The Politics of Expertise: Biased Learning in International Organizations

Amanda Kennard*

September 16, 2017

Abstract

Negotiators of international economic agreements must contend with uncertainty about how formal rules translate into changes in the behavior of signatory states and private actors. Information gathering to reduce this uncertainty is a key component of international negotiations. Outside observers can infer the results of information gathering through its influence on resulting agreements. Yet inference is imperfect since agreements are also shaped by the preferences of negotiating parties. I develop these arguments in the context of negotiations surrounding the accession of new members to the World Trade Organization (WTO). I analyze a formal model of sequential negotiations in which outcomes reflect novel information and agent preferences, demonstrating that the outcome of negotiations may be persuasive even if it is dominated by partisan preferences. I test the model's implications by studying states' participation in accession negotiations. I find consistent support for the argument, taking into account a range of alternative explanations.

*Ph.D. Candidate, Princeton University. For many helpful comments and suggestions I thank Marc L. Busch, Allison Channey, Christina Davis, Manfred Elsig, Matías Iaryczower, Federica Izzo, Helen Milner, Julia Morse, John Patty, Krzysztof Pelc, Maggie Penn, Arlo Poletti, Tyler Pratt, Kris Ramsay, Brandon Stewart, and Simon Wuethrich.

1 Introduction

The original architects of the General Agreement on Tariffs and Trade grappled with a central challenge. Trade liberalization requires disciplines on a range of policy instruments. Many of these instruments - tariffs or quantitative import restrictions among others - relate directly to international trade. Yet there are many ways to restrict trade using policies nominally unrelated to international exchange. Domestic taxes, regulation, and other behind the border measures can all erode the effectiveness of legal disciplines on trade-related measures. Which types of policies should be ruled out or disciplined in the pursuit of trade liberalization and which should be regarded as legitimate social or economic policy?[1]

This type of uncertainty - about the optimal design of institutions - is pervasive in international cooperation. While bargaining over the design of treaties or other agreements, negotiators must anticipate the likely impacts of design choices on the behavior of states and private actors. The impact of these choices frequently depends on unobserved - or unobservable - characteristics of negotiating partners, domestic interest groups, or even the global economy. Over the long run, the evolution of cooperation depends crucially on the accumulation of domain-specific knowledge.

Yet this accumulation of knowledge is complicated by the far-reaching distributional implications of international economic negotiations. The agreements which result from these negotiations are shaped as much by the material interests of influential participants as by the beliefs of technocrats about the optimal design of institutions. To take another example from the trade regime, an evolving technocratic consensus on the need for liberalization played an important role in shaping the outcome of services negotiations during the Uruguay Round. Policy makers, initially skeptical of the inclusion of services in the negotiating agenda, ultimately embraced this new form of cooperation following years of study and information

[1] As Howse notes, "There is no natural or self-evident baseline or rule that can solve this basic dilemma" (Howse, 2002).

dissemination by international organizations, academics, and members of the private sector.

At the same time, the final text of the General Agreement on Trade in Services (GATS) also reflects the material interests of the most influential party to the negotiations. Beginning in the 1970's the world's largest service-providing firms, headquartered in the United States, increasingly found their access to foreign markets restricted by a range of regulatory barriers. For U.S. negotiators these firms provided crucial political support for Uruguay Round negotiations, offsetting the growing skepticism of manufacturers towards increased liberalization (Drake and Nicolaidis, 1992).

I argue that this tension - between the resolution of technocratic uncertainty and the pursuit of self-interest - has profound implications for the bargaining behavior of states and the evolution of cooperation over time. The paper advances two main claims. First, the potential for international economic agreements to convey newly-gathered information about the efficient design of institutions leads to path dependency in equilibrium. Outside observers - unable to observe either the results of information gathering or the precise distribution of preferences among negotiating partners - rationally update their beliefs about the optimal design of institutions in light of past bargaining outcomes. Second, this path dependency creates added incentives for states to shape bargaining outcomes to their (far-sighted) material advantage.

I develop these arguments in the context of committee decision making in international organizations. While relatively under-studied, bureaucratic committees are a central feature of many international organizations.[2] IO committees are tasked with many, varied functions, yet nearly all engage regularly in the collection and dissemination of policy-relevant information.[3] This information can have far-ranging implications for the design of cooperation more

[2]For example, the United Nations General Assembly boasts twenty nine subsidiary committees, including committees dedicated to peace-keeping operations, south-south cooperation, and others. Members of the OECD meet regularly in committees and working groups numbering in the hundreds. Similarly, much of the day-to-day politics of the World Trade Organization takes place in committees and working parties: in 2016 alone the organization's twenty eight permanent committees and subcommittees held an average of nineteen meetings per month.

[3]One committee within the International Labor Organization, for example, is tasked with reviewing annual reports on the implementation of ILO standards and making specific recommendations for improving

broadly. Members' participation in IO committees should then reflect strategic consideration of these long-run effects, providing a valuable means of testing the arguments advanced here.

I begin by developing a game theoretic model of committee decision making, drawing on models of information aggregation and herding behavior.[4] In the model non-overlapping committees select policy outcomes sequentially. Before doing so, committee members observe a common signal about the state of the world and any prior policy choices. Committee members may be one of two types: *technocrat* types who wish to tailor the policy to the state of the world and *partisan* types whose preferred policy is unaffected by the state. Committee outputs are shaped by both novel information and the distribution of committee member types. Yet joint uncertainty about the state of the world and the composition of committees implies that prior decisions will be partially persuasive. In this sense precedent emerges endogenously in equilibrium.

In turn this creates powerful incentives for agents to participate in committees in order to influence the evolution of policy making more broadly. To explore this possibility I extend the model to consider the strategic decision of agents to participate in committee decision making. In equilibrium, agents with higher long-term value for the policy area are more likely to participate at any stage. In addition, all agents are more likely to participate in committee decision making early on in the game, when doing so enables them to influence not only the current outcome, but also future policy outcomes.

I test implications of the model in the context of negotiations surrounding the accession of new members to the World Trade Organization. These negotiations - each overseen by a specially constituted *ad hoc* working party - provide a valuable setting in which to test the model's implications since they exemplify the tension between technocratic uncertainty and narrow self-interest. In particular, the predominance of low- and lower-middle income

compliance. A subsidiary committee of the World Intellectual Property Organization (WIPO) is tasked with determining the need for, and negotiating the substance of, new treaty provisions relating to patent protection.

[4]See for example Ottaviani and Sorensen (2001) and Iaryczower (2007).

states among applicants to the WTO means that negotiators must navigate one of the most contentious questions in contemporary trade politics: how to design multilateral rules to support trade liberalization while respecting the developmental circumstances of member states.

In the course of accession negotiations, applicants undertake binding policy-related commitments the design of which is not limited to existing WTO agreements. Rather these policy commitments may diminish rights and protections previously enshrined in WTO law or extend the multilateral agreements into entirely new areas. From a technocratic perspective, balancing the goal of trade liberalization with developmental concerns requires extensive information gathering. Yet from the perspective of self-interested member states, by affecting the balance of concessions expected from WTO members at varying levels of development, the outcome of individual accession negotiations may have long-term distributional implications.

I test the model's predictions using text analysis of a novel corpus of negotiating documents and newly-collected data on the participation of existing WTO Members in all accession negotiations to date. Statistical tests support the argument that participation reflects farsighted concern about the evolution of global trade rules. Members are more likely to participate in negotiations when they touch upon subjects in which the member has previously revealed an interest, signaling that the policy has a high long-term salience to that Member. I also demonstrate that members are more likely to participate in negotiations early on: they participate at higher rates in negotiations which touch upon systemically novel subjects and are thus likely to have the greatest impact on future decision making. These results are robust to a range of alternative explanations.

This work contributes to a substantial literature on the rational design of international institutions (Koremenos et al., 2002). A number of studies have analyzed how the design of international institutions can mitigate or exacerbate underlying strategic obstacles to

cooperation (Martin, 1992). Scholars have also analyzed the allocative consequences of institutional design (Allee et al., N.d.; Krasner, 1991). The current work builds on these literatures by considering how the imperatives of technical efficiency and partisan interest jointly shape the evolution of international cooperation. As such, it also provides a novel account of the dynamics of international cooperation, particularly the path dependency often observed in international agreements (Alschner, 2016).

The paper is also closely related to literature on epistemic communities and the role of experts in shaping international cooperation (Haas, 1992). A number of papers analyze case studies of international agreements with the goal of determining the relative influence of a class of international technocrats and the parochial interests of negotiating parties over the eventual form of cooperation (Adler, 1992; Kapstein, 1992; Ikenberry, 1992).[5] This paper goes beyond existing studies by considering how uncertainty about this relative influence itself shapes cooperation over time. I also present a more nuanced view of expert knowledge itself as dynamic and shaped in part by the existing architecture of international cooperation.

In the following sections I develop and analyze a formal model of committee decision making. Section four describes my empirical strategy for testing implications of the model in the context of WTO accession negotiations. The empirical results are described in section five. A final section concludes.

2 Model

There are two periods, $t = 1, 2$. In each period a committee, C_t, chooses a policy outcome, $x_t \in \{0, 1\}$. Membership is non-overlapping across the two committees, and each committee consists of exactly k members.[6] The value of the two policy decisions depends on a single

[5] See Cross (2012) for a review of more recent work.

[6] The assumption of symmetric committee size is made for notational simplicity. Allowing commitee size to vary does not alter the results.

unobserved state, $\omega \in \{-1, 1\}$ where the probability that $\omega = 1$ is $\pi \in (0, 1)$. Before selecting a policy, all members of C_t observe a common signal, $s_t \in \{-1, 1\}$, which matches the state with probability p. Members of C_2 observe the outcome x_1 prior to making their decision, but not the first period signal or the record of first period voting.

After realization of the signal, s_t, a policy is chosen by unanimity rule where $x_t = 1$ if and only if all committee members vote in favor of that outcome. If any member dissents then the reversion policy, $x_t = 0$, is implemented. Throughout the analysis I restrict attention to undominated voting strategies and assume that members who are indifferent vote in favor of $x_1 = 1$ (Feddersen and Pesendorfer, 1997).

There are two types of agent, $o \in \{A, B\}$. Utility reflects both policy choices and varies according to type. Utility of agent i is,

$$u_i(x_1, x_2) = \begin{cases} \omega(x_1 + x_2) & \text{if } i \text{ is type } A \\ -(x_1 + x_2) & \text{if } i \text{ is type } B \end{cases}$$

That is, total payoffs are additive across policies and accrue only at the end of the game. Note that the ideal policy of type A agents depends on the state of the world. If the state is $\omega = 1$ an agent of type A prefers the policy $x_t = 1$ in both periods. If the state is $\omega = -1$, an A type prefers $x_t = 0$. In contrast, B types always prefer $x_t = 0$, regardless of the state of the world. We can think of A types as technocratically-minded participants who prefer to implement the "correct" solution and B types as pure partisans who always prefer a particular outcome regardless of circumstances. I assume that types are private information.

The probability that any given agent is a B type is θ^0. Given a committee of size k, denote the probability that there is at least one B type present by,

$$\theta = 1 - (1 - \theta^0)^k \tag{1}$$

Finally, I make the following assumption:

A1. $p > \pi > 1 - p$.

Assumption A1 ensures that the signal is informative relative to the prior. A **committee decision making game** is $\Gamma^c = (\theta^0, \pi, p, k)$. The solution concept is Perfect Bayesian Equilibrium.

3 Results

The focus of the following analysis is the endogenous emergence of precedent in committee decision making and the resulting incentives of agents to shape decision making to their (far-sighted) advantage. I first characterize optimal behavior in the committee decision making game. A subsequent section extends the model to consider agents' strategic participation in committees.

3.1 The Emergence of Precedent

In this section I provide an intuitive explanation of equilibrium behavior before introducing the main results concerning the emergence of precedent. A strategy for agent i consists of a voting rule which maps beliefs about the state of the world to a policy choice. Let $\mu(x_1, s_2)$ be the posterior probability that the state is one given first period outcome x_1 and signal s_2. The first result characterizes equilibrium of the participation game.[7]

Proposition 1. Γ^c *admits a unique pure strategy Perfect Bayesian Equilibrium:*

— *First period agents of type A vote in favor of* $x_1 = 1$ *if and only if* $s_1 = 1$,

— *Second period agents of type A vote in favor of* $x_t = 1$ *if and only if* $\mu(x_1, s_2) \geq \frac{1}{2}$, *and*

— *In either period, agents of type B vote in favor of* $x_t = 0$.

