions and the Content of Legal Opinions," *Journal of Law, Economics, and Organization*, 29(1), 78–114.

Breza, E., S. Kaur, and Y. Shamdasani (2015): "The Morale Effects of Pay Inequality," *Working Paper*.

Burgess, R., R. Jedwab, E. Miguel, A. Morjari, and G. Padró i Miquel (2015): "The Value of Democracy: Evidence from Road Building in Kenya," *The American Economic Review*, 105(6), 1817–1851.

Burns, J., L. Corno, and E. La Ferrara (2016): "Interaction, Stereotypes and Performance: Evidence from South Africa," *Working Paper*.

Bursztyn, L., and R. Jensen (2015): "How Does Peer Pressure Affect Educational Investments?," *The Quarterly Journal of Economics*, 130(3), 1329–1367.

Card, D., S. DellaVigna, and U. Malmendier (2011): "The Role of Theory in Field Experiments," *The Journal of Economic Perspectives*, 25(3), 39.

Casey, K., R. Glennerster, and E. Miguel (2012): "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *The Quarterly Journal of Economics*, 127(4), 1755–1812.

Christensen, C., J. R. Larson, A. Abbott, A. Ardolino, T. Franz, and C. Pfeiffer (2000): "Decision Making of Clinical Teams: Communication Patterns and Diagnostic Error," *Medical Decision Making*, 20(1), 45–50.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

Coate, S., and G. Loury (1993): "Will Affirmative-Action Policies Eliminate Negative Stereotypes?," *The American Economic Review*, 83(5), 1220–1240.

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2014): "Social Comparison and Effort Provision: Evidence from a Field Experiment," *Journal of the European Economic Association*, 12(4), 877–898.

Cross, F. B., and E. H. Tiller (1998): "Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals," *Yale Law Journal*, pp. 2155–2176.

Degan, A., and A. Merlo (2009): "Do Voters Vote Ideologically?," *Journal of Economic Theory*, 144(5), 1868–1894.

DellaVigna, S., B. Lindner, Attila andd Reizer, and J. F. Schmieder (2015): "Reference-Dependent Job Search: Evidence from Hungary," *Working Paper*.

DellaVigna, S., J. A. List, and U. Malmendier (2012): "Testing for Altruism and Social Pressure in Charitable Giving," *The Quarterly Journal of Economics*, 127(1), 1–56.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2015): "Voting to Tell Others," *Working Paper*.

——— (2016): "Estimating Social Preferences and Gift Exchange at Work," *Working Paper*.

Doraszelski, U., and A. Pakes (2007): "A Framework for Applied Dynamic Analysis in IO," *Handbook of industrial organization*, 3, 1887–1966.

Downs, A. (1957): *An Economic Theory of Democracy*. Harper and Row.

Duflo, E., R. Hanna, and S. Ryan (2012): "Incentives Work: Getting Teachers to Come to School," *The American Economic Review*, 102(4), 1241–1278.

Dupas, P. (2014): "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment," *Econometrica*, 82(1), 197–228.

Epstein, L., W. Landes, and R. Posner (2011): "Why (and When) Judges Dissent: A Theoretical and Empirical Analysis," *Journal of Legal Analysis*, 3(1), 101–137.

Feddersen, T., and W. Pesendorfer (1996): "The Swing Voter's Curse," *The American Economic Review*, 86(3), 408–424.

Feddersen, T., and W. Pesendorfer (1997): "Voting Behavior and Information Aggregation in Elections with Private Information," *Econometrica: Journal of the Econometric Society*, pp. 1029–1058.

——— (1998): "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting," *American Political science review*, 92(01), 23–35.

Fischman, J. B. (2011): "Estimating Preferences of Circuit Judges: A Model of Consensus Voting," *Journal of Law and Economics*, 54(4), 781–809.

——— (2013): "Interpreting Circuit Court Voting Patterns: A Social Interactions Framework," *Journal of Law, Economics, and Organization*, p. ews042.

Garicano, L., I. Palacios-Huerta, and C. Prendergast (2005): "Favoritism under social pressure," *Review of Economics and Statistics*, 87(2), 208–216.

Genz, A. (2004): "Numerical Computation of Rectangular Bivariate and Trivariate Normal and *t* Probabilities," *Statistics and Computing*, 14(3), 251–260.

Gerardi, D., and L. Yariv (2007): "Deliberative Voting," *Journal of Economic Theory*, 134(1), 317–338.

Gerber, A. S., D. P. Green, and C. W. Larimer (2008): "Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment," *American Political Science Review*, 102(01), 33–48.

Ginsburg, R. B. (1992): "Speaking in a Judicial Voice," *NYUL Rev.*, 67, 1185.

Goldin, C., and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *The American Economic Review*, 90(4), 715–741.

Grieco, P. L. (2014): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," *The RAND Journal of Economics*, 45(2), 303–340.

Hamlin, A., and C. Jennings (2011): "Expressive Political Behaviour: Foundations, Scope and Implications," *British Journal of Political Science*, 41(03), 645–670.

Hansen, S., and M. McMahon (forthcoming): "First Impressions Matter: Signalling as a Source of Policy Dynamics," *Review of Economic Studies*.

Hansen, S., M. McMahon, and C. V. Rivera (2014): "Preferences or Private Assessments on a Monetary Policy Committee?," *Journal of Monetary Economics*, 67, 16–32.

Harrison, G., and J. List (2004): "Field Experiments," *Journal of Economic Literature*, 42(4), 1009–1055.

Hinich, M. J., and M. C. Munger (1997): *Analytical Politics*. Cambridge University Press.

Huck, S., I. Rasul, and A. Shephard (2015): "Comparing Charitable Fundraising Schemes: Evidence from a Natural Field Experiment and a Structural Model," *American Economic Journal: Economic Policy*, 7(2), 326–69.

Iaryczower, M., X. Shi, and M. Shum (2013): "Words Get in the Way: The Effect of Deliberation in Collective Decision-Making," *Working Paper*.

Iaryczower, M., and M. Shum (2012): "The Value of Information in the Court: Get it Right, Keep it Tight," *The American Economic Review*, 102(1), 202–237.

Iskhakov, F., J. Rust, and B. Schjerning (2015): “Recursive Lexicographical Search: Finding all Markov Perfect Equilibria of Finite State Directional Dynamic Games,” *The Review of Economic Studies*, 83(2), 629–657.

Jung, A. (2011): “An International Comparison of Voting by Committeees,” *European Central Bank Working Paper No 1383*.

Keane, M. (2010): “Structural vs. Atheoretic Approaches to Econometrics,” *Journal of Econometrics*, 156(1), 3–20.

Keynes, J. (1936): “The General Theory of Employment, Interest and Money,” .

Kimhi, A. (1994): “Quasi Maximum Likelihood Estimation of Multivariate Probit Models: Farm Couples’ Labor Participation,” *American Journal of Agricultural Economics*, 76(4), 828–835.

Laibson, D., A. Repetto, and J. Tobacman (2007): “Estimating Discount Functions with Consumption Choices over the Lifecycle,” Discussion paper.

Lavy, V. (2008): “Do Gender Stereotypes Reduce Girls’ or Boys’ Human Capital Outcomes? Evidence from a Natural Experiment,” *Journal of Public Economics*, 92(10), 2083–2105.

Levy, G. (2005): “Careerist Judges and the Appeals Process,” *RAND Journal of Economics*, 36(2), 275–297.

——— (2007): “Decision Making in Committees: Transparency, Reputation, and Voting Rules,” *The American Economic Review*, 97(1), 150–168.

Lindquist, S. A. (2006): “Bureaucratization and Balkanization: The Origins and the Effects of Decision-Making Norms in the Federal Appellate Courts,” *University of Richmond Law Review*, 41, 659.

Liu, N., Q. Vuong, and H. Xu (2014): “Rationalization and Identification of Discrete Games with Correlated Types,” *Working paper*.

Lu, L., Y. C. Yuan, and P. L. McLeod (2012): “Twenty-Five Years of Hidden Profiles in Group Decision Making A Meta-Analysis,” *Personality and Social Psychology Review*, 16(1), 54–75.

