

Coercive Trade Policy[†]

By VINCENT ANESI AND GIOVANNI FACCHINI*

Coercion is used by one government (the “sender”) to influence the trade practices of another (the “target”). We build a two-country trade model in which coercion can be exercised unilaterally or channeled through a “weak” international organization without enforcement powers. We show that unilateral coercion may be ineffective because signaling incentives lead the sender to demand a concession so substantial to make it unacceptable to the target. If the sender can instead commit to the international organization’s dispute settlement mechanism, then compliance is more likely because the latter places a cap on the sender’s incentives to signal its resolve. (JEL D74, D82, F12, F53)

In international trade disputes, coercion is often used against governments whose trade practices are deemed unfair. Trade coercion occurs when a “sender government” makes a demand backed by threats to use retaliatory sanctions against a “target government” if the latter does not acquiesce to this demand. There are typically two distinct methods of trade coercion: it can be exercised unilaterally (e.g., Eaton and Engers 1992) or through multilateral institutions (e.g., GATT and WTO). In the case of unilateral coercion, the sender government makes a demand and (if necessary) retaliates one-sidedly, unconstrained by international obligations.¹ In the case of multilateral coercion, the sender uses instead an international institution’s framework for trade dispute resolution.²

While a small body of empirical literature has studied trade coercion—and in particular whether multilateral organizations like the GATT or WTO can

*Anesi: School of Economics, University of Nottingham, The Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom, and GEP (email: vincent.anesi@nottingham.ac.uk); Facchini: University of Nottingham, School of Economics, The Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom, CEPR, CES-Ifo, GEP, and IZA (email: giovanni.facchini@nottingham.ac.uk). Michael Ostrovsky was coeditor for this article. We wish to thank Kyle Bagwell, Mostafa Beshkar, Eric Bond, Giovanni Maggi, Doug Nelson, Emanuel Ornelas, Bob Staiger, and participants at the XIV SAET Conference in Tokyo, the Midwest International Trade Meetings in Rochester, and the 2016 Stanford Institute for Theoretical Economics for useful comments. We are also grateful to three referees of this journal, who helped us improve significantly the paper.

[†]Go to <https://doi.org/10.1257/mic.20170085> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

¹A typical example was Section 301 of the 1974 US Trade Act, which allowed the United States to impose unilateral sanctions on countries whose trade practices were found to be unfair to US interests. This clause was invoked in several occasions—for instance, in the much publicized dispute with Japan over automobiles in 1995, in which the United States essentially bypassed the WTO and imposed sanctions unilaterally (e.g., Puckett and Reynolds 1996, Schoppa 1999).

²The WTO Dispute Settlement Mechanism is the leading institution of this kind, and since its inception, it has handled hundreds of cases. Several preferential trade agreements also include similar institutions. See, for instance, NAFTA’s Dispute Settlement Process or MERCOSUR’s Dispute Settlement Mechanism.

increase the chance of a sender government obtaining concessions from targets (e.g., Busch 2000, Pelc 2010, Busch and Reinhardt 2000)—few if any formal analyses explicitly incorporate both the coercion process itself and the role of the institutions through which coercion is channeled. The purpose of this paper is to contribute toward filling this gap, by analyzing the strategic incentives underlying trade coercion under three different institutional settings. In particular, we ask two main questions: how can international trade institutions achieve their objectives if defendants can reject adverse rulings with impunity? Given international trade institutions' limited enforcement powers, is unilateral or multilateral coercion more effective in inducing the target country to concede?

To address these questions, we set up a model depicting a dispute between two states, Home and Foreign, in which the Foreign government is dissatisfied with the trade policy implemented by the Home government. A key feature of trade coercion is the target government's lack of information on the sender government's domestic political constraints (e.g., Eaton and Engers 1999, Busch and Reinhardt 2000, Bagwell and Staiger 2005, Beshkar and Bond 2017). To capture this idea, we assume that the political pressure exerted by the import-competing sector on the government in Foreign is private information and is only known by the Foreign government. This political pressure plays a key role in shaping its level of resolve—i.e., the severity of its trade sanctions against the Home government—in a potential trade war.

Appraising the actual effectiveness of an international organization (IO) in dispute settlement requires knowing what would happen if that institution did not exist—i.e., if there were no framework of rules governing trade coercion. For this reason, the first setting we examine is one in which unilateral coercion is the only option. The game begins with the Foreign government making a demand. The Home government can concede (ending the game with the implementation of the demanded tariff), or reject it (triggering a retaliatory trade war). In other words, it must decide which concessions are acceptable, that is, which tariff changes it would prefer to make rather than face Foreign's trade sanctions. Since the precise nature of these sanctions is uncertain and crucially depends on the privately observed level of resolve of the Foreign government, the latter has incentives to signal high levels of resolve by making excessive demands about the concessions required from Home to avoid retaliatory measures. Our characterization of equilibrium outcomes in this case reveals that such incentives lead the Foreign government to make requests that the Home government will not meet, thus causing a retaliatory trade war—even when there exist mutually advantageous policy concessions.