[7] All proofs are included in Appendix A below.

Proposition 1 establishes a linkage in decision making across committees. While the votes of technocrats in the first period are surely responsive to the observed signal, the same is not true in the second period. Instead, optimal second period voting depends on posterior beliefs. It is this dependence that links decision making over time reflecting the influence of x_1 on second period beliefs.

The role of posterior beliefs highlights the inferential challenge of technocrats in the second period. Since A types vote informatively in the first period, there is a positive probability that the outcome of decision making is likewise informative. Yet the votes of B types in the first period are strictly uninformative since they are unresponsive to the signal s_1. Unanimity rule implies that B types are also highly influential: if any B type is present in the first period committee, the outcome of decision making will surely be $x_1 = 0$. Technocrats in the second period would like to update their beliefs in light of x_1 if and only if it is informative about s_1 and hence the state of the world. Yet without knowing the distribution of types in C_1 they cannot be certain if an outcome $x_1 = 0$ reflects the signal itself or merely the presence of partisan interests.[8]

Consider equilibrium beliefs in the second period when $x_1 = 0$. If $s_2 = 1$,

$$\mu(0,1) = \frac{p(p\theta + (1-p))\pi}{p(p\theta + (1-p))\pi + (1-p)[(1-p)\theta + p](1-\pi)} \tag{2}$$

Beliefs incorporate both the possiblity that x_1 accurately transmits a negative signal $(1-p)$ and the possibility that a positive signal was blocked by the presence of B types $(p\theta)$. As θ approaches one, the expression above approaches $\mu(0,1) = \frac{p\pi}{p\pi+(1-p)(1-\pi)}$. When it is certain that a B type participated in the first period, beliefs are identical to those that would obtain if committee members did not observe x_1 at all, but *only* observed $s_1 = 1$. Second period agents rationally discount x_1 in this case as containing no information about the signal s_1.

[8] If $x_1 = 1$ then second period agents can perfectly infer that no B type was presence and that the signal was $s_1 = 1$.

At the other extreme, when $\theta = 0$, beliefs are $\mu(0,1) = \pi$. Second period agents perfectly infer that the first signal was $s_1 = -1$. Since the signals s_1 and s_2 offset one another agents are left with their prior belief π. Second period decision making then perfectly aggregates the information contained in both s_1 and s_2.

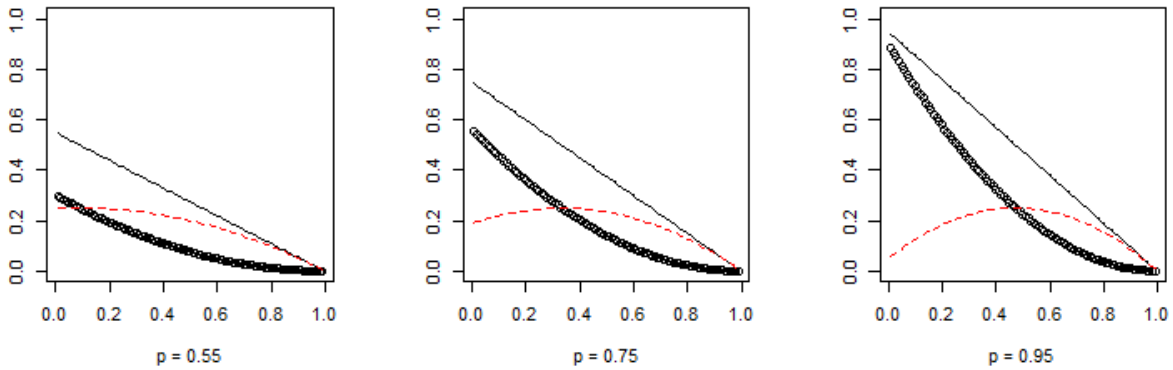
For intermediate values of θ the outcome is partial information transmission. In forming beliefs about the state of the world, agents in the second period weigh the probability that a zero first period outcome reflects the presence of B types versus the probability that it accurately reflects the signal $s_1 = -1$. Beliefs following $x_1 = 0$ will be strictly lower than if the first period outcome were unobserved for any $\theta \in (0,1)$. While second period agents know that x_1 may simply reflect partisan politics they nonetheless find it partially persuasive.

When will this persuasiveness matter for second period decision making? If agents are sufficiently optimistic about the state of the world then they are likely to remain responsive to their own signal even after observing $x_1 = 0$. However, if π is relatively low then after observing $x_1 = 0$ second period agents may prefer to vote for $x_2 = 0$ regardless of the signal they themselves observe. The following result establishes this formally.

Proposition 2. *There exists a $\tilde{\pi}$ such that when $\tilde{\pi} > \pi$ the following holds: if the first period outcome is zero, then the second period outcome is also surely zero.*

Note that if the same agents simply did not observe the first period outcome at all, then assumption A1 implies that their optimal outcome would surely be responsive to the signal. This suggests that in some cases, agents are made worse off by observing the outcome x_1 relative to observing only their own signal. For example let $\tilde{\pi} > \pi$ and suppose that the true state of the world is one. Consider the probability that either committee chooses the "correct" policy $x_t = 1$. In the first period this probability is $(1-\theta)p$ reflecting the probability that no B type is present and a signal of $s_1 = 1$ is observed. In the second period, the probability is $[(1-\theta)p]^2$, reflecting the fact that $x_2 = 1$ can only obtain if $x_1 = 1$. Otherwise, when $x_1 = 0$, with probability one $x_2 = 0$ as well.

Figure 1



If instead the second period committee does not observe the first period outcome the probability of $x_2 = 1$ is again $(1-\theta)p > [(1-\theta)p]^2$. Thus the probability that a committee chooses the correct policy is decreasing over time. Additionally the expected welfare of type A agents is strictly lower when they have the opportunity to observe x_1 than if they observe only their own signal. Figure 1 depicts this efficiency loss for several values of p. In each panel the solid line indicates the probability that $x_2 = 1$ if x_1 is unobserved. The dotted line indicates the probability that $x_2 = 1$ given that x_1 is observable. The difference between these two probabilities, corresponding to the expected utility loss for a type A player is depicted in red.

3.2 Endogenous Participation

Next, I consider how the influence of precedent shapes agent incentives to participate in committee decision making over time. Let there be a set of agents, $N_t = \{1, ..., n\}$, in each period who are eligible to participate in decision making upon paying fixed cost, $\phi_o > 0$ for $o \in \{A, B\}$. We assume that there is no overlap between the sets, $N_1 \cap N_2 = \emptyset$, so that there is no direct channel of information sharing across the two committees. Let the number of B type agents in each period be common knowledge and given by $b \in \{1, ..., n\}$. Each agent

has a salience parameter, ν_i, distributed uniformly on $[0,1]$. Salience parameters are private information. Payoffs for each type are now given by,

$$u_i(x_1, x_2) = \begin{cases} \omega(x_1 + x_2)\nu_i & \text{if } i \text{ is type } A \\ -(x_1 + x_2)\nu_i & \text{if } i \text{ is type } B \end{cases}$$

The realized number of participants in decision making each period, k_t, is unobservable to agents not in C_t. I make the following assumptions:

A2. $\mu(1) > \phi_o$ for $o \in \{A, B\}$

A3. $\hat{\pi} > \pi$ for some $\hat{\pi} \in (1-p, \frac{1}{2})$

Assumption A2 ensures that participation costs are low relative to an A type's expected gain from $x_t = 1$ given that she is the highest type and the signal $s_t = 1$ is observed. Assumption A3 ensures that π is sufficiently low that precedent may emerge on the equilibrium path. A **participation game** is $\Gamma^p = (G, \pi, p, N, b, \phi_A, \phi_B)$. Equilibrium strategies now consist of a participation rule and a voting rule for each type in each period. The next result characterizes optimal behavior in the participation game.

Proposition 3. Γ^p *admits a unique pure strategy Perfect Bayesian Equilibrium:*

— *In* $t = 1$,

- *An agent of type* o *joins the committee if and only if* $\nu_i \geq c_1^o$,
- *An agent of type* A *votes in favor of* $x_1 = 1$ *if and only if* $s_1 = 1$,

— *In* $t = 2$,

- *An agent of type* o *joins the committee if and only if* $\nu_i \geq c_{2,x_1}^o$, *for* $x_1 \in \{0, 1\}$,
- *An agent of type* A *votes in favor of* $x_2 = 1$ *if and only if* $\mu_t \geq \frac{1}{2}$,

— *In* $t = 1, 2$, *committee members of type* B *vote in favor of* $x_t = 0$.

— *The cutpoints* $\boldsymbol{c} \equiv (c_1^A, c_1^B, c_{2,0}^A, c_{2,0}^B, c_{2,1}^A, c_{2,1}^B)$ *are unique.*

Optimal participation strategies take the form of cutpoints such that an agent will participate in decision making if and only if ν_i is weakly greater than that threshold. In each case, the cutpoint is defined by the salience type that is exactly indifferent between participating and not participating. In the first period cutpoints differ by type (technocrat vs. partisan). In the second period cutpoints differ both by type and by the prior outcome x_1. The final result establishes the core intuition of the paper, that greater politicization is expected where long-run precedents are at stake.

Proposition 4. *There exists a* $\bar{\phi}_B$ *such that when* $\bar{\phi}_B > \phi_B$ *expected participation is decreasing over time.*

Note that given $\tilde{\pi} > \pi$, after an outcome $x_1 = 0$ all second period agents prefer that the policy $x_2 = 0$ be implemented. Since this is also the default policy, no agent of either type has incentive to incur cost ϕ_o. Thus $c_{2,0}^A = c_{2,0}^B = 1$, and the probability that any agent participates is zero. In other words, following an outcome of $x_1 = 0$ agents of both types agree to follow prior precedent despite the reality that x_1 may convey no information at all about the state of the world.

Very different results obtain following an outcome of $x_1 = 1$. In this case the preferences of technocrats remain responsive to the signal, s_2. Yet conditional on pivotality, a technocrat's expected gain from participation is strictly increasing over time reflecting growing confidence in the likelihood that $\omega = 1$. In many, though not all, cases higher expected participation by technocrats will lead in turn to higher participation by partisans. In other words, the outcome $x_1 = 1$ may *increase* politicization as technocrats become more certain that their preferred policy diverges from that of partisans. Nonetheless, the probability that $x_1 = 1$ is strictly decreasing in ϕ_B, the partisans' cost of participation. As a result, when costs are moderate $x_1 = 1$ is relatively unlikely and expected participation will be strictly decreasing over time.

4 Empirical Strategy

In the following sections I describe my empirical strategy for testing the model's implications in the context of WTO accession negotiations. I first describe the substantive context of accession negotiations and draw implications from the model for participation in accession committees over time. Next I discuss potential alternative explanations for participation before describing the data and variables used in the regression analysis.

4.1 WTO Accession Negotiations

Since its creation in 1995, thirty six states have requested and been granted membership in the WTO. Another twenty one are currently negotiating entry. All together over a quarter of the states in the world have acceded or will accede to the WTO via the procedure described below. These states account for thirty percent of the world's population though just twenty percent of its gross domestic product, reflecting the high proportion of low income countries who have joined or who will join in the coming years. Just ten of the thirty six states that have completed accession to the multilateral trade regime to date were classified as high or upper-middle income countries at the time of accession.[9]

The sole legal provision governing accession to the WTO is extraordinarily broad, noting only that applicants may join the organization "on terms to be agreed between it and the WTO" (Marrakesh Agreement Establishing the World Trade Organization, 1994). In practice, these terms consist of two distinct components: a package of policy commitments which an applicant will observe upon accession to the organization and a set of specific market access concessions in goods and services. Neither the depth nor the design of these terms is restricted by the obligations of existing WTO members. In fact the terms of WTO accession for new members are typically far more stringent than those adopted by the broader

[9] In contrast ten countries were classified as low income, and an additional sixteen as lower-middle income at the time of accession.

membership at the close of the Uruguay Round (Evenett and Primo Braga, 2005).