Mas, A., and E. Moretti (2009): “Peers at Work,” *The American Economic Review*, 99(1), 112–145.

Maskin, E., and J. Tirole (1988): “A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs,” *Econometrica*, pp. 549–569.

Merlo, A., and A. de Paula (forthcoming): “Identification and Estimation of Preference Distributions When Voters are Ideological,” *Review of Economic Studies*.

Morris, S., and H. Shin (2006): "Heterogeneity and Uniqueness in Interaction Games," in *The Economy as an Evolving Complex System, III: Current Perspectives and Future Directions*, ed. by L. E. Blume, and S. N. Durlauf, pp. 207–242. Oxford University Press.

Morris, S., and H. S. Shin (2003): "Global Games: Theory and Applications," *Econometric Society Monographs*, 35, 56–114.

Olken, B. A. (2015): "Promises and perils of pre-analysis plans," *The Journal of Economic Perspectives*, 29(3), 61–80.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011): "Strike Three: Discrimination, Incentives, and Evaluation," *The American Economic Review*, pp. 1410–1435.

de Paula, A., and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 80(1), 143–172.

Phelps, E. (1972): "The Statistical Theory of Racism and Sexism," *The American Economic Review*, 62(4), 659–661.

Posner, R. A. (1993): "What do Judges and Justices Maximize? (The Same Thing Everybody Else Does)," *Supreme Court Economic Review*, pp. 1–41.

——— (2008): *How Judges Think*. First Harvard University Press.

Price, J., and J. Wolfers (2010): "Racial Discrimination Among NBA Referees," *The Quarterly Journal of Economics*, 125(4), 1859–1887.

Rao, G. (2015): "Familiarity Does *Not* Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools," *Working Paper*.

Revesz, R. L. (1997): "Environmental Regulation, Ideology, and the DC Circuit," *Virginia Law Review*, pp. 1717–1772.

Riboni, A., and F. Ruge-Murcia (2014): "Dissent in Monetary Policy Decisions," *Journal of Monetary Economics*, 66, 137–154.

Schuessler, A. A. (2000): "Expressive Voting," *Rationality and Society*, 12(1), 87–119.

Seitz, C. J. (1991): "Collegiality and the Court of Appeals," *Judicature*, 75, 26.

Shearer, B. (2004): "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *Review of Economic Studies*, 71(2), 513–534.

Silvapulle, M. J., and P. K. Sen (2011): *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*, vol. 912. John Wiley & Sons.

Stasser, G., and W. Titus (1985): "Pooling of Unshared Information in Group Decision Making: Biased Information Sampling During Discussion," *Journal of Personality and Social Psychology*, 48(6), 1467.

Su, C. (2014): "Estimating Discrete-Choice Games of Incomplete Information: Simple Static Examples," *Quantitative Marketing and Economics*, 12(2), 167 – 207.

Sunstein, C. R. (2000): "Deliberative Trouble? Why Groups go to Extremes," *Yale Law Journal*, pp. 71–119.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.

Todd, P., and K. Wolpin (2006): "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *The American Economic Review*, 96(5), 1384–1417.

Visser, B., and O. Swank (2007): "On Committees of Experts," *The Quarterly Journal of Economics*, 122(1), 337–372.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, pp. 307–333.

Wan, Y., and H. Xu (2014): "Semiparametric Identification of Binary Decision Games of Incomplete Information with Correlated Private Signals," *Journal of Econometrics*, 182(2), 235–246.

Xu, H. (2014): "Estimation of Discrete Games with Correlated Types," *The Econometrics Journal*, 17(3), 241–270.

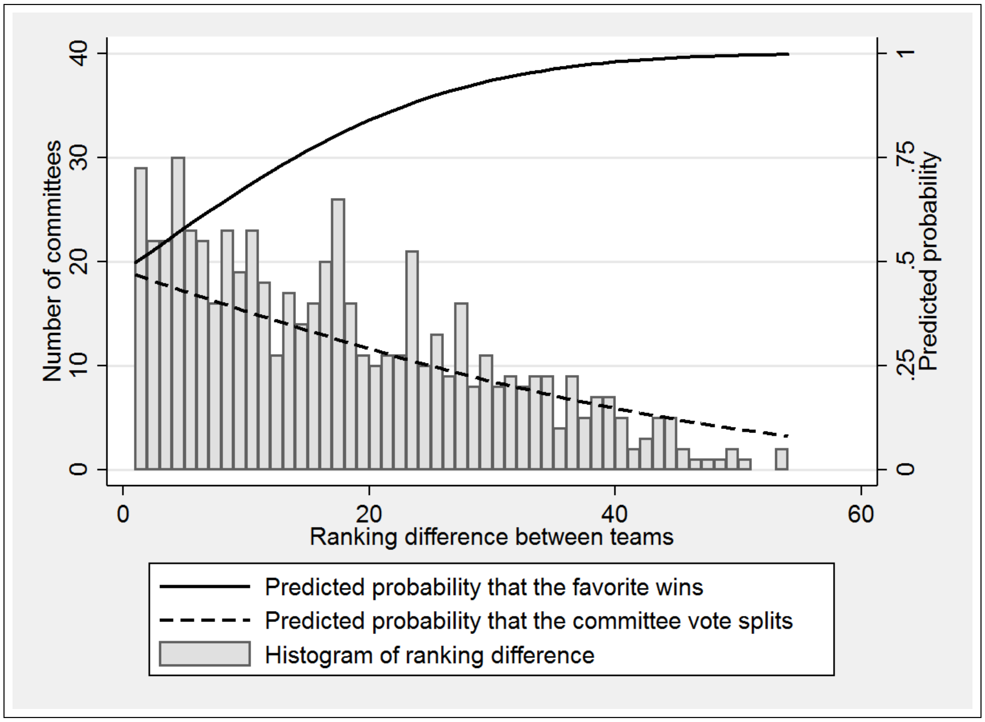
Figures and Tables

Table 1: **Description of key variables**

VARIABLE	MEAN	STD. DEV.	MIN.	MAX.
Class 1 votes for the favourite (dummy)	0.763		0	1
Class 2 votes for the favourite (dummy)	0.726		0	1
Class 3 votes for the favourite (dummy)	0.707		0	1
Committee votes for the favourite unanimously (dummy)	0.557		0	1
Committee votes for the favourite by 2-1 majority (dummy)	0.199		0	1
Committee votes against the favourite by 2-1 majority (dummy)	0.126		0	1
Committee votes against the favourite unanimously	0.118		0	1
Ranking difference between teams	17.0	12.1	1	54

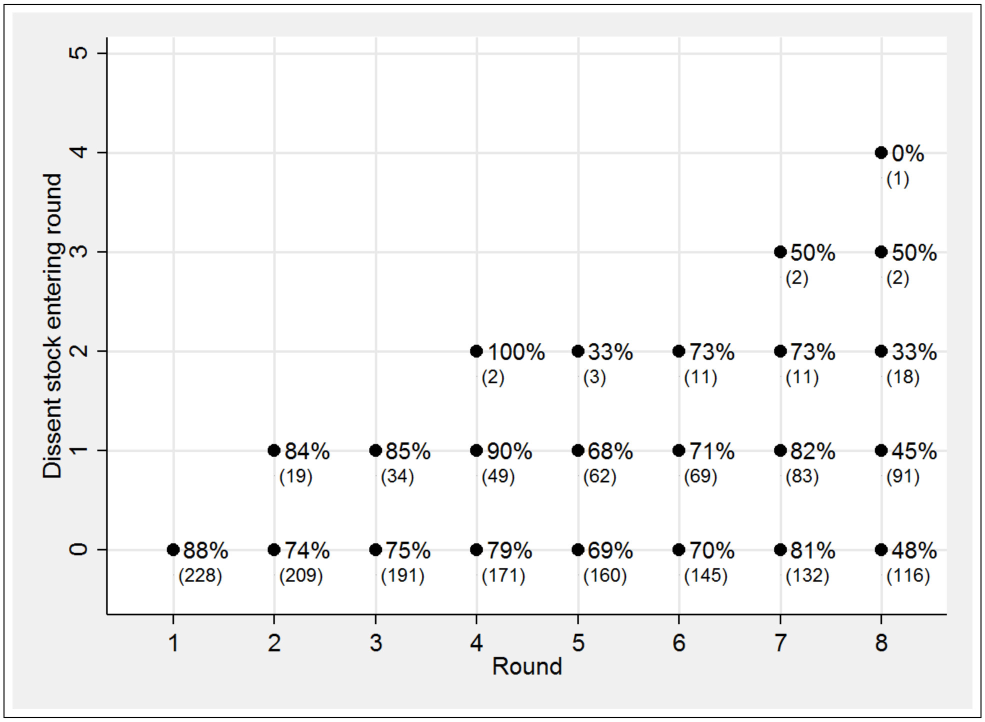
This table describes the key variables used in the analysis. We have 603 observations for each variable. Standard deviations for dummy variables are omitted for clarity.