As we will show, a key factor in determining whether concessions can be obtained with multilateral coercion is the extent to which the sender government can commit not to bypass the dispute settlement process of the IO through which coercion is channeled. To model the different strategic situations that may arise from differences in the sender's ability to commit to the IO, we will examine two distinct variants of the previous model. In the first, the Foreign government is not allowed to bypass the IO's dispute settlement process. As a result, multilateral coercion is its only option available. Dispute settlement is modeled by allowing the Foreign government to make a demand to the Home government prior to the IO

ruling. This assumption is intended to capture, e.g., the consultations stage of WTO disputes. If the Home government does not concede to the Foreign government's demand, the IO issues its ruling, whereas it remains inactive otherwise. As our aim is to investigate the effectiveness of weak international trade institutions—namely those that have no enforcement power and rely on the sender government itself to implement any retaliatory measures—the Home government is allowed not to comply with the ruling, thus triggering a trade war with the Foreign government. Our analysis shows that commitment to the IO's ruling makes concessions more likely. Intuitively, the potential IO ruling places a cap on the Foreign government's incentives to signal its resolve with high demands. This results in the latter making more moderate requests, which can be accepted by Home. Compared with the unilateral-coercion case above, this finding provides a possible explanation for the empirical evidence (e.g., Busch 2000, Pelc 2010) indicating that, although neither GATT nor WTO have enforcement powers, a target of trade coercion by the United States is significantly less likely to concede when coercion is unilateral than when it is multilateral.

In the second variant of the model, the Foreign government is only partially committed to the IO's dispute settlement process, in the sense that it can choose between unilateral and multilateral coercion in an additional stage at the beginning of the game, committing itself to that choice.³ This setting captures the environment created by Section 301 of the US Trade Act of 1974. In fact, this provision enabled the president to impose sanctions unilaterally against unfair trade practices, eliminating the need to observe existing international obligations (e.g., Puckett and Reynolds 1996). We show that the mere availability of the unilateral option prevents the foreign government from obtaining concessions in equilibrium. In fact, using multilateral coercion when unilateral coercion is available is perceived as a sign of the foreign government's weakness. Hence, incentives to signal higher levels of resolve to the Home government will lead the Foreign government to make unilateral demands which, as discussed above, cannot be accepted in equilibrium.

A large body of literature has studied international trade agreements as equilibria of infinitely-repeated prisoner-dilemma games, in which deviations from the (implicit) agreements are followed by indefinite play of high-tariff Nash equilibria. Papers in that literature study how international organizations' dispute settlement procedures can facilitate cooperation—e.g., Riezman (1991); Hungerford (1991); Maggi (1999); Ludema (2001); Klimenko, Ramey, and Watson (2008); Limão and Saggi (2008); Bagwell (2009); and Park (2011). Although strategic incentives in our paper do not rely on long-run interactions, our analysis shares some important features with Bagwell (2009) and Park (2011). In particular, Bagwell (2009) develops a two-country repeated game in which each government, as the sender in our model, possesses some private information about the extent of the political pressures it is confronted with. He shows that trade agreements with weak bindings (i.e., maximal tariff levels) are preferable to agreements with strong bindings (i.e., precise tariff levels). Interestingly, if private information is persistent through

³ As we will discuss in Section IB, this assumption is consistent with empirical evidence on US trade coercion.

time, then signaling incentives, which resemble those of our unilateral-coercion game, may undermine the Nash-reversion threat supporting a cooperative equilibrium. Indeed, in order not to be perceived as “weak,” governments may resist applying optimal tariffs, below their bound levels. Park (2011), like us, shows that an international trade organization without enforcement powers may play an effective role in preventing trade wars, but the reason is different. The author considers a two-country repeated game with imperfect private monitoring, in which each country only observes an (imperfect) private signal about the other country’s protection levels. In the absence of an IO, the private nature of the signals about their opponents’ policies restricts the punishment schemes that the countries can use to sustain cooperation in equilibrium. Observing both countries’ private signals, the IO decides whether a violation has occurred and then tells them to initiate a punishment phase based on its decision. Park shows that, despite its lack of coercive power, the IO can help sustain the cooperative equilibrium by changing the nature of the punishment-triggering signals from private to public, thus relaxing informational constraints on optimal punishment schemes.

Most papers in the literature reviewed so far model trade dispute settlement as a set of conditions imposed on the off-the-equilibrium-path punishments that follow deviations, not explicitly as a coercion game like ours. One notable exception is the model of sanctions developed by Eaton and Engers (1992). In their framework, a sender and a target country interact repeatedly under perfect information, and the sender is able to commit for a limited period of time to carry out a sanction, costly for both parties involved, if the target does not acquiesce to a sender’s minimum request. They find that the threat of sanctions can be effective in obtaining concessions from the target, even though sanctions are costly to the sender and not actually used in equilibrium. In a follow-up paper, Eaton and Engers (1999), they show how asymmetric information about senders and targets may generate equilibria in which sanctions are imposed by the sender on the equilibrium path, either to maintain its reputation for toughness or because it cannot discriminate between complaisant and stubborn targets. Our approach to coercion differs from theirs in two important respects. First, we ask a different question. They are concerned with explaining why sanctions, or the threat of sanctions, can be effective in achieving senders’ objectives *in the absence of third parties* to enforce agreements between governments. In contrast, we are concerned with explaining how the presence of even “weak” IOs can have a significant impact on coercion outcomes. Second, even our unilateral-coercion benchmark approaches coercion from a different perspective. Eaton and Engers’ (1992, 1999) arguments require repeated interactions: if the sender dealt with the target only once, then it would have no incentive to use costly sanctions or to maintain a reputation. Our aim is to capture different strategic incentives, not channeled through long-run interactions. In our framework, although the game will not be repeated, the sender government is prepared to use trade sanctions because it is dissatisfied with the target’s deviation from an existing trade agreement and, in the absence of a new agreement, increasing its tariffs would be profitable. As it seeks to obtain as favorable terms as possible, threats of sanctions signal its resolve in case a trade war could not be avoided. While such signaling incentives are conducive to inefficient trade wars, we show