Accessions are negotiated on a case-by-case basis giving negotiators leeway to tailor accession terms to the specific circumstances of each applicant. Market access concessions vary in the average depth of tariff cuts as well as in the total number of tariff bindings. Policy commitments vary across applicants both in number and substance.[10] Many commitments reiterate compliance with existing WTO agreements, while others specifiy novel disciplines on policy making ("WTO plus") or relinquish rights previously enshrined in the multilateral trade regime ("WTO minus"). Applicants may also negotiate transition periods during which policy reforms are phased in over a period of years following accession. The number and duration of these transition periods varies across applicants.[11]

The predominance of low- or lower-middle income states among WTO applicants means that accession negotiators must navigate one of the most contentious questions in contemporary trade politics: how to design multilateral rules to support trade liberalization while respecting the developmental circumstances of member states. Trade liberalization is widely accepted in academic and policy circles as a central component of economic development. Yet the adjustment costs associated with trade reform are particularly high in developing countries where economic activity may be concentrated in a small number of sectors and where there is little state capacity to assist those who lose out from globalization. As a result some call for a cautious approach to trade liberalization in the developing world (WTO, 2006; Rodrik, 2012).

The appropriate role of the multilateral trade regime in bringing about liberalization is particularly controversial. Existing global trade rules are widely viewed as imbalanced, disportionately protecting those industries such as agriculture and textiles in which developing

[10]The average number of policy commitments is 42 though with a standard deviation of 33.

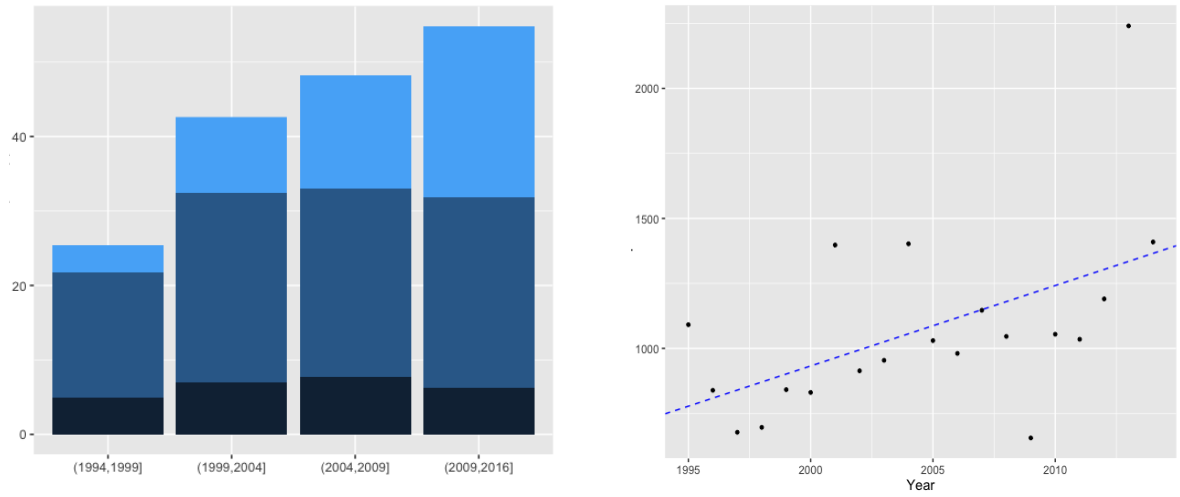
[11]The range and depth of these negotiations places considerable demands on the capacity of most applicants. During the ten years of its accession negotations Saudi Arabia conducted 365 rounds of bilateral and 14 rounds of multilateral negotiations, submitted around $7,600$ pages of documentation and issued over 70 orders, laws, and regulations in order to meet the demands of existing members (Evenett and Primo Braga, 2005).

country exports are most competitive, leading to skepticism about the institution's ability to constrain the partisan interests of its most powerful members. During the most recent round of multilateral trade negotiations, developing countries rejected many initiatives on the grounds that proposed changes to the global trade regime would unnecessarily constrain policy making space. Developing countries argue in favor of a more flexible approach to government intervention to correct market failures inherent in the early stages of economic growth, noting that developed economies have historically benefitted from similar policies (Gallagher, 2007).

These questions about trade liberalization and the optimal role of the WTO in developing economies are technocratic in nature, concerning the most effective means of encouraging economic growth in the developing world without imposing undue hardship. But as described in the case of services negotiations above, any resolution to these questions is complicated by the competing mercantilist incentives of the WTO's membership. For example, during the Doha Round the United States sought increased market access in developing economies to balance the domestic political costs of agricultural reform at the center of the negotiating agenda (USDA, 2006; Beattie, 2007). Against this backdrop of uncertainty, accession negotiations can provide valuable information about the optimal design of multilateral commitments in the developing world.

It is not surprising then that the contours of accession negotiations increasingly reflect the broader agenda of multilateral trade negotiations. The policy commitments negotiated during accessions often have direct bearing on key issues of contention among trade experts. For example, one WTO-plus provision adopted by the Seychelles during its accession requires the establishment of a domestic competition policy. The role of the WTO in regulating competition policy is controversial. While initially included in the Doha Round's negotiating mandate as one of four "Singapore issues," it was dropped from the work program in 2004 amid opposition from developing countries.

Figure 2: Growth in Number and Diversity of Accession Commitments



(a) Mean Commitments by Applicant (b) Unique Words Used Over Time

(a) Light blue indicates proportion of commitments relating to: TBT, SPS, Services, TRIPS, or Framework for Making and Enforcing Policies. Dark blue indicates proportion of commitments relating to: Customs Tariff, Quantitative Import Restrictions, Trade Remedies, Internal Taxes, or Export Subsidies. (b) Unique words contained in accession negotiating documents by year, 1995-2014. Scaled by total number of applicants negotiating in each year. Best fit line obtained via linear regression.

More generally, recent years have witnessed a dramatic growth in both the average number of commitments and the breadth of commitments across issue areas made by applicants to the WTO. The left-hand panel of Figure 2 depicts a sharp upward trend in the average number of commitments per applicant over the past twenty years. Many of these new policy commitments are related to the most contested issues in contemporary trade politics. This includes commitments related to an applicant's framework for making and enforcing policies, services, Trade Related Aspects of Intellectual Property Rights (TRIPS), Technical Barriers to Trade (TBT) and Sanitary and Phytosanitary measures (SPS).[12]

During the first five years of WTO accession negotiations, these five issues accounted for 15% of all accession commitments. In the most recent period, they have grown to account for 41%. The share of commitments related to these issues is shown in the lightest shaded

[12]The first of these relates most clearly to the growing emphasis on "behind the border" trade measures. Services and IP were key negotiating topics during the Doha Round. Measures related to TBT and SPS are widely seen as the cutting edge of trade negotiations and are being contested via the DSU at a growing rate.

regions of panel (a) of Figure 2. In contrast, the share of accession commitments related to more traditional trade issues has for the most part remained constant.

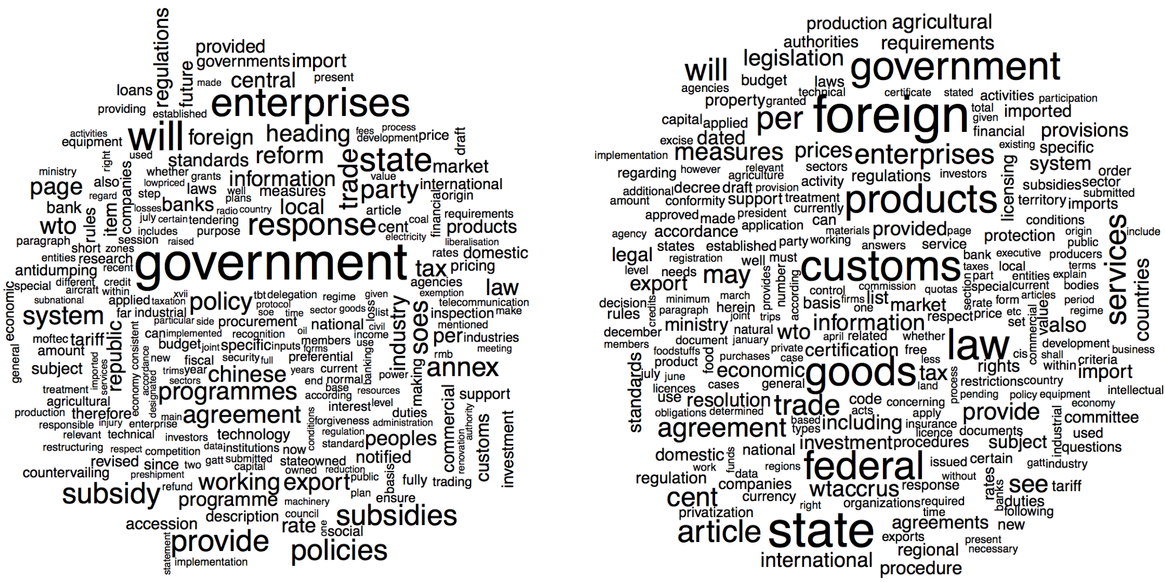
At the same time focusing on such fixed categorizations conceals the growing *diversity* of topics now subject to accession negotiations. To get a sense of the latter, the right-hand panel of Figure 2 depicts the average number of unique words appearing in a corpus of accession-related negotiating documents over time.[13] As shown in the figure, this exhibits a clear upward trend indicating that the number of subjects raised in the course of negotiations has grown consistently over time.

The impact of accession commitments on the consensus of trade policy experts can be seen directly in the path dependency of WTO-plus and -minus provisions negotiated during accessions. As Kennedy notes, the accession commitments of new members "tend to build on those of earlier accessions and [now] represent a substantial body of WTO law" Kennedy (2013). Accession commitments proscribing the use of export duties provide one example. In its accession protocol, China agreed to "eliminate all taxes and charges applied to exports" except for those specifically provided for in a separate annex (WT/ACC/CHN/49, 2001). Up until this point few members had made specific commitments proscribing the use of export duties. Yet following China's accession a number of subsequent applicants, including Ukraine, Montenegro, and Tajikistan, made similar commitments.

As in other areas of international cooperation, the negotiation of these accession terms requires that members engage in an extensive process of information gathering. The accession process itself includes a formal period of information gathering consisting of sequential rounds of fact-finding. In each round members submit written questions to the WTO Secretariat who subsequently transmits them to the applicant. The applicant's responses are then circulated among the membership. These questions and replies, as they are known, cover all aspects of

[13]This corpus is described in more detail below. In each case the number of words is scaled by the total number of concurrant accession negotiations under way in that year. Thus the figure depicts the average number of unique words per applicant by year.

Figure 3: Frequently Used Words in Accession Negotiations



(a) China

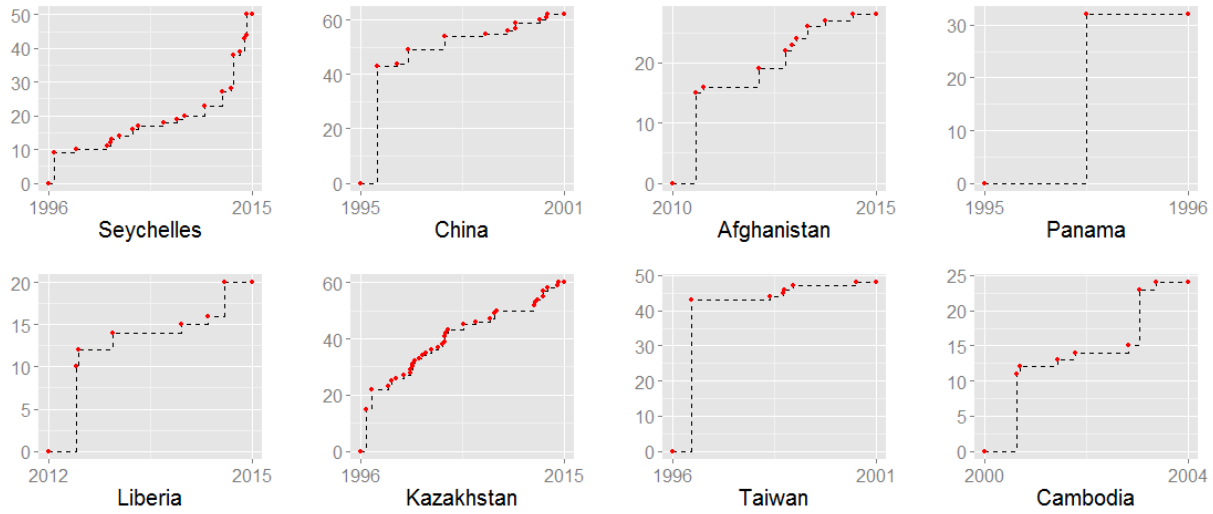
(b) Russia

an applicant's trade regime and evolve over the course of accession negotiations in response to the primary concerns of existing members or the revelation of new information in prior rounds.