Figure 1: **The role of ranking differences: Predicted values from probit estimations**



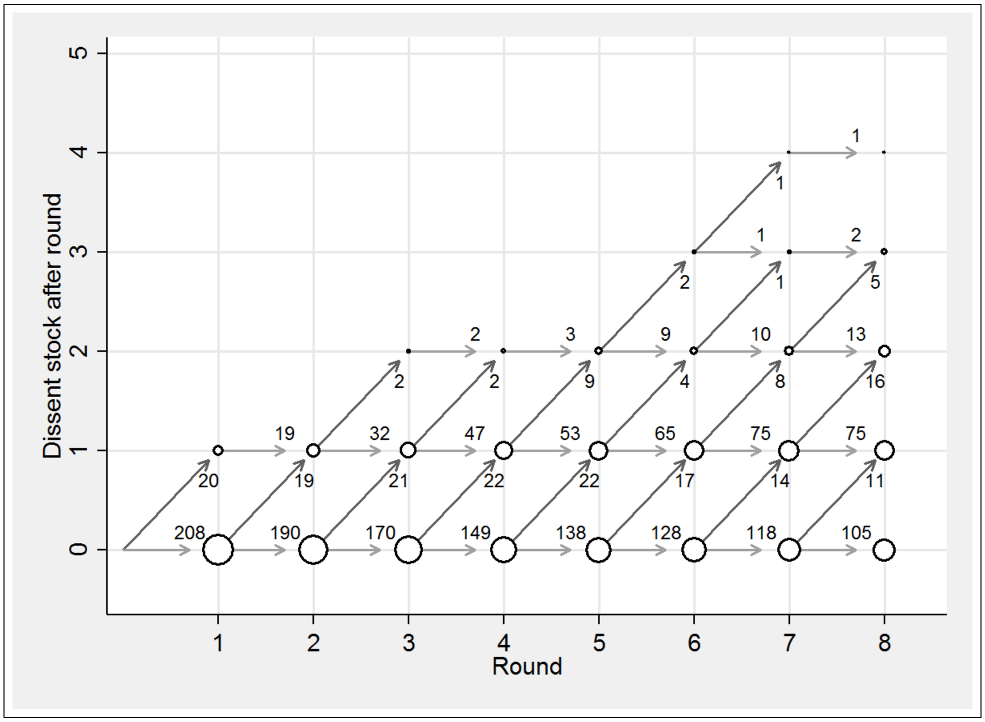
This figure shows the predictive value of the pre-tournament rankings. The x-axis shows the the difference in pre-tournament rankings between competing teams. The figure shows that the favourite is more more likely to win — and committees are more likely to vote unanimously — when the competing teams' pre-tournament rankings are closer.

Figure 2: **Sample probability of voting for the favorite by round and dissent stock**



This figure shows the sample probability of voting for the favorite, by round and dissent stock. The number of observations for each round-dissent combination is shown in parentheses below the sample probabilities.

Figure 3: **Stock of dissents by tournament rounds**



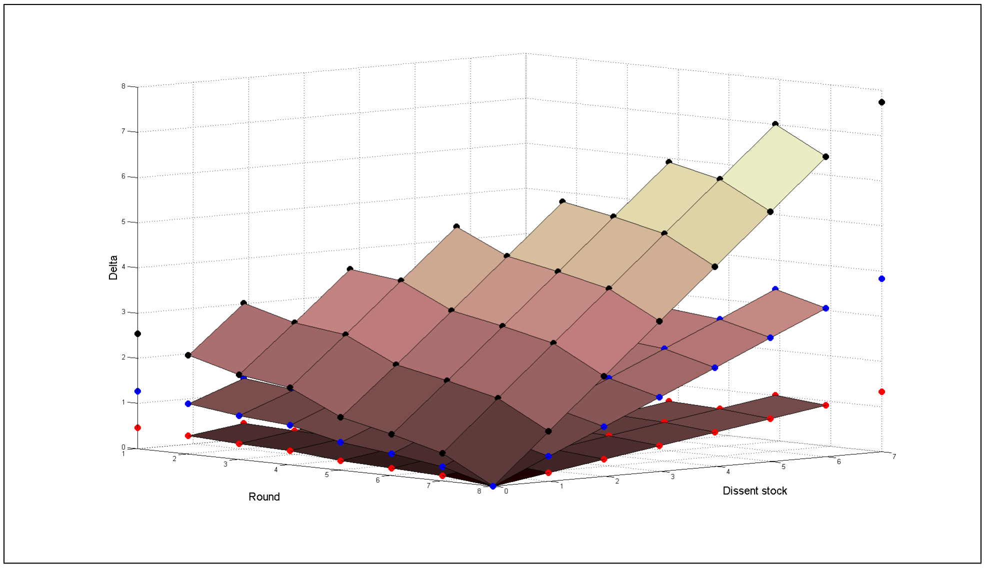
This figure shows the evolution of the dissent stock across tournament rounds. Circles are proportionate in size to the dissent stock in each round. Arrows show the number of judges transitioning between different points in the state space; the number at the head of each arrow shows the number of judges. (Note that the number of judges transitioning into a state does not always equal the number transitioning out; this is because some judges were rested in some rounds. The numbers refer to the judges actually participating in judging committees in a given round.)

Table 2: **Dynamic structural results**

PARAMETER	ESTIMATE	ℓ_r	LR	p-value
$\hat{\beta}_1$	0.026	-850.381	3.018	0.082*
$\hat{\beta}_2$	0.037	-860.070	22.395	0.000***
$\hat{\beta}_3$	0.019	-851.292	4.839	0.028**
$\hat{\rho}_{12}$	0.728	-909.551	121.357	0.000***
$\hat{\rho}_{13}$	0.517	-875.935	54.125	0.000***
$\hat{\rho}_{23}$	0.540	-879.378	61.011	0.000***
$\hat{\gamma}_1$	1.106	-851.054	4.363	0.018**
$\hat{\gamma}_2$	0.189	-849.298	0.851	0.178
$\hat{\gamma}_3$	0.547	-852.557	7.370	0.003***
LOG-LIKELIHOOD (ℓ_u)		-848.872		
H_0: $\rho_{12} = \rho_{13} = \rho_{13}$		-855.103	12.461	0.002***