that in the presence of an IO, an agreement is reached even without a ruling and sanctions are *not* implemented on the equilibrium path. As discussed in Section IV, this is consistent with empirical evidence on trade disputes.⁴

A recent literature has taken an incomplete-contracts approach to international trade agreements and dispute settlement—e.g., Bagwell, Mavroidis, and Staiger (2007); Beshkar (2010); Beshkar (2016); Horn, Maggi, and Staiger (2010); Maggi and Staiger (2011); and Beshkar and Bond (2017). Its main focus is on the design of optimal institutions for international trade and dispute settlement in various informational/contractual environments.⁵ In contrast, the IO's dispute settlement procedure is the main *exogenous* variable in our model. Our aim is not to study the normative aspects of trade institutions but, instead, to provide a positive theory of how commitment to such institutions may affect trade coercion outcomes.

In the related context of international conflict resolution, Hörner, Morelli, and Squintani (2015) also shows that an IO without enforcement powers can be effective. More precisely, the authors consider a dispute between two players over some positive surplus that shrinks if they engage in a war, the outcome of which is determined by their privately observed types (strong or weak). The authors compare the optimal mechanisms that minimize the probability of war under two different third-party-intervention settings: arbitration, under which a third party collects information privately and makes *binding* decisions on the outcome of the dispute; and mediation, under which the third party also collects information but can only make unenforceable recommendations. Interestingly they find that, despite her lack of enforcement power, a mediator is as effective as an arbitrator in preventing wars. In contrast to our model in which the IO's ruling is exogenously fixed, it is the mediator's ability to use sufficiently sophisticated recommendation strategies that allows her to circumvent the unenforceability constraint in their model.

The rest of the paper is organized as follows. Section I describes the model, while Section II presents the main results of our analysis. In Section III, we discuss the substantive implications of our results and relate them to the existing empirical evidence. Section IV concludes.

I. The Model

The goal of this section is twofold. We start by presenting the basic structure of the economy and lay out next a simple model of trade coercion.

⁴We discuss a dynamic extension of our model that would also capture the governments' long-run incentives in Section IV. Other recent examples are Bagwell and Staiger (2005), Martin and Vergote (2008), and Rosendorff (2005), who analyze repeated tariff games in which, as in our model, governments have private information about their relative valuations of import-competing sectors.

⁵Gilligan, Johns, and Rosendorff (2010) studies how variation in the strength of international courts affects dispute settlements in a model which, like ours, incorporates complainants' demands and defendants' (potential) concessions. Their model can be applied to international trade institutions, but while we focus on weak institutions, their interest is in the pros and the cons of strengthening such institutions.

A. The Economic Environment

We consider a model with two large countries, Home and Foreign, trading between each other, which has been used in several previous analyses of trade negotiations.⁶ Each economy is characterized by three sectors, $i = 0, 1, 2$. All goods are produced using a constant-returns-to-scale technology under perfect competition. Good 0 is freely traded and serves as the numeraire; it is produced using labor alone. We choose units so that its international and domestic prices are both equal to one, and the aggregate labor supply, $L = \bar{L}$, is assumed to be large enough to sustain production of a positive amount of good 0. As a result, in a competitive equilibrium the wage rate equals unity in each country. Goods 1 and 2 are produced instead using labor and a sector-specific input, which is available in fixed supply. Home is abundant in sector-specific input 2, whereas Foreign is abundant in sector-specific input 1. Consequently, Home imports good 1, while Foreign imports good 2. To simplify the analysis, we assume symmetry in factor endowments between the two countries. The domestic and international prices of a non-numeraire good i are denoted by p_i and π_i , respectively. The rent R_i , accruing to the specific factor used in sector i , depends only on the good's producer price, and can thus be expressed as $R_i(p_i)$. Industry supply is given by $Q_i(p_i) = \partial R_i / \partial p_i$.

Trade policies in the two countries take the form of ad valorem import tariffs or subsidies, denoted by τ and τ^* , and drive a wedge between domestic and international prices.⁷ Focusing on the Home country, the domestic price of good 1 is thus equal to $p_1 = (1 + \tau)\pi_1$, with $\tau > 0$ ($\tau < 0$) representing an import tariff (subsidy); the domestic price of the export good is instead equal to $p_2 = \pi_2$. In Foreign, domestic prices are given by $p_1^* = \pi_1$ and $p_2^* = (1 + \tau^*)\pi_2$.