Figure 3 depicts the most frequently-appearing words in all questions and replies pertaining to two high-profile accessions, those of Russia and China respectively. In the case of China, the most prominent words include "government," "subsidy," and "enterprise" reflecting concerns of the WTO membership about the active role of the Chinese government in its economy. In the case of Russia, the most prominent terms are related to more basic WTO obligations ("products," "goods," and "customs") indicating that Russia's accession negotiations were focused on more traditional barriers to trade.

Thus accession negotiations exhibit many of the key characteristics - uncertainty, information gathering, partisan interests - of international economic negotiations more broadly. As such they provide a novel testing ground for the implications of the argument.

Figure 4: Working Party Membership Over Time



Size of working party over time for selected WTO applicants. First observation represents initial formation of accession working party.

4.2 Participation in Accession Working Parties

As described above, accession negotiations provide a useful venue in which to test the model's implications in part because the membership structure of accession working parties provides a novel indicator of how interest in negotiations changes in response to the policies under negotiation. Accession negotiations are carried out by members of an ad hoc standing committee, the accession working party. Participation in each accession working party is open to all existing WTO members. There is no upper or lower limit on the number of countries that participate, and members may choose to join the working party at any point in the course of negotiations prior to the applicant's final accession. As depicted in Figure 4 membership in accession working parties varies significantly over time and across applicants.

In general, membership in an accession working party can be considered a costly signal of interest in negotiations, reflecting the significant demands that participation in any of the WTO's various committees and working parties places on the resources of an average member (Busch et al., 2009). Overall the WTO's thirty two permanent councils and committees alone held an average of nineteen meetings per month in 2016. Participation in the work of these

various bodies requires not only attendance at meetings, but also detailed preparatory work given the technical nature of the issues at stake. Yet in 2009, the last year for which data is available, member economies assigned, on average, fewer than six diplomats to represent their interests at the WTO across all substantive areas (Allee et al., N.d.). Participation in accession negotiations requires the dedication of scarce human resources and thus entails an opportunity cost.

The model makes two main predictions regarding the participation of WTO Members in accession negotiations. First, in equilibrium Members who place higher salience on the issue under negotiation will be more likely to participate at any point in negotiations (Proposition 3). To test this, I use members' own past behavior as a guide to the negotiating subjects which are most relevant to their interests. If members have previously participated in accession negotiations in which a certain topic has featured prominantly, this suggests that the topic is one over which they hold strong preferences. Thus members' reveal their preferences over negotiating topics through their previous participation in accession negotiations.[14]

H1. Members are more likely to join a working party when the substance of negotiations is similar to that of negotiations in which the member has previously participated.

Proposition 4 above establishes that, all else equal, members should be more likely to participate early on in negotiations over any given policy in order to have the greatest impact on future decision making. I test this directly by exploring how member states react to the emergence of topics not previously subject to accession negotiations. The outcome of negotiations will have the largest impact on future policy choices when it concerns relatively novel issues.[15]

H2. Members are more likely to join a working party when negotiations touch on subjects

[14]This may be the case because states self-select into negotiations where these topics are under debate or because the member itself has raised the topics in the accessions in which it previously participated.

[15]Note that neither of these hypotheses distinguish which members of the organization are expected to behave as technocrat or partisan types. As discussed above, the majority of member states should be expected to display characteristics of both according to the particular setting of negotiations.

which are systemically novel.

I next consider several potential alternative explanations for members' participation in accession working parties.

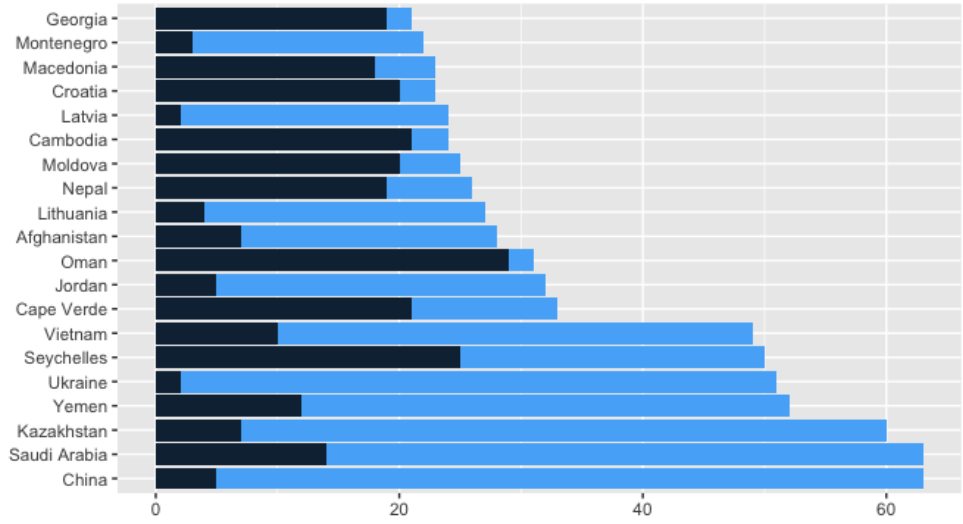
4.3 Alternative Explanations

The clearest alternative explanation for participation in accession negotiations is direct economic self-interest. Members seeking to extract particular tariff or service sector concessions must join the accession working party as a precursor to engaging in bilateral negotiations (Neumayer, 2013). Existing literature finds that the value of tariff concessions demanded from an applicant reflects the overall size of its import market, supporting the idea that a key objective of negotiators is to secure valuable export opportunities for existing members (Pelc, 2011). Yet there are reasons to doubt that participation in these negotiations is driven only, or even primarily, by the desire to extract concessions.

Institutional features of accession negotiations coupled with the capacity constraints described above provide members with strong incentive to free-ride on the negotiating efforts of others. The WTO's Most Favored Nation principle requires that any concessions granted by the applicant in the course of accession apply not only to the individual member who negotiated the concession but also to all existing members of the organization. This multilateralization of concessions means that states may rightly expect to enjoy the benefits of negotiations without investing their own precious time or resources (Accominotti and Flandreau, 2008). The United States makes a particularly attractive target for free-riding given its unparalleled technical capacity and vigorous participation in every accession negotiation.

This potential for free-riding appears to be well-recognized by developing states themselves. As one Moroccan negotiator explained during a regional workshop on WTO accession, "Morocco has never made any demand on developing countries in their stage of acces-

Figure 5: Distribution of Working Party Membership



Number of working party members for selected WTO applicants, 1995-2016. Members with positive reported trade shown in dark blue. Members with no reported trade shown in light blue (Source: UN COMTRADE).

sion...industrialized countries are doing a much better job than we can do! (Achy, 2004)." Nonetheless, Morocco has participated as a member of the working party in nearly a third of all accessions completed to date. This pattern of countries participating in working parties yet not pursuing economic concessions is widespread.[16] It is also consistent with the fact that many working party members have limited or no trade ties with the applicant. Figure 5 depicts the proportion of working party members with positive bilateral trade during the five years preceding WTO accession. While this describes a majority of working party members in a few cases, in many more the applicant's trade partners make up only a small minority of the working party's membership.

Nonetheless I control for export interests in my analysis below. An alternative channel through which direct economic self-interest may influence working party participation is through fear of export competition. Members may fear the accession of export competitors

[16] Anecdotal evidence suggests that of the 38 countries who joined the working party on the accession of Tajikistan, only 13 engaged in direct bilateral negotiations over tariff concessions (World Trade Organization, 2012a). The working party on the accession of Laos attracted 66 members in total, yet only nine of those pursued bilateral tariff negotiations (World Trade Organization, 2012b).

who can use their newfound market access to undermine the member's exports to other WTO members. Members of an accession working party can indefinitely delay or obstruct a competitor's accession potentially creating strong incentives to participate (Neumayer, 2013).

Alternatively political factors may also play a role in motivating participation. Davis and Wilf (2017) and Copelovitch and Ohls (2011) argue that a prospective applicant's geopolitical orientation and regime type are key determinants of the decision to join the GATT/WTO and of the duration of accession negotiations.

Finally, the decision to participate in WTO accession negotiations may reflect a desire on the part of the existing member to develop its knowledge of trade policy and, relatedly, its negotiating capacity. Existing literature argues that members are far more likely to pursue cases under the Dispute Settlement Understanding if they have already participated in the past, suggesting a "learning by doing" approach to participation (Davis and Bermeo, 2009). In the context of accession negotiations, this may be a particularly appealing strategy for those members who have recently acceded themselves and are thus relatively new to the multilateral trading system.

4.4 Estimation of Key Variables

To characterize the substance of negotiations over the course of a country's accession I employ automated text analysis using a corpus of declassified accession documents. I use Latent Dirichlet Allocation (LDA) to estimate the subjects which recur frequently in these documents and thus in the course of accession negotiations. LDA is a commonly-used topic model, a statistical tool which estimates unobserved clusters of language (topics) based on their co-occurrence within a body of text (Blei et al., 2003). In addition to estimating the topics contained within the corpus as a whole, LDA estimates the distribution of these topics within each individual document. This document-level distribution has a direct substantive

interpretation: it summarizes the various subjects under negotiation at any given moment as well as their relative prominence in those negotiations.

The hypotheses above make claims about the similarity or dissimilarity of subjects under negotiations. Thus my independent variable consists of a measure of the similarity or dissimilarity of these document-level distributions. First, for each document and each potential working party member I calculate the Kullback-Liebler (KL) divergence between the document and all prior documents related to accessions in which the potential member has previously participated. This quantity represents the substantive (dis)similarity of negotiations to those in which the potential member has previously revealed an interest. In line with H1 I predict that the probability of participation will be increasing in similarity, thus decreasing in the magnitude of the KL divergence.

Second, I estimate the KL divergence between each document and all prior documents across all previous accessions. The resulting quantity represents a measure of the substantive (dis)similarity of negotiations relative to all previous negotiations. In line with H2 I predict that the probability of participation for all members will be increasing in the magnitude of this divergence. I describe the calculation of these quantities in more detail below.

The corpus employed for the analysis comprises a novel collection of accession negotiating documents. In particular, I collect all questions and replies circulated in the course of accession negotiations completed before August 1, 2016.[17] As described above these documents are structured into individual questions submitted by working party members and written responses from the applicant country. Each of these questions is relatively short, with a mean length of 101 words, and concentrates on a particular aspect of the applicant's trade regime. To aid in the estimation of topics I analyze each question and corresponding response as a separate "document." The final corpus is composed of 326 unique documents yielding $30,062$ questions. These questions detail the accessions of 33 of the 36 countries that have joined

[17] I analyze both parent documents and documents labeled addenda provided that the latter are structured into questions and replies as described in the text.

the WTO since 1995.[18] Table 1 describes the number of documents and individual questions associated with each applicant.

Table 1: Documents and Questions by Applicant

Applicant	Documents	Questions	Applicant	Documents	Questions
Afghanistan	10	775	Mongolia	1	8
Albania	15	349	Montenegro	13	971
Armenia	1	127	Nepal	2	465
Cambodia	5	458	Oman	8	803
Cape Verde	10	899	Panama	3	226
China	2	73	Russia	20	2284
Croatia	14	915	Samoa	11	898
Estonia	6	425	Saudi Arabia	16	1199
Georgia	5	508	Seychelles	9	1057
Jordan	6	917	Taiwan	4	216
Kazakstan	22	1601	Tajikistan	13	1380
Lao PDR	14	1237	Tonga	6	410
Latvia	4	275	Ukraine	36	3392
Liberia	2	155	Vanuatu	3	488
Lithuania	8	634	Vietnam	20	3385
Macedonia	6	823	Yemen	13	1154
Moldova	8	854			

Using this question-level corpus I estimate a one hundred topic LDA.[19] To obtain an aggregate distribution over topics corresponding to a particular date, I average across the estimated topic proportions for all questions associated with that date. As described above, I employ the Kullback-Leibler divergence to characterize the similarity of topic distributions at various points in time. The KL divergence is a measure of information content commonly used in natural language processing and provides a measure of the extent to which two probability distributions differ. Formally, for two word frequency distributions, P and Q, the KL divergence from Q to P is:

$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)} \tag{3}$$

[18]No questions and replies are available for Bulgaria, Ecuador or the Kyrgyz Republic, three of the first countries to accede to the WTO.

[19]Topics estimated using the topicmodels in package in R.

In the present context, P represents the distribution of word frequencies in a particular document, while Q represents the word frequency distribution across a set of prior documents.