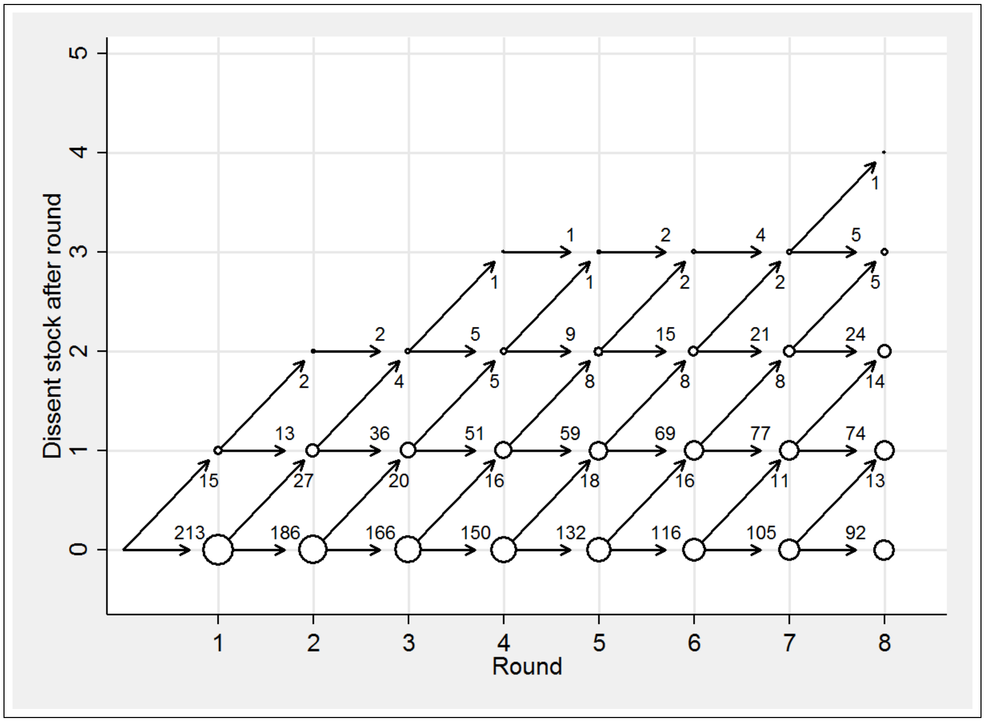
Confidence: *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Figure 4: **Coordination preferences over time**



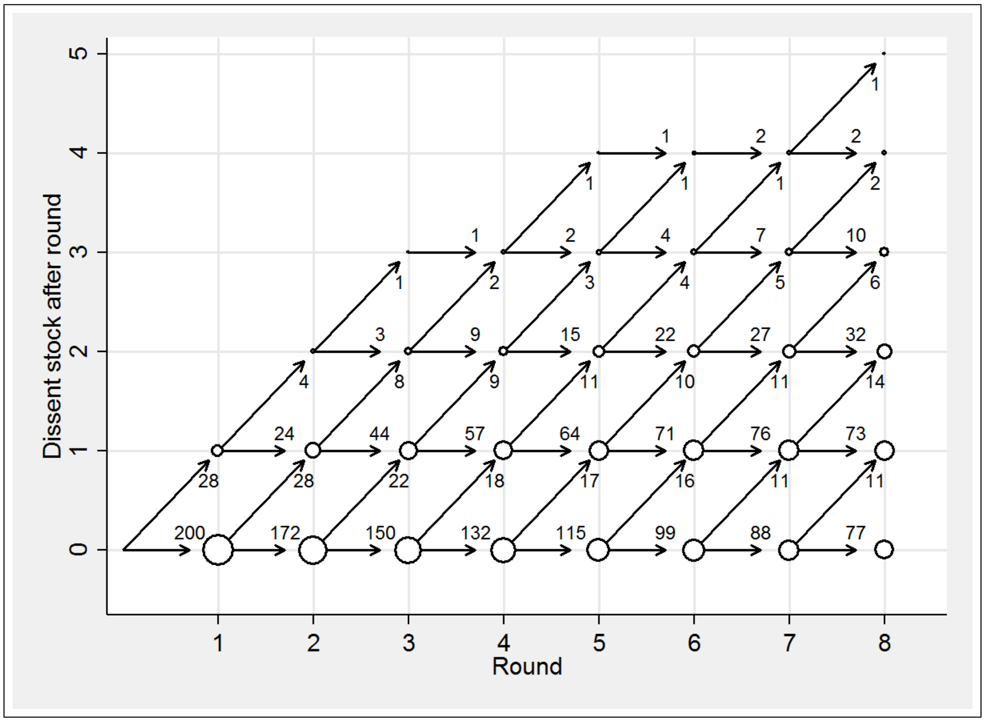
This figure shows the estimated coordination preference within each committee (i.e. δ) *as a function of tournament round and dissent stock (applied to the 2012 tournament). The top surface represents class 1 judges, the second surface represents class 3, and the lowest surface, class 2. The figure is useful for understanding the intuition of the dynamic model. The profile of the functions for tournament round 8 is the function* $\kappa(s_{iT}+1)-\kappa(s_{iT})$*, for each class of judge; note the identifying assumption* $\kappa(1)-\kappa(0)=0$*. As we would expect, the function increases monotonically in dissent stock — but, holding the stock constant, the function decreases monotonically in tournament round.*

Figure 5: **Stock of dissents by tournament rounds: structural results**



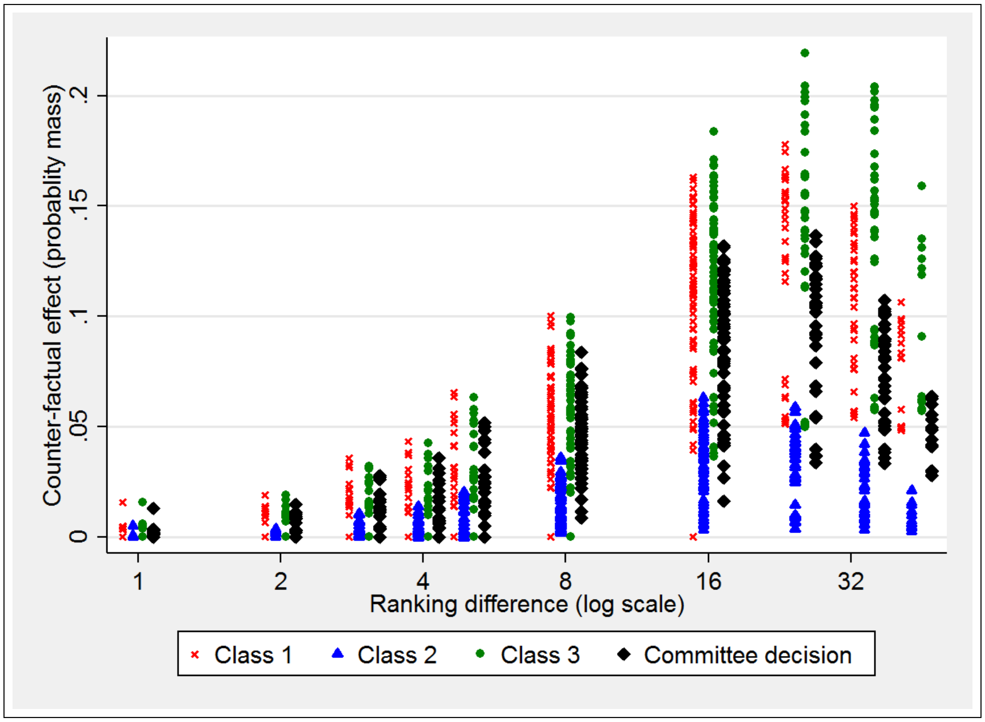
This figure shows the evolution of the dissent stock across tournament rounds, for the estimates from the dynamic structural model. Circles are proportionate in size to the dissent stock in each round. Arrows show the number of judges transitioning between different points in the state space; the number at the head of each arrow shows the number of judges.