The economy is populated by a continuum of agents, and we set the population's size equal to one. Each agent shares the same quasi-linear and additively separable preferences, which can be written as

$$(1) \quad u(c_0, c_1, c_2) \equiv c_0 + \sum_{i=1}^2 u_i(c_i),$$

where c_0 is the consumption of the numeraire good, and c_i , $i = 1, 2$ represent instead the consumption of the other goods. The sub-utility functions are assumed to be twice differentiable, increasing, and strictly concave.

Provided that income always exceeds the expenditure on the numeraire good, the domestic demand for good $i \in \{1, 2\}$ can be expressed as a function of price alone, $D_i(p_i)$. Imports of good 1 by Home can then be expressed as $M_1(p_1) = D_1(p_1) - Q_1(p_1)$, while exports are instead given by $X_2(p_2) = Q_2(p_2) - D_2(p_2)$.

⁶ See, for instance, Mayer (1981) and Conconi, Facchini, and Zanardi (2012).

⁷ This allows us to describe the preferences of the two countries in the tariff space (τ, τ^*) and to easily characterize trade negotiations between them. As argued by Levy (1999), export subsidies and taxes are rarely used.

The international markets for goods 1 and 2 are in equilibrium when

$$(2) \quad M_1((1 + \tau)\pi_1) - X_1^*(\pi_1) = 0,$$

$$(3) \quad M_2^*((1 + \tau^*)\pi_2) - X_2(\pi_2) = 0.$$

From (2) and (3), we can derive an expression for international equilibrium prices as a function of the trade policies implemented in the two countries, i.e., $\pi_1(\tau)$, $\pi_2(\tau^*)$. Tariff revenues in Home are given by

$$(4) \quad \mathbf{T}(\tau) = \tau\pi_1(\tau)M_1(\tau)$$

and are assumed to be redistributed uniformly among all domestic residents.

In this model, individuals derive income from several different sources: they all supply one unit of labor and earn wages; they also receive the same lump sum transfer (possibly negative) of trade policy revenues from the government and they own some share of the specific inputs used in the production of goods 1 and 2. We assume that the Home government seeks to maximize a social welfare function, which is defined as a weighted sum of all citizens' income (total labor income, industry rents, and government revenues), plus consumer surplus, i.e.:

$$(5) \quad W(\tau, \tau^*) = 1 + \alpha R_1(\tau) + R_2(\tau^*) + \mathbf{T}(\tau) + \Omega(\tau) + \Omega(\tau^*),$$

where $\Omega(\tau) \equiv u(D_1(\tau)) - p_1 D_1(\tau)$ and $\Omega(\tau^*) \equiv u(D_2(\tau^*)) - p_2 D_2(\tau^*)$, i.e., the first term describes the surplus from the consumption of good 1 and the second from the consumption of good 2. We assume that $\alpha > 1$ to capture a protectionist bias in the setting of trade policy (Grossman and Helpman 2005).⁸ To simplify notation, we do not include α as an argument of the welfare function, as it will remain constant throughout the analysis.

The preferences of the Foreign government are given by

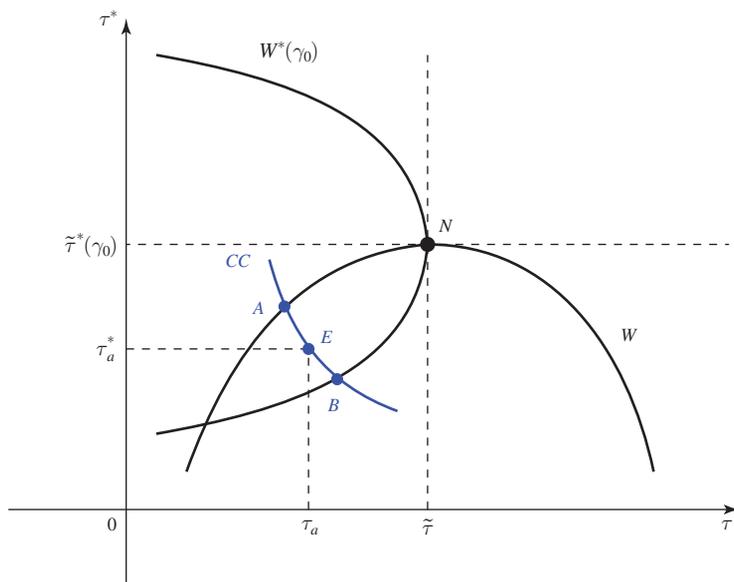
$$(6) \quad W^*(\tau, \tau^*, \gamma) = 1 + R_1^*(\tau) + \gamma R_2^*(\tau^*) + \mathbf{T}^*(\tau^*) + \Omega(\tau^*) + \Omega(\tau),$$

where γ captures the Foreign protectionist bias, and its initial value is given by $\gamma_0 > 1$. In the remainder of the paper, γ will be referred to as the Foreign government's type.⁹

As is standard in this class of models (e.g., Rosendorff 2005), we make the following natural assumptions about both governments' objective functions. First, for any given level of Home tariff τ [resp. of Foreign tariff τ^* and type γ], $W(\tau, \cdot)$ [resp. $W^*(\cdot, \tau^*, \gamma)$] strictly decreases with τ^* [resp. with τ]. This simply ensures that, in each country, the losses incurred by domestic export firms when the other country raises its tariff always outweigh the benefits to domestic consumers.