Consider a set of I applicants, $i \in \{1, ..., I\}$, J potential members, $j \in \{1, ..., J\}$, and T time periods $t \in \{1, ..., T\}$. Let $I(j,t)$ denote the set of applicants of whose working party country j is already a member at time t. The vector of topic proportions for applicant i at time t is $\boldsymbol{\theta}_{i,t}$. To test the first hypothesis I construct a measure `KL` consisting of the KL divergence from the distribution of topics at time t to the distribution of topics across all previous dates in all accessions in which the member has previously participated. Denote the former distribution by $P_{i,t} = \boldsymbol{\theta}_{i,t}$ and the latter by,

$$Q_{j,t} \propto \sum_{i' \in I(j,t)} \sum_{t' < t} \boldsymbol{\theta}_{i',t'}$$

The first independent variable is defined as,

$$\texttt{KL} = D_{KL}(P_{i,t} || Q_{j,t})$$

To test the second hypothesis I construct an additional measure `KL_All` consisting of the KL divergence from the distribution of topics in each document and the distribution over all prior documents (across all applicants). As before, let the former distribution be denoted by $P_{i,t}$ and the latter by,

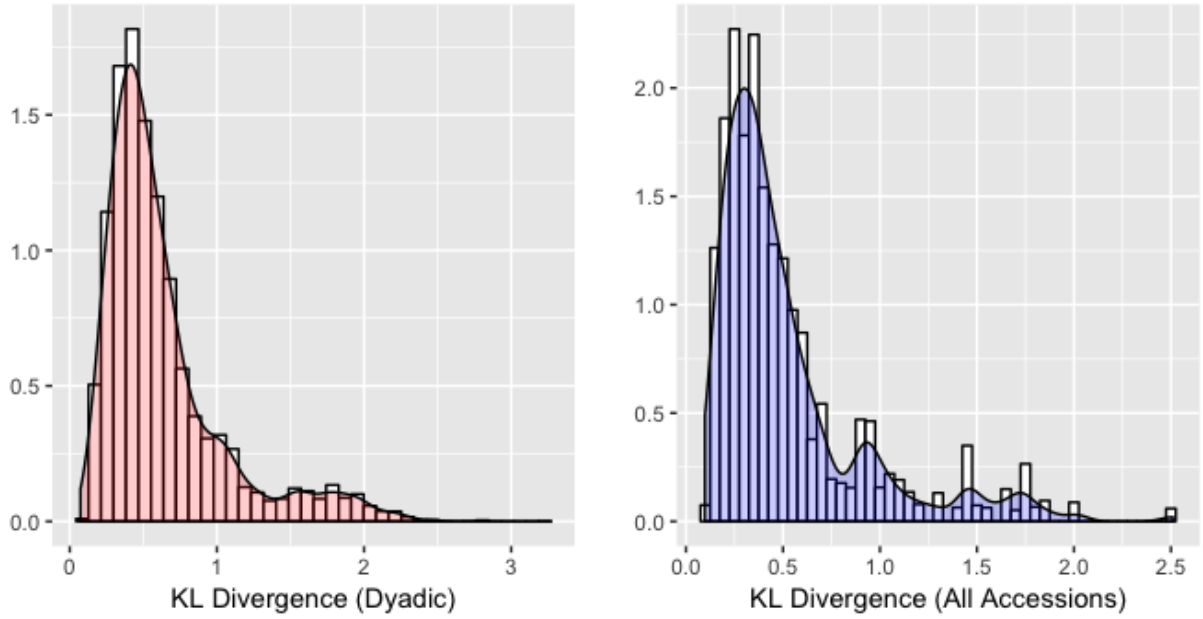
$$S_t \propto \sum_{i'} \sum_{t' < t} \boldsymbol{\theta}_{i',t'}$$

The second independent variable is then defined as,

$$\texttt{KL\_All} = D_{KL}(P_{i,t} || S_t)$$

The distribution of both variables is depicted in Figure 6. Both are skewed left with 95% of observations falling below 1.9 and 1.5 respectively. To ensure that results are not driven by a

Figure 6: Distribution of Key Variables



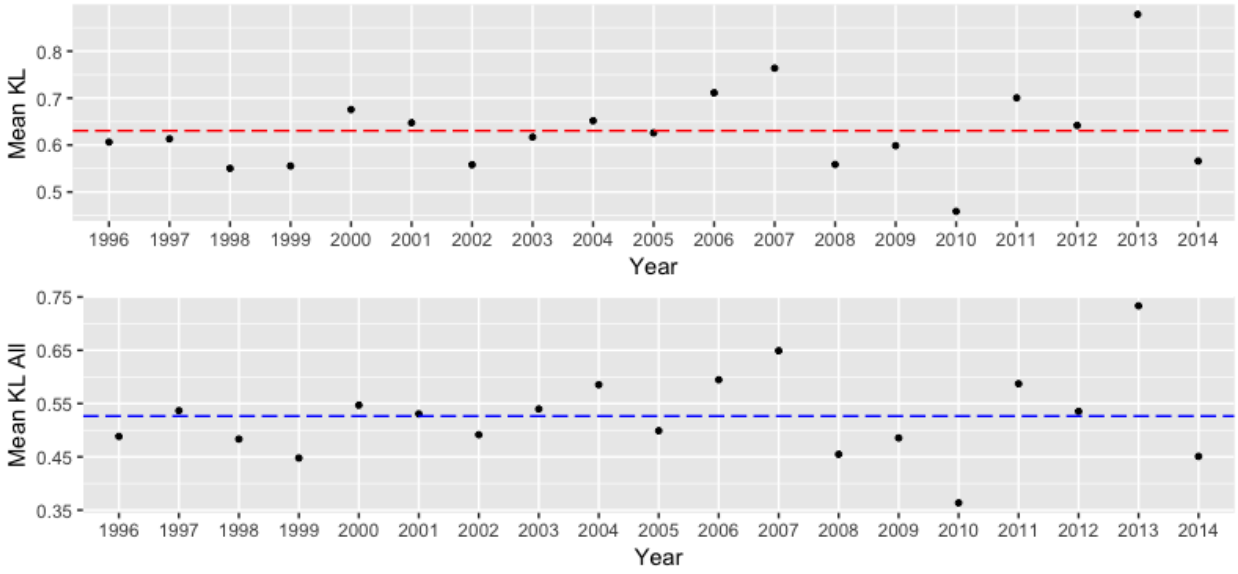
small number of outliers in the regression analysis below I employ the log of both variables.[20]

One additional concern with the independent variables is that the KL divergence may be sensitive to the number of documents released prior to time t. That is, during the early years of the WTO when few negotiations had taken place it is possible that any topic would appear to be comparatively novel. It's not clear that this would necessarily undermine the results of the analysis below. Afterall what is identified as novel in the data would also have appeared novel to members at the time suggesting that the same logic should apply. Nonetheless I drop the first year in the sample, 1995. Figure 7 plots the mean of both independent variables across the remaining years. This mean is clearly not monotonic across time, nor do early observations appear to be systematically more novel than later observations.[21] Figure 8 in the Appendix plots the mean of `KL_All` by country.

[20]Results are not sensitive to employing the original variables or to dropping the top five percent of outliers instead.

[21]Dropping 1996 from the sample also does not change the results below.

Figure 7: Mean of Key Variables by Year



4.5 Additional Data and Variables

As a dependent variable I employ a dichotomous measure `Joins6` which takes a value of 1 if potential member j joins applicant i's working party in the six months following date t (or prior to the next round of questions and replies if this follows within six months of t).[22] My primary measures of economic self-interest are the log of applicant i's imports from and exports to potential member j, `lImportsIJ` and `lExportsIJ`.[23],[24] I also include a count of the number of members who join a working party immediately following its formation, `OriginalMembers`. This provides a proxy of the *ex ante* level of economic interest in the applicant's accession before substantive negotiations are under way.

To account for strategic delay I construct measures of export similarity between applicant i and potential member j following Neumayer (2013). In particular I define `ProductSimilarityIJ`

[22]Employing a three month or nine month window following date t instead does not alter the results.
[23]This follows the specification in Pelc (2011).
[24]Trade data for these and the competition variables below comes from the UN COMTRADE database.

as:

$$\sum_{k} Min[X^k_{i,t}, X^k_{j,t}] \tag{4}$$

where k is a product group and $X^k_{i,t}$ and $X^k_{j,t}$ are the share of exports of product k relative to total exports in year t for applicant i and potential working party member j respectively.[25] The resulting measure ranges from zero to one with higher values indicating greater similarity in export profile. I also calculate a measure of export market similarity again defined as in equation 4, but letting X^k_{it} (X^k_{jt}) denote the share of country i's (j's) total exports bound for country k in year t. The final measure of export similarity is defined as

$$\texttt{TotalSimilarityIJ} = \texttt{ProductSimilarityIJ} \times \texttt{ExportMarketSimilarityIJ}.$$

To account for diplomatic capacity constraints I include a count of the number of working parties up to date t in which member j has previously participated, `MembershipsToDate`, and the size of j's diplomatic mission in Geneva, `MissionSize` (Allee et al., N.d.).[26] To measure regime type I employ polity IV scores for both applicant and potential member, `PolityI` and `PolityJ` (Davis and Wilf, 2017; Pelc, 2011; Copelovitch and Ohls, 2011). My measures of geopolitical alignment consist of UN ideal points, `IdealPointI` and `IdealPointJ`, estimated in Bailey et al. (2015). To account for learning as a motivation for participation I construct an additional dichotomous variable `RAM` which takes a value of 1 if potential member j has joined the WTO since 1995.

All specifications include (logged) GDP and GDP per capita for both applicants and potential members, a dichotomous variable equal to 1 if a PTA exists between the two, and a dichotomous variable, `Quad`, equal to 1 if the potential member is one of the United States, Canada, or Japan.[27],[28] This last variable reflects the historical role of quadrilateral negotia-

[25] I define product groups at the two digit HS level to minimize missingness.

[26] I am especially grateful to Manfred Elsig for sharing the data on members' mission size.

[27] GDP and population data comes from the Penn World Tables. PTA data comes from the DESTA dataset as described in Dür et al. (2013).

[28] Members of the European Union are excluded from the dataset altogether since they participate primarily

tions at the GATT/WTO.[29] Descriptive statistics for all variables are included in Table 4 of Appendix C.

5 Regression Analysis

I begin by testing both hypotheses in a baseline model, including only a set of basic covariates. I employ logistic regression due to the dichotomous nature of the dependent variable, `Joins6`, and cluster standard errors at the document level. Covariates include (log) GDP and GDP per capita for both applicant and potential working party member, `OriginalMembership`, `MembershipsToDate`, `PTA`, and `Quad`. The results of this analysis are included in the first column of Table 2.

I find a negative and statistically significant association between `KL` and `Joins6` indicating that member j's probability of joining a working party is decreasing in the divergence between the subjects under discussion and those subjects in which j has previously revealed an interest. The relationship between `KL_All` and `Joins6` is positive and also statistically significant. Thus WTO members are more likely to participate in accession negotiations when they touch on subjects which are systemically novel. These finding support both H1 and H2. While I do not report coefficient estimates for the covariates they are for the most part signed as expected.

In the remaining columns of Table 2 I report the results of several specifications including a range of trade-related covariates. I test the role of export interests by first including `lImportsIJ` and `lExportsIJ`. Surprisingly while i's exports to country j appear to be positively associated with j's participation the same is not true for i's imports from j. In the next column I replace these variables with the log of total trade between i and j. The coeffi-

as a bloc.

[29] Additionally all four members of the "quad" have participated in the working party of every accession to date.

Table 2: Main Results

	Model 1	Model 2	Model 3	Model 4
(Intercept)	−10.42*	−6.68*	−8.57*	−6.37*
	(1.00)	(1.10)	(1.07)	(1.13)
(Log) KL	−2.11*	−2.20*	−2.18*	−2.04*
	(0.36)	(0.36)	(0.36)	(0.36)
(Log) KL_All	1.88*	1.96*	1.92*	1.80*
	(0.33)	(0.34)	(0.34)	(0.34)
(Log) $\text{Imports}_{i,j}$		−0.01		
		(0.01)		
(Log) $\text{Exports}_{i,j}$		0.07*		
		(0.01)		
(Log) $\text{TotalTrade}_{i,j}$			0.04*	0.12*
			(0.01)	(0.02)
(Log) TotalImports_i				−0.07*
				(0.01)
(Log) TotalExports_j				0.00
				(0.01)
$\text{TotalTrade}_{i,j}/\text{GDP}_j$				0.00*
				(0.00)
$\text{TotalTrade}_j/\text{GDP}_j$				−0.00*
				(0.00)
N	11097	11097	11097	11097
AIC	3805.18	3738.91	3782.69	3699.40
BIC	4127.02	4119.26	4133.79	4167.52

Logistic regression with standard errors clustered at document level. Dependent variable takes a value of 1 if potential member_j joins working party_i within six months following release of $\text{document}_{i,t}$. * indicates significance at $p < 0.05$. All models include (log) GDP_i, (log) GDP_j, GDPpc_i, GDPpc_j, OriginalMembership, MembershipsToDate, PTA, and Quad.

cient estimate is positive and statistically significant. In the final model I include a range of additional measures intended to capture the extent of trade ties between the applicant and potential working party member as well as trade-related characteristics of each individually. The results suggest that economic interests do play a role in incentivizing working party membership. Across all models though the coefficient estimates on the main variables of interest, `KL` and `KL_All`, remain stable and statistically significant.