Figure 6: **Stock of dissents by tournament rounds: counter-factual with no coordination preference**



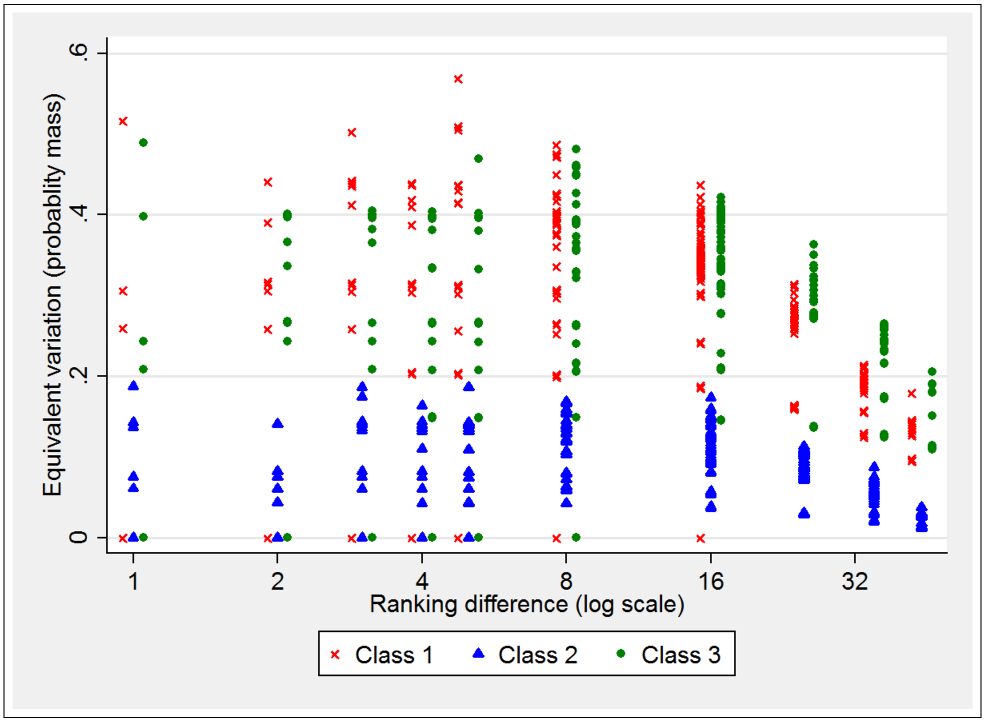
This figure shows the evolution of the dissent stock across tournament rounds for a counter-factual in which we set the coordination preference to zero ($\kappa_1 = \kappa_2 = \kappa_3 = 0$). Circles are proportionate in size to the dissent stock in each round. Arrows show the number of judges transitioning between different points in the state space; the number at the head of each arrow shows the number of judges.

Figure 7: **Mass shifted by coordination preferences**



This figure shows the estimated difference between the probability of voting for the favorite with a coordination preference and the probability of voting for the favorite with no coordination preference; we show this by each judge class separately, and then for the overall committee outcome. We show this for each committee separately, where we show the ranking difference on a log scale.

Figure 8: **Estimated equivalent variation, measured by the 'Faustian option'**



This figure shows the equivalent variation as a function of judge class and ranking difference. This is calculated by imagining a hypothetical 'Faustian option', in which each judge can sell his or her integrity in exchange for the agreement of his or her peers; the figure shows the proportion of cases in which each class of judge would take that option. This illustrates starkly the magnitude of the estimated preferences for coordination: in many committees, judges in class 1 and class 3 would be willing to abandon between approximately 20% and 50% of their decisions to avoid the risk of dissent.

Table 3: **Static structural results**

PARAMETER	ESTIMATE	ℓ_r	LR	p-value
$\hat{\beta}_1$	0.044	-952.089	201.499	0.000***
$\hat{\beta}_2$	0.042	-942.372	182.065	0.000***
$\hat{\beta}_3$	0.034	-913.067	123.455	0.000***
$\hat{\rho}_{12}$	0.726	-911.200	119.721	0.000***
$\hat{\rho}_{13}$	0.523	-878.584	54.489	0.000***
$\hat{\rho}_{23}$	0.545	-882.070	61.461	0.000***
$\hat{\delta}_1$	0.000	-851.340	0.001	0.487
$\hat{\delta}_2$	0.000	-851.339	0.000	1.000
$\hat{\delta}_3$	0.673	-852.557	2.435	0.059*
LOG-LIKELIHOOD (ℓ_u)		-851.339		
H_0: $\rho_{12} = \rho_{13} = \rho_{23}$		-857.123	11.567	0.003***

Confidence: *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Table 4: **Dynamic structural results: uncorrelated signals**

PARAMETER	ESTIMATE	ℓ_r	LR	p-value
$\hat{\beta}_1$	0.039	-967.559	39.082	0.000***
$\hat{\beta}_2$	0.042	-970.225	44.414	0.000***
$\hat{\beta}_3$	0.024	-955.623	15.210	0.000***
$\hat{\gamma}_1$	0.196	-948.386	0.736	0.195
$\hat{\gamma}_2$	0.000	-948.018	0.000	1.000
$\hat{\gamma}_3$	0.556	-951.154	6.272	0.006***
LOG-LIKELIHOOD (ℓ_u)		-948.018		

Confidence: *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

Table 5: **Testing for latent signal heterogeneity**

CLASS	$\hat{\psi}^2 \cdot (1+\hat{\psi}^2)^{-1}$	$H_0 : \psi = 0$ (p-value)
1	0.048	0.170
2	0.004	0.471
3	0.063	0.113

Appendix: Proofs

Proof of Proposition 1

Given that the support of x_i is the real line, it is sufficient for judge i to have a unique cutoff x_i^* that the difference in utility between $a_i = 0$ and $a_i = 1$ is monotonically increasing in x_i, holding fixed x_j^* and x_k^*, the cutoff points of the other judges.
In our setting, that difference is:

$$\begin{aligned} x_i^* + &\left\{ \Phi_2 \left[-\alpha_j(x_i^*), -\alpha_k(x_i^*), \omega_{jk} \right] - \Phi_2 \left[\alpha_j(x_i^*), \alpha_k(x_i^*), \omega_{jk} \right] \right\} \delta_i \\ &+ \left\{ \Phi_2 \left[-\alpha_j(x_i^*), \alpha_k(x_i^*), \omega_{jk} \right] - \Phi_2 \left[\alpha_j(x_i^*), -\alpha_k(x_i^*), \omega_{jk} \right] \right\} (\delta_{ij} - \delta_{ik}), \end{aligned}$$

The derivative of this difference with respect to x_i is:

$$\begin{aligned} 1 + \delta_i \cdot \sqrt{1 - \omega_{jk}^2} \cdot &\left[\frac{\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}} \cdot \phi\left(\alpha_j(x_i^*)\right) + \frac{\rho_{ik}}{\sqrt{1 - \rho_{ik}^2}} \cdot \phi\left(\alpha_k(x_i^*)\right) \right] \\ + \left(\delta_{ij} - \delta_{ik}\right) \cdot \sqrt{1 - \omega_{jk}^2} \cdot &\left[\frac{\rho_{ij}}{\sqrt{1 - \rho_{ij}^2}} \cdot \phi\left(\alpha_j(x_i^*)\right) - \frac{\rho_{ik}}{\sqrt{1 - \rho_{ik}^2}} \cdot \phi\left(\alpha_k(x_i^*)\right) \right]. \end{aligned}$$

Rearranging, this becomes: $1 + \left(\delta_i + \delta_{ij} - \delta_{ik}\right) \cdot \sqrt{1 - \omega_{jk}^2} \cdot \frac{\rho_{ij}}{\sqrt{1-\rho_{ij}^2}} \cdot \phi\left(\alpha_j(x_i^*)\right) + \left(\delta_i - \delta_{ij} + \delta_{ik}\right) \cdot \sqrt{1 - \omega_{jk}^2} \frac{\rho_{ik}}{\sqrt{1-\rho_{ik}^2}} \cdot \phi\left(\alpha_k(x_i^*)\right)$. For this to be positive, it is sufficient that $\delta_i \geq \delta_{ij} \geq 0$ and $\delta_i \geq \delta_{ik} \geq 0$. ■

Proof of Proposition 2

Our proof of Proposition 2 is an extension of the proof in Morris and Shin (2006) to three players. The proof requires some additional notation not included in the main text. Noting that Proposition 1 guarantees the existence of a threshold strategy, we now restrict our attention to such strategies.[56] Let $u_i(a_i, \Gamma_i(x_j^*, x_k^*, x_i), x)$ be judge i's expected payoff if their action is a_i; $\Gamma_i(x_j^*, x_k^*, x_i)$ is a bivariate normal cdf capturing their belief about the other two judge's actions given they believe the other judges to be following cutoff stategies of x_j^*, x_k^* and they observe signal x_i; and their signal is x. A strategy profile is triple $\mathbf{x} = (x_i^*, x_j^*, x_k^*)$.