⁸For instance, this could be due to the fact that the import-competing industry is politically organized and lobbies the government in a model à la Grossman and Helpman (1994).

⁹See Bagwell and Staiger (2005) for a similar setting.

FIGURE 1. POLICY PREFERENCES WHEN $\gamma = \gamma_0$

Second, for any given level of τ^* [resp. of τ and γ], $W(\cdot, \tau^*)$ [resp. $W^*(\tau, \cdot, \gamma)$] first increases and then decreases with τ [resp. with τ^*]. This ensures that $W(\cdot, \tau^*)$ and $W^*(\tau, \cdot, \gamma)$ have unique maximizers, which we denote by $\tilde{\tau}$ and $\tilde{\tau}^*(\gamma)$ respectively. (Additive separability in (5) and (6) implies that $\tilde{\tau}$ is independent of τ^* , and that $\tilde{\tau}^*(\gamma)$ is independent of τ .)

Tariffs $\tilde{\tau}$ and $\tilde{\tau}^*(\gamma)$ are clearly those which would be implemented if governments chose their policies non-cooperatively—or, using the language of the previous literature, if they engaged in a “trade war.” Figure 1 provides an illustration for governments’ preferences: W (resp. W^*) describes an indifference curve for the Home (resp. Foreign) government. A downward (resp. leftward) shift leads to higher values of the government’s objective function. The policy pair $(\tilde{\tau}, \tilde{\tau}^*(\gamma))$ lies at the intersection between the two curves and describes the coordinates of point N . Clearly the two governments could make themselves better off if they could agree on any tariff pair lying within the lens described by the two indifference curves. In particular, the portion AB of the contract curve CC identifies the set of Pareto optimal tariff pairs that improve upon the Nash equilibrium N . Note that, as long as α and γ_0 are sufficiently large, AB belong to the positive orthant, and we will assume this to be the case. We further assume that the initial trade agreement between the two governments, (τ_a, τ_a^*) , belongs to AB .

B. A Simple Model of Trade Coercion

Our goal is to investigate whether and how different institutional arrangements affect the outcome of trade coercion. To do so, we develop a model with two active

agents, the Home and the Foreign governments, which possibly interact with an IO. In the preliminary stage, the Foreign government privately observes the new value of its protectionist bias γ , which is drawn from a cumulative distribution function F_0 . We assume that F_0 has a continuous and strictly positive density over some interval $[\underline{\gamma}, \bar{\gamma}]$. Note that for (and only for) expositional clarity, we assume that there is no asymmetry of information about the Home government's protectionist bias, α .¹⁰

In order to justify the use of coercion by the Foreign government in the next stages of the model, we need to make two assumptions. First, we assume that $\tilde{\tau}^*(\gamma) > \tau_a^*$. This ensures that the Foreign government is always prepared to raise its tariff above the initial agreement τ_a^* if the Home government does not acquiesce to its demands for lower Home tariffs. In addition, to accommodate cases of multilateral coercion (i.e., within the framework of the IO), the Foreign government's coercive actions must be justified by a violation of the original agreement, (τ_a, τ_a^*) . To this end, we further assume that a shock has also occurred in the Home country, which causes its government to seek an adjustment to its trade policy away from the original agreed level, τ_a . Several motives can be invoked for this desired change. For example, it could be the result of the election of a new government in Home, which prefers less trade than its predecessor (as in Bown 2002). Alternatively, it could be driven by a real or simply perceived import "surge," etc.¹¹ We assume that, as a consequence of this shock, the Home government has increased its tariff on imports to some $\tau_0 > \tau_a$. Note that this increase may not necessarily be in definite violation of the existing agreement between the two countries: as in Maggi and Staiger (2011), one can think of the agreement as an incomplete contract containing "gray areas" about the circumstances in which exceptions should be granted. As a result, the Home government might well deem its choice of tariff as perfectly legal. At the same time, however, the Foreign government is dissatisfied with Home's policy change and has decided to use coercion to reduce Home's tariff, i.e., it threatens to increase τ_a^* if Home does not implement a new trade policy $\tau < \tau_0$.¹² More generally, since our focus in this paper is on the strategic aspects that underlie trade disputes themselves, irrespective of the events that prompted them, we will follow Rosendorff (2005) and treat the status quo policies $(\tau_0, \tau_0^*) = (\tau_0, \tau_a^*)$ as given.

¹⁰This assumption is totally innocuous. Indeed, additive separability of the governments' preferences ensures that the Foreign trade-war tariff does not depend on Home's preferences and, therefore, that the Home government's actions would not imply any relevant belief updating about α . Thus, including uncertainty about α would add nothing but notation: when making demands, the Foreign government would anticipate the expected (as opposed to the known) response from the Home government. Moreover, our results remain intact if we assume that the Foreign government's set of types is a finite set.

¹¹See, for example, the debate around the 30 percent US steel tariff introduced by the Bush administration in 2002 (see, e.g., Rosendorff 2005).

¹²Note that since we are interested in developing a model of coercion as opposed to bargaining, we do not allow the Foreign government to use its tariff as a bargaining instrument when it formulates its demand. In other words, in this model, as in Bagwell, Mavroidis, and Staiger (2007), the import-competing sector is only a "retaliation-good sector" in the sense that the Foreign government can only use its tariff τ^* as a retaliation instrument when coercion is unsuccessful.