Next, I turn to testing the robustness of these results against the alternative explanations described above. Table 3 contains the results of this analysis. In each model I reproduce model 3 from the previous table, while adding variables corresponding to each alternative

explanation one by one. The main variables of interest retain their statistical significance across all models. Additionally the magnitidue of each estimated coefficient remains relatively stable. I find strong support for the role of strategic delay as well as geopolitical factors and learning.

Further robustness checks are included in Table 5 in Appendix C. There I demonstrate that the results are robust to the inclusion of year and applicant fixed effects, to alternative specifications of the dependant variable (using 3 and 9 month windows instead of 6), and to dropping the most extreme values for `KL` and `KL_All` rather than taking the log. Overall, the regression analysis provides strong and consistent support for the theoretical model.

6 Conclusion

I argue that negotiators of international agreements face uncertainty about the outcome of particular architectural choices and thus engage in information gathering as a precursor to bargaining. Outsiders to negotiations cannot perfectly observe the results of this information gathering but can infer the beliefs of negotiators from the final form of international cooperation which results. This inference is imperfect though since the form of international cooperation may alternatively reflect the parochial self-interest of negotiating parties rather than efficiency concerns. Thus the parochial interests of influential states may shape the evolution of beliefs about what forms of cooperation are most efficient. Furthermore, this creates incentives for states to manipulate the formation of agreements in order to shape the evolution of international cooperation to their material advantage.

I formalize these arguments in a model of sequential negotiations and demonstrate the endogenous emergence of path dependency in the decisions of committee members. This path dependence has important distributional implications: by blocking adoption of the optimal policy early on, partisans can manipulate perceptions of future committee members about

Table 3: Alternative Explanations

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	−8.47*	−8.68*	−8.36*	−9.39*	−9.60*	−8.32*
	(1.08)	(1.24)	(1.32)	(1.17)	(1.09)	(1.65)
(Log) KL	−2.14*	−2.62*	−2.61*	−2.08*	−2.10*	−2.93*
	(0.36)	(0.41)	(0.40)	(0.39)	(0.37)	(0.46)
(Log) KL_All	1.89*	2.30*	2.34*	1.89*	1.84*	2.66*
	(0.34)	(0.38)	(0.37)	(0.36)	(0.34)	(0.43)
(Log) $\text{TotalTrade}_{i,j}$	0.02*	0.03*	0.04*	0.04*	0.04*	0.03*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\text{TotalSimilarity}_{i,j}$	3.77*					2.22
	(1.03)					(1.22)
MissionSize		−0.04*				−0.02
		(0.02)				(0.02)
Polity_i			0.00			−0.03
			(0.01)			(0.01)
Polity_j			0.04*			0.00
			(0.01)			(0.01)
IdealPoint_i				0.28*		0.40*
				(0.07)		(0.09)
IdealPoint_j				0.33*		0.32*
				(0.06)		(0.08)
RAM					0.71*	0.17
					(0.14)	(0.20)
N	11056	9314	9578	10281	11097	8220
AIC	3760.00	3228.09	3296.42	3462.74	3762.76	2753.09
BIC	4140.15	3599.33	3697.78	3868.08	4143.11	3286.18

Logistic regression with standard errors clustered at document level. Dependent variable takes a value of 1 if potential member_j joins working party_i within six months following release of $\text{document}_{i,t}$. * indicates significance at $p < 0.05$. All models include (log) GDP_i, (log) GDP_j, GDPpc_i, GDPpc_j, OriginalMembership, MembershipsToDate, PTA, and Quad.

the true state of the world. This creates powerful incentives for agents to participate in negotiations in order to influence the evolution of policy making more broadly. I test these claims in the context of WTO accession negotiations and find consistent support for the argument. WTO members are more likely to participate in accession negotiations when they expect these negotiations to have valuable long-term influence on subsequent cooperation.

This argument has implications for literature on the design of international institutions and the role of epistemic communities in shaping international cooperation. It emphasizes the challenge of uncertainty in the design of international institutions and how the resolution

of this uncertainty is inevitably shaped by parochial self-interest. Future work will explore how the evolution of learning is shaped by the presence of partisan interests in other areas of international cooperation, particularly in the delegation of enforcement authority to an international court which may yet be influenced by the parochial interests of powerful member states (Brutger and Morse, 2015).

References

Olivier Accominotti and Marc Flandreau. Bilateral treaties and the most-favored-nation clause the myth of trade liberalization in the nineteenth century. *World Politics*, 60(2), 2008.

Emanuel Adler. The emergence of cooperation: national epistemic communities and the international evolution of the idea of nuclear arms control. *International Organization*, 46 (01):101, 1992. doi: 10.1017/s0020818300001466.

Todd Allee, Manfred Elsig, and Andrew Lugg. The ties between the world trade organization and preferential trade agreements: A textual analysis. *Journal of International Economic Law*, N.d.

Wolfgang Alschner. Mapping the universe of international investment agreements. *SSRN Electronic Journal*, 2016. doi: 10.2139/ssrn.2801608.

Michael Bailey, Anton Strezhnev, and Erik Voeten. Estimating dynamic state preferences from united nations voting data. *Journal of Conflict Resolution*, 2015.

Alan Beattie. India calls for us and eu concessions on doha. *Financial Times*, Jan 2007. URL `https://www.ft.com/content/a8b3c8be-abca-11db-a0ed-0000779e2340`.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 2003.

Ryan Brutger and Julia C. Morse. Balancing law and politics: Judicial incentives in wto dispute settlement. *The Review of International Organizations*, 10(2):179–205, 2015. doi: 10.1007/s11558-015-9216-x.

Marc L Busch, Eric Reinhardt, and Gregory Shaffer. Does legal capacity matter? a survey of wto members*. *World Trade Review*, 8(4):559–577, 2009.

Mark S. Copelovitch and David Ohls. Trade, institutions, and the timing of gatt/wto ac-

cession in post-colonial states. *The Review of International Organizations*, 7(1):81–107, 2011. doi: 10.1007/s11558-011-9129-2.

Maia K. Davis Cross. Rethinking epistemic communities twenty years later. *Review of International Studies*, 39(01):137–160, Nov 2012. doi: 10.1017/s0260210512000034.

Christina L. Davis and Sarah Blodgett Bermeo. Who files? developing country participation in gatt/wto adjudication. *The Journal of Politics*, 71(3):1033–1049, 2009. doi: 10.1017/s0022381609090860.

Christina L. Davis and Meredith Wilf. Joining the club: Accession to the gatt/wto. *The Journal of Politics*, 79(3):964–978, 2017. doi: 10.1086/691058.

William J Drake and Kalypso Nicolaidis. Ideas, interests, and institutionalization: "trade in services" and the uruguay round. *International Organization*, 46(1):37–100, 1992.

Andreas Dür, Leonardo Baccini, and Manfred Elsig. The design of international trade agreements: Introducing a new dataset. *The Review of International Organizations*, 9(3): 353–375, 2013. doi: 10.1007/s11558-013-9179-8.

Simon J Evenett and Carlos A Primo Braga. Wto accession : lessons from experience. *The World Bank Group, International Trade Department*, Jun 2005.

Timothy Feddersen and Wolfgang Pesendorfer. Voting behavior and information aggregation in elections. *Econometrica*, 65(5):1029–1058, 1997.

Kevin P. Gallagher. Understanding developing country resistance to the doha round. *Review of International Political Economy*, 15(1):62–85, 2007. doi: 10.1080/09692290701751308.

Peter M. Haas. Introduction: epistemic communities and international policy coordination. *International Organization*, 46(01):1, 1992. doi: 10.1017/s0020818300001442.

Robert Howse. From politics to technocracy-and back again: The fate of the multilat-

eral trading regime. *The American Journal of International Law*, 96(1):94, 2002. doi: 10.2307/2686127.

Matias Iaryczower. Strategic voting in sequential committees. 2007.

G. John Ikenberry. A world economy restored: expert consensus and the anglo-american postwar settlement. *International Organization*, 46(01):289, 1992. doi: 10.1017/s002081830000151x.

Ethan Barnaby Kapstein. Between power and purpose: central bankers and the politics of regulatory convergence. *International Organization*, 46(01):265, 1992. doi: 10.1017/s0020818300001508.

Matthew Kennedy. The integration of accession protocols into the wto agreement. *Journal of World Trade*, 47(1):45–75, 2013.

Barbara Koremenos, Charles Lipson, and Duncan Snidal. The rational design of international institutions. *The Rational Design of International Institutions*, page 1–40, 2002. doi: 10.1017/cbo9780511512209.003.

Stephen D. Krasner. Global communications and national power: Life on the pareto frontier. *World Politics*, 43(03):336–366, 1991. doi: 10.2307/2010398.

Lisa L. Martin. Interests, power, and multilateralisml. *International Organization*, 46(4): 765–792, 1992.

Eric Neumayer. Strategic delaying and concessions extraction in accession negotiations to the world trade organization: An analysis of working party membership. *World Trade Review*, 12(4):669–692, 2013.

Marco Ottaviani and Peter Sorensen. Information aggregation in debate: who should speak first? *Journal of Public Economics*, 2001.

Krzysztof J. Pelc. Why do some countries get better wto accession terms than others? *International Organization*, 65(04):639–672, 2011. doi: 10.1017/s0020818311000257.

Dani Rodrik. *The globalization paradox: democracy and the future of the world economy*. W.W. Norton Co., 2012.

USDA. Global agriculture and the doha round—market access is the key. *https://www.ers.usda.gov/amber-waves/2006/september/global-agriculture-and-the-doha-round-market-access-is-the-key/*, 2006.

WT/ACC/CHN/49. *Report of the Working Party on the Accession of China*. 2001.

WTO. Lamy: It's time for a new "geneva consensus" on making trade work for development. *WTO News*, Oct 2006.

Appendix A: Proofs

Proof of Proposition 1

I begin by characterizing beliefs for members of committee two. Below I refer to *ex ante* beliefs to mean a committee member's beliefs after observing x_1 and *ex post* beliefs to mean a committee member's beliefs after observing both x_1 and s_2. *Ex post* beliefs are given by a function, $\mu : \{0,1\} \times \{-1,1\} \to (0,1)$ mapping the product of first committee's policy space and the signal space into the unit interval. Given equilibrium strategies, beliefs for each pair (x_1, s_2) are,

$$\mu(1,1) = \frac{p^2\pi}{p^2\pi + (1-p)^2(1-\pi)} \tag{5}$$

$$\mu(1,-1) = \pi \tag{6}$$

$$\mu(0,1) = \frac{p(p\theta + (1-p))\pi}{p(p\theta + (1-p))\pi + (1-p)[(1-p)\theta + p](1-\pi)} \tag{7}$$

$$\mu(0,-1) = \frac{(1-p)(p\theta + (1-p))\pi}{(1-p)(p\theta + (1-p))\pi + p[(1-p)\theta + p](1-\pi)} \tag{8}$$

Note that beliefs following $x_1 = 0$ incorporate both the probability that the signal was $s_1 = 0$ (that is $1-p$) and the probability that $s_1 = 1$ but that there were B types present in the first period committee ($p\theta$).

Next we consider optimal voting behavior in the second committee. Type A members of the committee optimally vote in favor of proposal $x_2 = 1$ if and only if,

$$\mu(x_1, s_2) - (1 - \mu(x_1, s_2)) \geq 0$$

or $\mu(x_1, s_2) \geq \frac{1}{2}$. Otherwise she optimally votes for $x_1 = 0$. Type B members always prefer $x_1 = 0$ to $x_1 = 1$ and so vote in favor of the former. Note that A types may or may not vote

responsively to the signal s_2. There are three possible cases.