Define $\Pi_i(\Gamma_i(x_j^*, x_k^*, x_i), x_i) = u(1, \Gamma_i(x_j^*, x_k^*, x_i), x_i) - u(0, \Gamma_i(x_j^*, x_k^*, x_i), x_i)$ — *i.e.* the expected difference in utilities between $a_i = 1$ and $a_i = 0$. Then $\mathbf{x}$ is an equilibrium iff $x_i > x_i^* \iff \Pi_i(\hat{\Gamma}_i(s_j, s_k, x_i), x_i) > 0 \; \forall \; i, \; x_i$.

Note that the problem set up in this way has the following properties:

(i) Uniformly positive sensitivity to the state: for $\underline{\kappa} \geq 1$, if $x \geq x'$,

$$[u(1, \Gamma, x) - u(0, \Gamma, x)] - [u(1, \Gamma, x') - u(0, \Gamma, x')] \geq \underline{\kappa}(x - x').$$

(ii) Uniformly bounded sensitivity to opponents' actions: for $\bar{\kappa} \geq 2\delta_i$

$$[u(1, \Gamma, x) - u(0, \Gamma, x)] - [u(1, \Gamma', x) - u(0, \Gamma', x)] \geq \bar{\kappa}(\Gamma - \Gamma'),$$

where

$$|\Gamma - \Gamma'| = sup|\Gamma(x_j^*, x_k^*, x_i) - \Gamma'(x_j^*, x_k^*, x_i)|.$$

(iii) Uniformly bounded marginals on differences: for $\nu \geq \sqrt{\frac{1-\omega_{jk}^2}{2\pi}} \left(\sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}} + \sqrt{\frac{1-\rho_{ik}}{1+\rho_{ik}}} \right)$, for all x, Δ_1 and Δ_2, $\frac{d}{dx_i} F_i(x_i + \Delta_1, x_i + \Delta_2 | x_i) \leq \nu$ where $F_i(x_j, x_k | x_k)$ is the conditional CDF of x_j and x_k given x_i.

By standard arguments in supermodular games (see Morris and Shin (2006)), there exist largest ($\bar{\mathbf{x}}$) and smallest ($\underline{\mathbf{x}}$) pure strategy profiles that satisfy iterated deletion of dominated strategies. We proceed to show by contradiction that if $\underline{\kappa} > \bar{\kappa}\nu$, those largest and smallest strategy profiles are the same.
Suppose $\bar{\mathbf{x}} \neq \underline{\mathbf{x}}$. Then translate the cutoffs of $\underline{\mathbf{x}}$ left until each judge's cutoff lies to the left of their cutoff under $\bar{\mathbf{x}}$, but that one translated cutoff is equal under the two profiles. Let z be the amount of the translation, and without loss of generality let player i with type $\hat{\mathbf{x}}_i$ be the player whose cutoff is the same under the two profiles. Write $\tilde{\mathbf{x}}$ for the translated strategy profile, and note that $\underline{x}_j = \tilde{x}_j + z$ for all j, and $\bar{x}_i = \tilde{x}_i$. We then have:

$$\begin{aligned} \Pi_i(\hat{\Gamma}_i(\underline{x}_j, \underline{x}_k, \underline{x}_i), \underline{x}_i) &= \Pi_i(\hat{\Gamma}_i(\underline{x}_j, \underline{x}_k, \bar{x}_i + z), \bar{x}_i + z) \\ &\geq \Pi_i(\hat{\Gamma}_i(\underline{x}_j, \underline{x}_k, \bar{x}_i + z), \bar{x}_i) + \underline{\kappa} z \\ &\geq \Pi_i(\hat{\Gamma}_i(\tilde{x}_j, \tilde{x}_k, \bar{x}_i), \bar{x}_i) + \underline{\kappa} z - \bar{\kappa}\nu z \\ &\geq \Pi_i(\hat{\Gamma}_i(\bar{x}_j, \bar{x}_k, \bar{x}_i), \bar{x}_i) + (\underline{\kappa} - \bar{\kappa}\nu)\, z \end{aligned}$$

[56] Note also that Proposition 1 is a three-player analogue to the final condition in section 3 of Xu (2014) which guarantees that the unique strategy profile is monotone strategies. Like Xu (2014), we leave extension of Xu (2014)'s more general proof of uniqueness to more than two parties for future work.

Since $\Pi_i(\hat{\Gamma}_i(\underline{x}_j, \underline{x}_k, \underline{x}_i), \underline{x}_i) = 0$ by definition of $\underline{\mathbf{x}}$, this implies that if $\underline{\kappa} > \bar{\kappa}\nu$, then $\Pi_i(\hat{\Gamma}_i(\underline{x}_j, \underline{x}_k, \underline{x}_i), \underline{x}_i) < 0$ and $\bar{\mathbf{x}}$ is not an optimal strategy. We thus have a contradiction and $\bar{\mathbf{x}} = \underline{\mathbf{x}}$. Therefore, a sufficient condition for a unique equilibrium is $\underline{\kappa} > \bar{\kappa}\nu$. Expanding and re-arranging gives Proposition 2:

$$\delta_i < \sqrt{\frac{\pi}{2(1-\omega_{jk}^2)}} \cdot \left(\sqrt{\frac{1-\rho_{ij}}{1+\rho_{ij}}} + \sqrt{\frac{1-\rho_{ik}}{1+\rho_{ik}}} \right)^{-1}.$$

■

Proof of Proposition 3

Start by considering i's decision where $\delta_j = \delta_k = 0$, and therefore $x_j^* = x_k^* = 0$. If $\delta_i = 0$, then $x_i^* = 0$ straightforwardly. At $x_i^* = 0$, with $x_j^* = 0$, note that if $\mu_j - \rho_{ij}\mu_i < 0$, $\alpha_j(x_i^*) = \frac{x_j^* - \mu_j - \rho_{ij}(x_i^* - \mu_i)}{\sqrt{1-\rho_{ij}^2}} < 0$, and similarly if $\mu_k - \rho_{ik}\mu_i$, $\alpha_k(x_i^*) < 0$. In that case, we have $\Phi_2\left[-\alpha_j(x_i^*), -\alpha_k(x_i^*), \omega_{jk}\right] - \Phi_2\left[\alpha_j(x_i^*), \alpha_k(x_i^*), \omega_{jk}\right] > 0$ and $\frac{\partial}{\partial \delta_i}\left[\Phi_2\left[-\alpha_j(x_i^*), -\alpha_k(x_i^*), \omega_{jk}\right] - \Phi_2\left[\alpha_j(x_i^*), \alpha_k(x_i^*), \omega_{jk}\right]\right] > 0$. It follows that if $x_i^* = 0$, $x_j^* = 0$ and $\mu_j - \rho_{ij}\mu_i < 0$, $\frac{\partial x_i^*}{\partial \delta_i} < 0$ for any value of $\delta_i \geq 0$.

Note that, so far, this proof does not rely on uniqueness. Let us now impose it, and consider increasing δ_{ij} and δ_{ik} to be greater than zero as well. We then need to consider the sign of the second and third parts of the equation that defines x_i^*:

$$\left\{\Phi_2\left[-\alpha_j(x_i^*), -\alpha_k(x_i^*), \omega_{jk}\right] - \Phi_2\left[\alpha_j(x_i^*), \alpha_k(x_i^*), \omega_{jk}\right]\right\} \cdot \delta_i$$
$$+\left\{\Phi_2\left[-\alpha_j(x_i^*), \alpha_k(x_i^*), -\omega_{jk}\right] - \Phi_2\left[\alpha_j(x_i^*), -\alpha_k(x_i^*), -\omega_{jk}\right]\right\} \cdot (\delta_{ij} - \delta_{ik}).$$