The sequence of events that follow the realization of γ depends on Foreign's institutional arrangements for trade coercion:

- (i) *Absence of IO Membership.*—Suppose first that Foreign is not a member of the IO, so that coercion must be unilateral. In this case, the Foreign government threatens to increase its tariff unless the Home government acquiesces to a demand $\tau \leq \tau_0$. If the Home government concedes, reducing its tariff from τ_0 to τ , then Foreign does not impose any sanction. Then the policy vector (τ, τ_0^*) is implemented. If the Home government stands firm, then the Foreign government carries out its threat, thereby triggering a trade war.
- (ii) *Full Commitment to the IO.*—Suppose now that Foreign is a member of the IO and is fully committed to its dispute settlement process—so that coercion must be conducted multilaterally. The process through which disputes are settled in international trade organizations is usually long and complex. It typically involves consultations between the sender and target (and potentially third parties and/or mediators) to reconcile their differences by themselves, IO panels' hearings and parties' rebuttals, several reports from the IO panel to the parties and, in the absence of an early settlement, rulings and appeals. Our aim here is to focus on the effects of incomplete information on multilateral negotiation outcomes and, therefore, to abstract away from any other complexity that such a situation might entail. To this end, and to ease comparison with the previous framework, we model proceedings as follows.

First, both parties observe the realization of the IO panel's "interpretation" of the trade agreement, τ^{io} . The Foreign government then makes a demand $\tau \leq \tau_0$. The Home government can concede to this demand (ending the game with the implementation of the policy pair (τ, τ_0^*)) or reject it. In the latter case, the IO issues ruling τ^{io} . The Home government reacts to the ruling in one of two ways: compliance (ending the game with the implementation of policy pair (τ^{io}, τ_0^*)) or noncompliance. If it fails to comply with the ruling, then the IO authorizes Foreign to retaliate and a trade war ensues.

Although this is a highly abstract version of GATT-WTO proceedings, it contains all the elements needed to study the impact of incomplete information and IO membership on trade coercion, which is the main focus of the present paper.

- (iii) *Partial Commitment to the IO.*—As explained in the introduction, it is interesting to consider also an intermediate case in which the Foreign government initially decides whether to coerce unilaterally or multilaterally. The remainder of the game is as in (i) if it chooses to coerce unilaterally, and as in (ii) otherwise. This setting captures, for instance, the working of the Section 301 provision of the 1974 US Trade Act, under which action on a dispute could be unilateral or accompanied by a GATT-WTO complaint (e.g., Busch and Reinhardt 2000, Pelc 2010).

Before we proceed with the analysis, we need to discuss three of the assumptions of the model. First, we treat the IO ruling τ^{io} as exogenous. Note that our main goal is to study how countries' commitment to international dispute settlement mechanisms affect trade coercion outcomes. Consequently, in our model, the IO dispute settlement process is taken as given. The value of τ^{io} can simply be interpreted as the governments' (common) expectations about the ideal ruling of the decisive IO-panel member. More specifically, one can think of the IO as an organization with its own social welfare function (maximized by τ^{io}), which is unaffected by the political pressure from domestic actors.¹³ Beyond intrinsic policy preferences concerning the current situation, this objective function may also be influenced by other external factors—e.g., consistency with previous rulings and setting precedents in anticipation of potential future disputes. A second assumption of the model is that, once the Foreign government has filed a complaint with the IO, it always complies with the IO ruling and empirical evidence supports this view. In fact, as observed by Pelc (2010), "... once the United States began GATT proceedings, it did not turn back to unilateralism." In particular, the United States never retaliated unilaterally nor threatened to do so after a panel decision was reached. Finally, we assume that, even in the case of full commitment to the IO, noncompliance to a ruling leads to a trade war. This evidently does not mean that the IO falls apart whenever a defendant spurns its ruling. In reality, the WTO only authorizes the complainant to retaliate on a noncomplying defendant within certain limits.¹⁴ However, even such constrained retaliatory trade sanctions might cause the target to retaliate in turn, leading to escalation into further sanctions. We thus assume—for simplicity—that a trade war follows noncompliance. It is important to note though that all of our qualitative results carry over to alternative settings with constrained retaliation. In fact, consider the case in which the IO imposes a limit, say Δ , on the increase in the Foreign tariff, to capture the idea contained in World Trade Organization (2002) art. 22.4 that "... the level of the suspension of the concessions or other obligations authorized by the Dispute Settlement Body shall be equivalent to the level of the nullification or impairment." In this case, unsuccessful multilateral coercion would lead to a "constrained trade war," in which the type- γ Foreign government would implement a tariff of $\tilde{\tau}^*(\gamma) \equiv \min\{\tilde{\tau}^*(\gamma), \tau_0^* + \Delta\}$ (instead of $\tilde{\tau}^*(\gamma)$). Nothing in the logic of the

¹³ The growing judicialization of the WTO has led to the introduction of an Appellate Body, made up of independent legal experts, selected to deal with all the disputes that will arise in their four-year terms. Several observers have argued (e.g., Zangl 2008) that its rulings are less likely to be biased in favor of one of the parties involved than under the preexisting GATT arrangements.