First, suppose that $\mu(1,-1) > \frac{1}{2}$, that is after observing $x_1 = 1$ and $s_2 = -1$ A types still believe it more likely than not that the state is $\omega = 1$. In this case A types in the second period prefer to vote in favor of $x_2 = 1$. Since $\mu(1,1) > \mu(1,-1)$ it must also be the case that $\mu(1,1) > \frac{1}{2}$ implying that A types will never vote for $x_2 = 0$ after observing a first period outcome $x_1 = 1$. This condition obtains whenever $\pi \geq \frac{1}{2}$.

In the second case, $\frac{1}{2} > \mu(0,1)$ implying that after observing $x_1 = 0$ and $s_1 = 1$, A types in the second period strictly prefer $x_2 = 0$. Since this also implies that $\frac{1}{2} > \mu(0,-1)$, A types in the second period will never vote for $x_1 = 1$. This condition obtains whenever $\tilde{\pi} > \pi$ where $\tilde{\pi} \equiv \frac{1}{1+\phi}$ and,

$$\phi = \frac{p[p\theta + (1-p)]}{(1-p)[(1-p)\theta + p]}$$

Note that assumption A1 ensures that $\frac{1}{2} > \tilde{\pi}$. We can then consider final case, in which $\pi \in [\tilde{\pi}, \frac{1}{2})$. It must be that $\mu(1,1) \geq \frac{1}{2} > \mu(1,-1)$ and that $\mu(0,1) \geq \frac{1}{2} > \mu(0,-1)$. Then A types will vote responsively to s_2 regardless of the first period outcome. If $s_2 = 1$, A types optimally vote for $x_2 = 1$. Otherwise an A type optimally votes for $x_2 = 0$.

Turning to the first period, equilibrium beliefs are given by a function $\mu : \{-1, 1\} \to (0, 1)$ mapping the signal space to the unit interval. Beliefs are,

$$\mu(1) = \frac{p\pi}{p\pi + (1-p)(1-\pi)} \tag{9}$$

$$\mu(0) = \frac{(1-p)\pi}{(1-p)\pi + p(1-\pi)} \tag{10}$$

The analysis of optimal voting in the first committee is similar that of the second committee, with the exception that votes in the first period may also influence the outcome of the second through $\mu(\cdot,\cdot)$. In addition to conditioning on pivotality in the first period, agents must take

into account how their vote will impact the likely second period outcome. There are three cases to consider corresponding to the conditions identified above.

Case I: $\pi > \frac{1}{2}$

In the first case if $x_1 = 1$ then A types in the second period surely prefer $x_2 = 1$ while B types continue to prefer $x_1 = 0$. If $x_2 = 0$ then the preferences of A types will follow the signal s_2. Expected utility of an A type in the first period from outcome $x_1 = 1$ is,

$$2[\mu(s_1) - (1 - \mu(s_1))]$$

Expected utility of outcome $x_1 = 0$ is,

$$(1 - \theta)Pr(s_2 = 1|s_1)\big[2Pr(\omega = 1|s_1, s_2 = 1) - 1\big]$$

Given $\pi \geq \frac{1}{2}$ this quantity is always (weakly) greater than zero. Agent i prefers to vote in favor of $x_1 = 1$ if and only if,

$$\mu(s_1) > \frac{1}{2} + \frac{(1 - \theta)Pr(s_2 = 1|s_1)\big[Pr(\omega = 1|s_1, s_2 = 1) - \frac{1}{2}\big]}{2}$$

Note that this is a higher threshold than that required by agents in the second period. This reflects conservatism on the part of A types in the first period who fear shutting off the responsiveness of second period agents to their own signal. This means that first period agents may believe it more likely than not that the state is $\omega = 1$, and thus strictly prefer $x_1 = 1$ considered in isolation. Yet they still prefer to vote in favor of $x_1 = 0$ in order to reap the benefits of information gathering in the second period.

Type A agents will nonetheless vote responsively if,

$$\mu(1) > \frac{1}{2} + \frac{(1-\theta)[\mu(1)p + (1-\mu(1))(1-p)][Pr(\omega = 1|s_1 = 1, s_2 = 1) - \frac{1}{2}]}{2}$$

$$\text{and} \qquad \mu(0) < \frac{1}{2} + \frac{(1-\theta)[\mu(0)p + (1-\mu(0))(1-p)][Pr(\omega = 1|s_1 = 0, s_2 = 1) - \frac{1}{2}]}{2}$$

As $\theta \to 1$ the right hand side of both conditions approaches $\frac{1}{2}$ while the left hand side remains constant. By assumption A1, there exists a $\tilde{\theta}$ such that if $\theta \in (0, \tilde{\theta})$ an equilibrium exists in which first period A types vote responsively.

Case II: $\tilde{\pi} > \pi$

In this case, after observing $x_1 = 0$ type A's in the second period prefer $x_2 = 0$ irrespective of the signal s_2. Expected utility of an A type in the first period if $x_1 = 0$ is surely zero in both periods. Expected utility if $x_1 = 1$ is,

$$\mu(s_1) + (1 - \mu(s_1)) + (1-\theta)Pr(s_2 = 1|s_1)[Pr(\omega = 1|s_1, s_2 = 1) - (1 - Pr(\omega = 1|s_1, s_2 = 1))]$$

Thus an A type in the first period, prefers $x_1 = 1$ if,

$$\mu(s_1) \geq \frac{1}{2} - (1-\theta)Pr(s_2 = 1|s_1)\left[Pr(\omega = 1|s_1, s_2 = 1) - \frac{1}{2}\right]$$

An A type in the first period will vote responsively if and only if,

$$\mu(1) > \frac{1}{2} - (1-\theta)Pr(s_2 = 1|s_1)\left[Pr(\omega = 1|s_1, s_2 = 1) - \frac{1}{2}\right] \tag{11}$$

$$\text{and} \qquad \mu(0) < \frac{1}{2} - (1-\theta)Pr(s_2 = 1|s_1)\left[Pr(\omega = 1|s_1, s_2 = 1) - \frac{1}{2}\right] \tag{12}$$

Note that when $s_1 = 1$, $Pr(\omega = 1|s_1 = 1, s_2 = 1) > \frac{1}{2}$ implying that the right hand side

of (11) is strictly less than $\frac{1}{2}$. But by A1 $\mu(1) > \frac{1}{2}$ so the first condition is implied. When $s_1 = -1$, $Pr(\omega = 1|s_1 = -1, s_2 = 1) = \pi$ which is strictly less than $\frac{1}{2}$. Thus the right hand side of (12) is strictly greater than $\frac{1}{2}$. Then (12) also holds by assumption A1. In this case first period agents of type A vote responsively for any value of θ.

Case III: $\pi \in [\tilde{\pi}, \frac{1}{2}]$

In this case second period agents of type A always vote with the signal s_2. In this case, there is no possibility that the first period outcome will affect x_2. Thus agents of type A prefer to vote in favor of $x_1 = 1$ if and only if $\mu(s_1) \geq \frac{1}{2}$. By assumption A1 this ensures that for any $\theta \in (0, 1)$, A types will vote responsively in the first period. In all three cases, agents of type B prefer to vote in favor of $x_1 = 0$. ■

Proof of Proposition 2.

Let $\tilde{\pi} > \pi$, where $\tilde{\pi}$ is defined as above, and $x_1 = 0$. Suppose that $s_2 = 1$. Beliefs are,

$$\frac{p(p\theta + (1-p))\pi}{p(p\theta + (1-p))\pi + (1-p)[(1-p)\theta + p](1-\pi)} \tag{13}$$

By the definition of $\tilde{\pi}$ this is bounded above by,

$$\frac{p(p\theta + (1-p))}{p(p\theta + (1-p)) + (1-p)[(1-p)\theta + p]\phi} = \frac{1}{2} \tag{14}$$

Since $\frac{1}{2} > \mu(0, 1)$ type A committee members prefer the outcome $x_2 = 0$ so this will obtain regardless of the presence of B types in the committee. ■

Proof of Proposition 3

I begin by comparing equilibrium beliefs and voting behavior in Γ^p with that in Γ^c. Beliefs in the first period of the participation game are identical to those in the decision making game. If the first period outcome is $x_1 = 1$ beliefs in the second period are again identical to the baseline model. If the first period outcome is instead $x_1 = 0$, beliefs depend on the

probability that the outcome is *uninformative*. This reflects both the probability that a B type participated and the probability no agent of either type participated (since $x_1 = 0$ whenever $k_1 = 0$).

In the baseline model, x_1 is uninformative whenever at least one B type participates. This occurs with probability θ, where θ is given exogenously given. In contrast, in the participation game the probability of a signal being uninformative is determined endogenously by equilibrium strategies. In particular, an outcome is responsive if no B type joins, but at least one A type joins. The probability that an outcome is unresponsive is the complement of this probability. For any strategy profile $\boldsymbol{c} = (c_{A,1}, c_{B,1}, c_{A,2}^{x_1}, c_{B,2}^{x_1})$, this is,

$$\theta' = 1 - (c_{B,1})^b[1 - (c_{A,1})^{N-b}]$$

When $x_1 = 0$ second period beliefs are equivalent to beliefs in the baseline model but with θ' replacing θ. Note that in the participation game the threshold $\hat{\pi}$ is defined identically to $\tilde{\pi}$ above, but with θ' replacing θ.

While in the second period optimal voting strategies remain as before, in the first period voting strategies must take into account not only the direct influence of x_1 on x_2, but also its indirect influence through the participation strategies of both agent types. By assumption A3 technocrat types in the second period will vote responsively to the signal if and only if $x_1 = 1$. Then, a first period agent of type A prefers to vote in favor of $x_1 = 1$ if and only if,

$$\mu(s_1) \geq \frac{1}{2} - (1 - \theta')Pr(s_2 = 1|s_1)\left[Pr(\omega = 1|s_1, s_2 = 1) - \frac{1}{2}\right]$$

This differs from the equivalent condition in the baseline model since the probability of responsiveness is now given by $(1 - \theta')$. As before though Assumption A1 ensures that type A agents vote responsively for any $\theta' \in (0, 1)$. As before, B types prefer $x_1 = 0$ and vote accordingly.

Next, we consider optimal participation strategies for each type of agent, beginning in $t = 2$. We assume throughout that $\tilde{\pi} > \pi$. It is immediate that if $x_1 = 0$ then no agents of either type participate in the second period. All agents strictly prefer $x_1 = 0$ so are unwilling to incur any cost to participate implying $c^0_{A,2} = c^0_{B,2} = 1$. If instead $x_1 = 1$, an agent of type B prefers to participate if and only if,

$$\nu_i (c^1_{B,2})^{b-1}[1 - (c^1_{A,2})^{N-b}][\mu(1)p + (1 - \mu(1))(1 - p)] \geq \phi_B$$

That is, an agent of type B prefers to participate if his expected gain from doing so, weighted by the probability of pivotality, exceeds the cost, ϕ_B. The probability that an agent of type B is pivotal is determined by the joint probability that no other B type participates, at least one A type does, and the realized signal is $s_2 = 1$. The equilibrium cutpoint is determined by the type, ν_i, for which the above condition holds with equality,

$$c^1_{B,2} = \left[\frac{\phi_B}{[1 - (c^1_{A,2})^{N-b}][\mu(1)p + (1 - \mu(1))(1 - p)]} \right]^{\frac{1}{b}}$$

Similarly an agent of type A prefers to participate if and only if,

$$\nu_i (c^1_{A,2})^{N-b-1} (c^1_{B,2})^{b} [\mu(1)p + (1 - \mu(1))(1 - p)]\mu(1, 1) \geq \phi_A$$

where the equilibrium cutpoint is determined as the type for which the above holds with equality,

$$c^1_{A,2} = \left[\frac{\phi_A}{(c^1_{B,2})^{b}[\mu(1)p + (1 - \mu(1))(1 - p)]\mu(1, 1)} \right]^{\frac{1}{N-b}}$$

Solving the system of equations yields the following equilibrium cutpoints,

$$c^1_{A,2} = \left[\frac{\phi_A}{\phi_A + \frac{\phi_B p^2 \pi}{p^2\pi + (1-p)^2(1-\pi)}}\right]^{\frac{1}{N-b}}$$

$$c^1_{B,2} = \left[\frac{p\pi + (1-p)(1-\pi)}{p^2\pi + (1-p)^2(1-\pi)} \cdot \left(\frac{p^2\pi + (1-p)^2(1-\pi)}{p^2\pi} \cdot \phi_A + \phi_B\right)\right]^{\frac{1}{b}}$$

Next, consider $t = 1$. If a type B agent is pivotal in the first period, he alters the outcome x_1 with certainty. In addition with positive probability he also alters the outcome x_2 since second period agents will all strictly prefer $x_2 = 0$ and no longer vote responsively to that period's signal. Taking both possibilities into account, a B type prefers to participate in the first period if and only if,

$$\nu_i (c_{B,1})^{b-1}[1 - (c_{A,1})^{N-b}][\pi p + (1-\pi)(1-p)] \cdot$$
$$\left[1 + (c^1_{B,2})^b[1 - (c^1_{A,2})^{N-b}][\mu(1)p + (1-\mu(1))(1-p)]\right] \geq \phi_B$$

The first line in the above represents the probability that i is pivotal in the first period (that is, he is the only B type to participate, at least one A type also participates, and the signal is $s_1 = 1$), weighted by his salience parameter. The second line represents his expected gain from participation. This is equal to one with certainty in the first period. His expected second period gain is equal to the probability that, if he had not participated, the outcome $x_2 = 1$ would have been realized. The cutpoint is defined as above by the type for which this condition holds with equality.