The fact that $\alpha_j(^*_i)$ and $\alpha_k(^*_i)$ are negative ensures the first difference is positive, but does not guarantee that the second difference is positive. However, it does guarantee that the first difference is larger in magnitude than the second difference, noting that the standard bivariate normal is symmetric. Since we have assumed $\delta_i > \delta_{ij} > \delta_{ik} > 0$, it follows that the entire sum *is* positive, so x_i^* is negative provided that $\mu_n - \rho_{in}\mu_i > 0 \; \forall \; n \in j, k$. Finally, consider the case where $\delta_j > 0$ and $\mu_n - \rho_{jn}\mu_j > 0 \; \forall \; n \in i, k$. By the argument above, $x_j^* < 0$. Because x_j^* enters positively in $\alpha_j(x_i^*)$, this makes $\alpha_j(x_i^*)$ more negative, and so the arguments above still hold: if $\mu_n - \rho_{in}\mu_i > 0 \; \forall \; n \in j, k$, $x_i^* \leq 0$. This result extends to the case where both δ_j and δ_k are greater than 0. ■

Proof of Proposition 4

The model is identified if we could solve for $(\rho_{12}, \rho_{13}, \rho_{23}, \beta_1, \beta_2, \beta_3, \delta_1, \delta_2, \delta_3)$, were we to observe the following eight conditional probabilities:

$$\begin{gathered}
\Pr(a_{1c} = 0, a_{2c} = 0, a_{3c} = 0 \mid R^c, D_1^c, D_2^c, D_3^c); \\
\Pr(a_{1c} = 1, a_{2c} = 0, a_{3c} = 0 \mid R^c, D_1^c, D_2^c, D_3^c); \\
\vdots \\
\Pr(a_{1c} = 1, a_{2c} = 1, a_{3c} = 1 \mid R^c, D_1^c, D_2^c, D_3^c).
\end{gathered}$$

Step 1: Identification of $(\rho_{12}, \rho_{13}, \rho_{23}, \beta_1, \beta_2, \beta_3)$. Consider the eight conditional probabilities for the case in which $D_{1c} = D_{2c} = D_{3c} = 0$. In that case, it is trivial that $x_1^{c*} = x_2^{c*} = x_3^{c*} = 0$. Then the conditional probabilities are wholly determined by a trivariate probit, and identification of $(\rho_{12}, \rho_{13}, \rho_{23}, \beta_1, \beta_2, \beta_3)$ is straightforward and well understood: see Ashford and Sowden (1970). By way of illustration, note that β_i is identified by the probability of $a_i = 1$, conditional on R^c: $\Pr(a_i = 1 \mid R^c, D_1^c = D_2^c = D_3^c = 0) = \Phi\left(\beta_i \cdot R^c\right)$. The parameters $(\rho_{12}, \rho_{13}, \rho_{23})$ are identified through correlations in observed outcomes, conditional on R^c. (Note that these parameters can even be identified simply through pairwise correlations: see Kimhi (1994).)

Step 2: Identification of $(\delta_1, \delta_2, \delta_3)$. Consider the eight conditional probabilities for the three cases in which $D_i = 1; D_j = D_k = 0 \ \forall\ j, k \neq i$. Note that, in those three cases, $\delta_i^c = \delta_i$ and $\delta_j^c = 0$, and therefore, $x_j^{c*} = 0$. Then the first part of Proposition 3 shows that there is a one-to-one relationship between δ_i and x_i^{c*}, as the comparative static with respect to δ_i, holding j and k's cutoffs constant, is monotone. Then we are done if we can identify x_i^*. We can identify x_i^* from the probability that $a_i = 1$, treating β_i as known and choosing any fixed value for R^c: $\Pr(a_i = 1 \mid R^c, D_i^c = 1, D_{j \neq i}^c = 0, D_{k \neq i}^c = 0) = \Phi\left(\beta_i \cdot R_c - x_i^*\right)$. ■

Online Appendices

Online Appendix A: Pre-registration of structural predictions

Model validation is an important challenge for structural analysis.[57] In this paper, we take a novel approach to structural model validation: before proceeding to develop our dynamic model, we used our stage game to make testable predictions about the 2013 World Schools Debating Championships, and registered the model and predictions at the J-PAL Hypothesis Registry.[58] To our knowledge, this approach to out-of-sample validation of a structural model has not been used before.

Table 6 summarises both in-sample and out-of-sample predictions; the out-of-sample predictions (the final three columns) relate directly to the registered hypotheses. In each case, we took the particular tournament structure (*i.e.* the assigned match-ups between opponents of differing pre-tournament rankings), and ran 1000 simulated versions of the tournament. We then compare moments between actual and simulated tournaments. In each case, we compare the observed moment to the mean of the simulated distribution; we also report the corresponding percentile of the simulated distribution.[59] We underline measured moments lying outside a 90% confidence interval in our simulated distribution (that is, values whose percentile is less than 5% or greater than 95%).

In general, the static model performed reasonably well in matching cross-sectional moments, both in-sample and out-of-sample. Of the 72 moments reported, nine lie outside the 90% confidence interval from the simulated data; of the 24 moment predictions made for the 2013 tournament, three lie outside the same interval. Insofar as the model performs poorly, it does so in predicting the probability that Class 2 dissents. This appears to be driven by an unusual variation in Class 2 behavior between tournaments; something that lies beyond the scope of our model framework.

[57] See, for example, Keane (2010, page 18), who argues, "It has often been treated as a feat worthy of praise to simply estimate a structural model, regardless of whether the model can be shown to provide a good fit to the data, or perform well in out-of-sample predictive exercises. I see no reason why an estimated structural model should move my priors about, say, the likely impact of a policy intervention, if it fits the in-sample data poorly and/or has not been shown to perform reasonably well in any validation exercises. Structural econometricians need to do a much better job in this area in order for structural work to gain wider acceptance."

[58] See `http://www.povertyactionlab.org/Hypothesis-Registry`. Note that, in that document, we estimated a subtly different version of the stage game, in which we allowed the ranking difference to affect the signal mean through a quadratic specific, rather than a linear specification. This does not substantially change the model implications, which is why we constrained to the linear specification for our main estimations.

[59] Our registered document contained graphs of the empirical CDFs from our simulation exercise; in this way, we registered not merely a predicted mean of each moment, but its entire distribution.

Table 6: GOODNESS OF FIT: IN-SAMPLE AND OUT-OF-SAMPLE STRUCTURAL PREDICTIONS

	IN-SAMPLE									OUT-OF-SAMPLE		
	2010			**2011**			**2012**			**2013**		
MOMENT (PROBABILITY)	*pred.* (%)	*actual* (%)	*perc.* (%)	*pred.* (%)	*actual* (%)	*perc.* (%)	*pred.* (%)	*actual* (%)	*perc.* (%)	*pred.* (%)	*actual* (%)	*perc.* (%)
Class 1												
judge dissents	8.8	8.9	**59.4**	9.0	8.9	**53.7**	8.9	8.4	**45.1**	8.8	10.7	**82.3**
…if just dissented	8.6	7.7	**50.8**	8.5	0.0	**31.2**	8.5	8.3	**52.4**	8.4	0.0	**32.3**
…if not just dissented	8.8	8.9	**54.3**	9.0	9.3	**57.5**	9.0	8.4	**41.3**	8.8	11.8	**90.9**
judge votes for the favorite	76.4	76.0	**39.4**	75.6	76.6	**64.3**	75.8	78.5	**81.3**	76.3	74.3	**22.6**
…if just dissented	76.3	46.2	**1.1**	75.1	90.0	**92.0**	75.7	66.7	**24.1**	76.1	76.5	**48.0**
…if not just dissented	76.4	78.2	**71.4**	75.6	75.8	**51.8**	75.8	79.3	**88.2**	76.3	74.1	**21.6**
Class 2												
judge dissents	8.4	4.7	**3.6**	8.6	9.4	**68.3**	8.7	12.0	**95.3**	8.5	13.4	**98.3**
…if just dissented	7.7	16.7	**89.7**	8.4	7.1	**44.9**	7.8	14.3	**83.4**	8.2	10.0	**65.5**
…if not just dissented	8.5	3.9	**0.6**	8.6	9.6	**66.2**	8.7	11.8	**90.4**	8.6	13.8	**98.9**
judge votes for the favorite	74.7	75.0	**56.4**	73.6	72.9	**37.5**	73.7	71.7	**24.5**	74.2	75.9	**69.7**
…if just dissented	74.9	75.0	**50.5**	73.6	57.1	**8.3**	74.0	71.4	**36.4**	74.4	75.0	**51.7**
…if not just dissented	74.6	75.0	**53.5**	73.6	74.2	**57.3**	73.7	71.8	**25.6**	74.2	76.1	**72.1**
Class 3												
judge dissents	15.0	18.2	**89.3**	15.4	15.6	**51.8**	15.4	12.6	**13.3**	15.2	15.5	**53.0**
…if just dissented	12.6	5.0	**15.6**	13.2	14.3	**59.7**	13.1	26.7	**97.7**	13.1	19.1	**82.1**
…if not just dissented	15.4	19.8	**94.3**	15.7	15.8	**53.7**	15.8	11.4	**4.2**	15.5	15.1	**44.8**
judge votes for the favorite	71.9	71.9	**51.9**	70.8	72.9	**71.3**	70.9	67.0	**11.4**	71.5	71.7	**50.1**
…if just dissented	77.0	70.0	**21.5**	75.7	66.7	**16.9**	75.8	80.0	**68.4**	76.4	57.1	**2.1**
…if not just dissented	71.1	72.1	**58.4**	70.1	73.7	**85.5**	70.2	65.9	**11.8**	70.7	73.5	**76.6**