¹⁴ A well-known case in which the complainant accepted the WTO ruling, whereas the defendant did not, is the Beef-Hormone dispute between Canada and the United States on the one hand and the EU on the other. In 1989, the EU banned the importation of meat containing six artificial growth hormones approved instead for use in the United States. Under WTO rules, this type of ban is allowed, but only if the part introducing the ban is able to provide valid scientific evidence that the measure had been introduced because of a health and safety concern. In 1997, the Dispute Settlement Body ruled against the EU, and the ruling was confirmed by the Appellate Body in 1998. As the EU did not initially comply, an arbitrator appointed by the WTO Dispute Settlement Body authorized the United States and Canada to impose, respectively, a tariff of US\$116.8 and US\$8 million per year on the EU (Kerr and Hobbs 2002). The controversy on the matter is still ongoing. In September 2009, the EU and the United States signed a memorandum of understanding, which established a new duty-free quota for grain-fed, high-quality beef as a compromise. As of December 2016 though, based on continued concerns on US beef access to the EU market, the United States has taken steps to reinstate retaliatory tariffs.

arguments we will develop in Section III would be affected, since all the key incentives would be preserved: the Home government would still be better off facing low rather than high Foreign government types, and trade wars would still be more damaging to low than to high types. We return to this in more detail when we provide intuitions for our results (see footnotes 17 and 20). As we will see, what matters for the IO to have an impact on coercion outcomes is that it allows the Home government to secure the tariff τ^{io} when it believes that the Foreign government's type is high.

Each variant of the model describes a sequential game of incomplete information. We solve it by looking for (pure strategy) perfect Bayesian equilibria, which are defined as follows: (i) the Home government's beliefs are generated by Bayesian updating whenever possible, and (ii) in each stage, governments' actions are optimal, given their beliefs and their opponents' strategies. In order to eliminate equilibria, which rely on implausible beliefs off the equilibrium path, we use criterion D1 from Cho and Kreps (1987). Intuitively, this refinement requires that if the set of Home government's actions that make some foreign government's type γ willing to deviate is strictly smaller than the set of actions that make some other type γ' willing to deviate, then the Home government should believe that type γ' is infinitely more likely to deviate than γ is.¹⁵ In the remainder of the paper, any reference to an "equilibrium" is to a perfect Bayesian equilibrium consistent with criterion D1.¹⁶

II. International Trade Institutions and Coercion Outcomes

In this section, we characterize in turn the equilibria that will emerge from the three institutional settings described in the previous section.

A. Benchmark: Coercive Trade Policy in the Absence of the IO

Both because it is empirically relevant and because it provides a benchmark to compare outcomes with those possible when the Foreign government can coerce multilaterally, we start by analyzing the case in which the Foreign country is not a member of the IO.

Trade Wars and Reservation Demands.—To solve the game, we begin with the last stage in which the two governments engage in a trade war. Although this continuation game may involve the presence of asymmetric information, it always has a unique equilibrium outcome: the Home government adopts its ideal tariff $\tilde{\tau}$, irrespective of the Foreign government's policy choice; likewise, the type- γ Foreign government adopts its ideal tariff $\tilde{\tau}^*(\gamma)$, irrespective of the Home government's policy choice.

¹⁵ This is a strengthening of the Intuitive Criterion, which has no bite in this game. See the supplementary online Appendix for the formal definition.

¹⁶ In order to limit the number of possible cases (without affecting the paper's conclusions), we also assume that in case of a tie, a player will prefer to agree than to disagree with the other player or the IO.

Given the outcome of a trade war, consider now the Home government's decision of whether to concede to the Foreign government's demand τ . Suppose that its beliefs about γ are given by some cdf F . It will concede to demand τ if and only if its payoff from conceding exceeds its expected payoff from triggering a trade war; that is

$$(7) \quad W(\tau, \tau_0^*) \geq \int_{\underline{\gamma}}^{\bar{\gamma}} W(\tilde{\tau}, \tilde{\tau}^*(\gamma)) dF(\gamma).$$

Let the smallest value of τ that satisfies the above inequality be denoted by $T(F)$. This is the Home government's "reservation demand," or the minimum demand it will accept rather than engage in a trade war. In what follows, we will sometimes indulge in a slight abuse of notation and denote by $T(\gamma)$ the Home government's reservation tariff when its beliefs assign probability 1 to type γ . Similarly, the type- γ Foreign government's reservation demand $T^*(\gamma)$ —that is, the Home tariff at which the Foreign government is indifferent between settling and engaging in a trade war—is defined as the largest value of tariff τ that satisfies

$$(8) \quad W^*(\tau, \tau_0^*, \gamma) \geq W^*(\tilde{\tau}, \tilde{\tau}^*(\gamma), \gamma)$$

(recall that $W^*(\tau, \tau_0^*, \gamma)$ decreases as τ increases).