Finally, an agent of type A prefers to participate in the first period if and only if,

$$\nu_i (c_{A,1})^{N-b-1}(c_{B,1})^b[\pi p + (1-\pi)(1-p)] \cdot$$
$$\left[\mu(1) + (c^1_{B,2})^b[1 - (c^1_{A,2})^{N-b}][\mu(1)p + (1-\mu(1))(1-p)]\mu(1,1)\right] \geq \phi_A$$

In contrast to an agent of type B, an A type is pivotal only when no agents of either type besides herself participate in the first period. In addition, her expected gain from participation reflects her beliefs about the state of the world conditioning on pivotality in each period. Again, solving the system of equations we obtain expressions for each cutpoint,

$$c_{A,1} = \left[\frac{\phi_A(1+\phi_B)}{\phi_A(1+\phi_B) + \phi_B\left(\frac{p\pi}{p\pi+(1-p)(1-\pi)} + \frac{\phi_B p^2\pi}{p^2\pi+(1-p)^2(1-\pi)}\right)} \right]^{\frac{1}{N-b}}$$

$$c_{B,1} = \left[\left(\frac{\phi_A}{\frac{p\pi}{p\pi+(1-p)(1-\pi)} + \frac{\phi_B p^2\pi}{p^2\pi+(1-p)^2(1-\pi)}} + \frac{\phi_B}{1+\phi_B} \right) \cdot \left(\frac{1}{\pi p + (1-\pi)(1-p)} \right) \right]^{\frac{1}{b}}$$

■

Proof of Proposition 4.

Note that, as mentioned in the text, the expected participation of A types is increasing if $x_1 = 1$. This is the case if and only if,

$$(N-b)\left[1 - \underbrace{\left[\frac{\phi_A}{\phi_A + \phi_B\mu(1,1)}\right]^{\frac{1}{N-b}}}_{c^1_{A,2}}\right] >$$

$$(N-b)\left[1 - \underbrace{\left[\frac{\phi_A(1+\phi_B)}{\phi_A(1+\phi_B) + \phi_B\left(\mu(1) + \phi_B\mu(1,1)\right)}\right]^{\frac{1}{N-b}}}_{c_{A,1}}\right]$$

Re-arranging this expression yields,

$$\phi_B\mu(1,1) > \frac{\phi_B[\mu(1) + \phi_B\mu(1,1)]}{1+\phi_B}$$

or,

$$\frac{p^2\pi}{p^2\pi + (1-p)^2(1-\pi)} > \frac{p\pi}{p\pi + (1-p)(1-\pi)}$$

which is true by assumption $A1$. Turning to the main result, expected participation in the first period is,

$$\begin{aligned}\mathbb{E}[k_1] =& \mathbb{E}[k_1^A] + \mathbb{E}[k_1^B] \\ =& (N-b)(1-c_{A,1}) + b(1-c_{B,1})\end{aligned}$$

Recalling that $c_{A,2}^0 = c_{B,2}^0 = 1$, expected second period participation is,

$$\begin{aligned}\mathbb{E}[k_2] =& Pr(x_1 = 1)\left(\mathbb{E}[k_2^A|x_1 = 1] + \mathbb{E}[k_2^B|x_1 = 1]\right) \\ =& (c_{B,1})^b[1-(c_{A,1})^{N-b}][\pi p + (1-\pi)(1-p)]\left[(N-b)(1-c_{A,2}^1) + b(1-c_{B,2}^1)\right]\end{aligned}$$

From the proof of Proposition 3 first period cutpoints are,

$$c_{A,1} = \left[\frac{\phi_A(1+\phi_B)}{\phi_A(1+\phi_B) + \phi_B\left(\frac{p\pi}{p\pi+(1-p)(1-\pi)} + \frac{\phi_B p^2\pi}{p^2\pi+(1-p)^2(1-\pi)}\right)}\right]^{\frac{1}{N-b}}$$

$$c_{B,1} = \left[\left(\frac{\phi_A}{\frac{p\pi}{p\pi+(1-p)(1-\pi)} + \frac{\phi_B p^2\pi}{p^2\pi+(1-p)^2(1-\pi)}} + \frac{\phi_B}{1+\phi_B}\right)\cdot\left(\frac{1}{\pi p + (1-\pi)(1-p)}\right)\right]^{\frac{1}{b}}$$

Note that as $\phi_B \to 0$, $c_{A,1} \to 1$, but $c_{B,1} \to \left[\frac{\phi_A}{[\mu(1)]^2}\right]^{1/b}$ which is strictly less than one whenever,

$$[\mu(1)]^2 > \phi_A$$

This is true by Assumption A2. Then as $\phi_B \to 0$, $Pr(x_1 = 1) \to 0$ while $\mathbb{E}[k_1]$ is strictly positive. Second period cutpoints are,

$$c^1_{A,2} = \left[\frac{\phi_A}{\phi_A + \frac{\phi_B p^2 \pi}{p^2\pi + (1-p)^2(1-\pi)}} \right]^{\frac{1}{N-b}}$$

$$c^1_{B,2} = \left[\left(\frac{1}{\mu(1)p + (1-\mu(1))(1-p)} \right) \cdot \left(\frac{\phi_A}{\mu(1,1)} + \phi_B \right) \right]^{\frac{1}{b}}$$

As $\phi_B \to 0$, $c^1_{A,2} \to 1$ again while,

$$c^1_{B,2} \to \left[\left(\frac{1}{\mu(1)p + (1-\mu(1))(1-p)} \right) \left(\frac{\phi_A}{\mu(1,1)} \right) \right]^{\frac{1}{b}}$$

which is strictly less than one by assumption A2. In summary as $\phi_B \to 0$, $\mathbb{E}[k_1]$ is strictly positive while $\mathbb{E}[k_2]$ approaches zero. Since all cutpoints are continuous in ϕ_B, there must exist a $\bar{\phi}_B$ such that for any $\phi_B \in (0, \bar{\phi}_B)$, expected participation is strictly decreasing over time: $\mathbb{E}[k_1] > \mathbb{E}[k_2]$. ■

Appendix B: Additional Empirical Results

Figure 8: Mean KL_All by Applicant

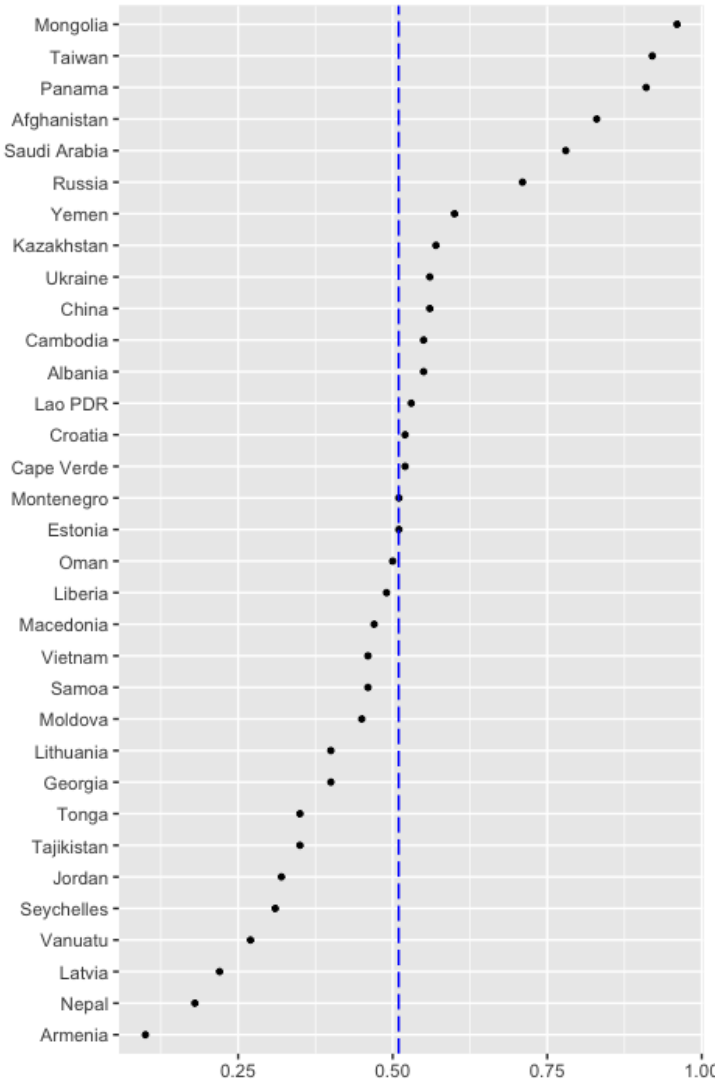


Table 4: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Joins6	8,220	0.05	0.22	0	1
KL	8,220	0.64	0.43	0.10	3.27
KL_All	8,220	0.53	0.40	0.10	2.51
OriginalMembership	8,220	19.37	6.33	8	42
MembershipsToDate	8,220	6.24	5.65	1	32
LogGDPI	8,220	4.54	0.41	3.76	5.64
LogGDPJ	8,220	4.66	0.35	3.64	5.71
GDPpcI	8,220	9,534.49	10,706.02	1,291.84	44,686.05
GDPpcJ	8,220	16,093.52	25,550.93	574.31	146,215.80
pta	8,220	0.06	0.24	0	1
Quad	8,220	0.01	0.07	0	1
PolityI	8,220	0.70	6.59	−10	10
PolityJ	8,220	4.29	6.05	−10	10
RAM	8,220	0.11	0.32	0	1
MissionSize	8,220	4.81	3.02	0	24
lImportsIJ	8,220	8.08	6.78	0.00	20.51
lExportsIJ	8,220	6.29	7.05	0.00	21.32
TotalSimilarityIJ	8,220	0.03	0.04	0.00	0.38
IdealPointI	8,220	−0.02	0.89	−1.50	1.57
IdealPointJ	8,220	−0.21	0.74	−1.65	2.60

Table 5

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	−40.07	−4.54*	−4.30*	−9.52*	−5.10*
	(972.39)	(1.64)	(1.66)	(1.41)	(1.56)
lKL	−1.95*	−1.88*	−1.93*	−1.98*	
	(0.38)	(0.34)	(0.38)	(0.31)	
lKL_All	1.91*	1.77*	1.65*	1.54*	
	(0.35)	(0.32)	(0.36)	(0.29)	
KL					−4.04*
					(0.67)
KL_All					3.61*
					(0.66)
lTotalTradeIJ	0.12*	0.06*	0.08*	0.03*	0.06*
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
OriginalMembership	0.98	0.06*	0.09*	0.11*	0.08*
	(36.01)	(0.02)	(0.02)	(0.01)	(0.02)
SimilarityUS	−1.26*	−1.13*	−1.30*	−1.02*	−1.10*
	(0.46)	(0.44)	(0.45)	(0.37)	(0.42)
N	5268	5268	5268	5261	4831
Dependent Variable	$Joins6$	$Joins6$	$Joins3$	$Joins9$	$Joins6$
Applicant Fixed Effects	Yes	No	No	No	No
Year Fixed Effects	No	Yes	No	No	No
Top 5% outliers removed	No	No	No	No	Yes

Logistic regression with standard errors clustered at document level. Dependent variable takes a value of 1 if potential member joins working party$_i$ within six, three, or nine months following release of document$_{i,t}$. * indicates significance at $p < 0.05$. All models include (log) GDPI, (log) GDPJ, GDPpcI, GDPpcJ, OriginalMembership, MembershipsToDate, PTA, and Quad.