This table reports in-sample and out-of-sample predictions for the 2010, 2011, 2012 and 2013 World Schools Debating Championships. Predictions are formed from 1000 simulations. 'Pred' reports the mean predicted statistic. 'Actual' reports the actual statistic. 'Perc' reports the percentile for the actual statistic in the empirical distribution of 1000 simulations. 'Judge dissents' reports the unconditional probability of a judge dissenting. 'Judge dissents…if just dissented' reports the probability of a judge dissenting, conditional on having just dissented in the previous round. 'Judge dissents…if not just dissented' reports the probability of a judge dissenting, conditional on not having just dissented in the previous round. The same structure applies for statistics reporting 'votes for the favorite'. All data — except for the final two columns — were reported in our 'Structural Predictions' document, registered at http://www.povertyactionlab.org/Hypothesis-Registry *on 26 January 2013. That document provides more information — for example, it plots the Empirical CDFs from which the last column is generated.*

Online Appendix B: An example ballot

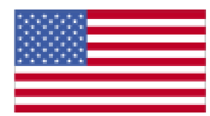
Figure 9 shows an example of a judge's ballot, taken from a debate between Kuwait and the United States. This judge voted for the United States (by a margin of 243 marks to 231). Note that the total score is simply the sum of the individual speaker totals, and that each individual speaker total is simply the sum of marks for the categories 'Style', 'Content', 'Strategy', and 'Points of Information' ('P.o.I'). 'Style' refers to the way that a speaker presents: for example, whether the the speech is delivered at an appropriate speed, and whether the speaker makes eye contact with the audience. 'Content' refers to the substance of the arguments: for example, whether arguments are logical, and substantiated with persuasive evidence. 'Strategy' refers to the speaker's identification of the key issues in the debate, and the consistency of the speaker's material with the material of his or her teammates. 'Points of Information' are short interjections that speakers are permitted to make during their opponents' speeches.

< Figure 9 here. >

Figure 9: An example ballot from one judge

Proposition **KUWAIT**

Speakers	Style	Content	Strategy	P.o.I.	Total
1 [Speaker name here]	28	26	13	0	67
2 [Speaker name here]	27	25	12.5	0	64.5
3 [Speaker name here]	28	25	13	0	66
R [Speaker name here]	14	13	6.5		33.5
Total					**231**



Opposition **USA**

Speakers	Style	Content	Strategy	P.o.I.	Total
1 [Speaker name here]	29	27	13	0	69
2 [Speaker name here]	30	28	14	0	72
3 [Speaker name here]	27	26	13	0	66
R [Speaker name here]	15	14	7		36
Total					**243**

Online Appendix C: Solving and estimating the model

We use the following algorithm to estimate the log-likelihood for a given parameter vector $\boldsymbol{\theta}$:

(i) Start with tournament 1 (Doha).

(ii) Use the function κ to calculate V_{T+1}, for all values of $s_{iT} \in \{0, \ldots, T\}$.

(iii) For judges of class 1 with dissent stock $s_{1T} = 0$, loop over the elements of the Cartesian product of (i) the ranking conditions observed in the data for period T (that is, R_T), (ii) the dissent stock for judges of class 2 entering period T (that is, S_{2T}) and (iii) the dissent stock for judges of class 3 entering period T (that is, S_{3T}). For each iteration of this loop, solve the stage game. Save the cutoff for the class 1 judge, $x^*_{1T}(r, s_{1T}, s_{2T}, s_{3T}; \boldsymbol{\theta})$. Repeat this exercise for all *possible* values of s_{1T} (*i.e.* $s_{1T} \in \{0, \ldots, T-1\}$, even if not all $s_{1T} \in \{0, \ldots, T-1\}$ are observed in the data).

(iv) Use the Bellman equation (equation 8) to calculate V_T for judges of class 1, for all values of $s_{iT} \in \{0, \ldots, T-1\}$.

(v) Repeat this process, symmetrically, for judges of classes 2 and 3.

(vi) Repeat steps (ii), (iii) and (iv), iterating backwards from period $T+1$ to period 1. (This implies that we assume — for simplicity and for tractability — that each judge participates in the full eight rounds of competition, and that this is known with certainty. Empirically, this is a very reasonable assumption in this context, since most judges do participate in all or almost all rounds; specifically, almost 90% of decisions were made by judges who judged six, seven or eight rounds.)

(vii) For committee c in round t, we observe a ranking difference (r_{ct}), a dissent stock for each class of judges ($s_{1t,c}$, $s_{2t,c}$ and $s_{3t,c}$), and a binary outcome for each judge recording whether the judge voted for the favorite (a_{1ct}, a_{2ct} and a_{3ct}). Having solved the model for the entire equilibrium path, we extract the values of $x^*_{1t}(r, s_{1t,c}, s_{2t,c}, s_{3t,c}; \boldsymbol{\theta})$ actually observed in the data. The log-likelihood for committee c in round t, the log-likelihood is therefore:

$$
\begin{aligned}
&\ell_{ct}\left(\boldsymbol{\theta}; a_{1ct}, a_{2ct}, a_{3ct} \,|\, r_{ct}, s_{1t,c}, s_{2t,c}, s_{3t,c}\right) \\
&\quad = \ln \Phi_3 \left[(2a_{1ct} - 1) \cdot \left(\beta_1 \cdot r_{ct} - x^*_1(r_{ct}, s_{1t,c}, s_{2t,c}, s_{3t,c})\right), \right. \\
&\qquad\qquad (2a_{2ct} - 1) \cdot \left(\beta_2 \cdot r_{ct} - x^*_2(r_{ct}, s_{1t,c}, s_{2t,c}, s_{3t,c})\right), \\
&\qquad\qquad \left. (2a_{3ct} - 1) \cdot \left(\beta_3 \cdot r_{ct} - x^*_3(r_{ct}, s_{1t,c}, s_{2t,c}, s_{3t,c})\right), \boldsymbol{\Omega}\right],
\end{aligned}
$$

where Φ_3 denotes the *cdf* of the trivariate normal, and $\boldsymbol{\Omega}$ denotes the signal covariance matrix). We approximate the *cdf* of the trivariate normal using the method of Genz (2004).

(viii) We treat draws of (x_1, x_2, x_3) as independent across committees, so the sample log-likelihood is $\ell(\boldsymbol{\theta}) = \sum_{c,t} \ell_{ct}\left(\boldsymbol{\theta}; a_{1ct}, a_{2ct}, a_{3ct} \,|\, r_{ct}, s_{1t,c}, s_{2t,c}, s_{3t,c}\right)$.

(ix) Repeat the entire exercise for tournament 2 (Dundee) and tournament 3 (Cape Town). The total sample log-likelihood is the the sum of the log-likelihoods for each of the three tournaments.

(x) We nest this entire algorithm in a Sequential Quadratic Program and iterate until convergence.