It can be easily shown that $T(\gamma)$ and $T^*(\gamma)$ are both strictly decreasing in γ . An increase in γ causes the trade-war tariff of the Foreign government, $\tilde{\tau}^*(\gamma)$, to rise. As $W(\tilde{\tau}, \tilde{\tau}^*(\gamma))$ decreases with $\tilde{\tau}^*(\gamma)$ (and therefore with γ), the Home government is willing to implement a lower tariff to avoid a trade war. In contrast, applying the Envelope Theorem reveals that $W^*(\tilde{\tau}, \tilde{\tau}^*(\gamma), \gamma)$ increases with γ , so that greater political pressure from its import-competing sector makes the Foreign government less willing to tolerate high tariffs applied by Home.¹⁷ We assume throughout our analysis that $T^*(\underline{\gamma}) < T(\underline{\gamma})$. Intuitively, this means that, facing the "weakest" type of Foreign government, the Home country is only willing to make small compromises that are not even acceptable to that type of Foreign government: the latter's level of resolve is too high for it to tolerate Home tariffs above $T(\underline{\gamma})$ but is at the same time too low for the Home government to concede to demands below $T^*(\underline{\gamma})$.¹⁸ As explained in Section IV, this assumption is made purely for technical convenience as it allows us to avoid equilibrium existence issues that arise if $T^*(\underline{\gamma}) > T(\underline{\gamma})$.

The Ineffectiveness of Unilateral Coercion.—Can the Foreign government obtain a concession from the Home government in equilibrium? This question is answered in the following.

¹⁷ It is readily checked that these key properties of $T(\gamma)$ and $T^*(\gamma)$ would still hold if the IO imposed a cap $\tau_0^* + \Delta$ on the Foreign government's retaliatory tariff, which would then be equal to $\zeta^*(\gamma) \equiv \min\{\tilde{\tau}^*(\gamma), \tau_0^* + \Delta\}$ for each type γ . As $\zeta^*(\gamma)$ may be constant on some interval of types, the resulting reservation-tariff function for the Home government, $T(\gamma)$, would only be weakly decreasing. However, as $W^*(\tilde{\tau}, \zeta^*(\gamma), \gamma)$ is strictly increasing in γ , the Foreign reservation tariff, now denoted by $T^*(\gamma)$, would remain strictly decreasing.

¹⁸ Note that we only need this condition to hold when $\gamma = \underline{\gamma}$, which is a measure-zero event.

PROPOSITION 1: *Suppose that there is no IO—so that coercion must be unilateral. There exists an equilibrium, and in any equilibrium, the Foreign government always fails to obtain a concession from the Home government.*

PROOF:

See Appendix A. ■

To understand the intuition for this result, note that upon observing the demand τ by the Foreign government, the Home government—uninformed about the level of political pressure γ that has emerged in the Foreign country—updates its beliefs. Given these new beliefs, say F , it concedes to τ if and only if $\tau \geq T(F)$. As its reservation demand $T(\gamma)$ is decreasing in γ , the best strategy for the Foreign government is to signal high values of γ by requiring a low level of τ . Indeed, trade wars are less costly to Foreign governments that are very sensitive to the well-being of the import sector (characterized by a high- γ)—the Foreign reservation demand $T^*(\gamma)$ decreases with γ —and the Foreign government is therefore more likely to risk a trade war when γ is large. Understanding this, the Home government rationally infers higher values of γ from a demand for a lower tariff. Such beliefs lead the foreign policymaker to go too far, however, and to make requests which the Home government is not prepared to meet. This signaling spiral leads all types of Foreign government to make unsuccessful demands, and a trade war will ensue in every equilibrium.

More specifically, it is straightforward to see that there is always a separating equilibrium in which no concession can be obtained. In such an equilibrium, all types of Foreign government make different unacceptable demands (i.e., demands below $T(\gamma)$) and the Home government believes that any deviation is the action of the weakest type $\underline{\gamma}$. This implies that the Home government would only be prepared to concede to demands $\tau > T(\underline{\gamma})$. As the Foreign reservation demand $T^*(\gamma)$ decreases with γ (so that $T^*(\gamma) \leq T^*(\underline{\gamma}) < T(\underline{\gamma})$), deviating to such demands is not profitable for any type of Foreign government. In addition, an equilibrium in which concessions may arise would have to be pooling—otherwise, the Foreign types that do not obtain concessions or only modest concessions would mimic the type(s) which obtain the best concession. What remains to establish, therefore, is that concessions in a pooling equilibrium are impossible. Assume toward a contradiction that a subset of types, say $[\underline{\gamma}_0, \bar{\gamma}_0] \subseteq [\underline{\gamma}, \bar{\gamma}]$, obtain a concession τ from the Home government in some equilibrium.¹⁹ This implies that $T^*(\bar{\gamma}_0) > \tau \geq T(F') > T(\bar{\gamma}_0)$, where F' represents the updated beliefs of the Home government conditional on observing demand τ . (The last inequality follows from the fact that the Home government is prepared to make more concessions if it believes with certainty that Foreign's type is $\bar{\gamma}_0$ than if it believes that any type in $[\underline{\gamma}_0, \bar{\gamma}_0]$ is possible.) Now consider a deviation from a request τ to a smaller tariff τ' , running the risk to trigger a trade war. As

¹⁹The assumption that this subset is an interval is only made for expositional convenience. The proof of the proposition considers all possible subsets of $[\underline{\gamma}, \bar{\gamma}]$.

higher types have stronger incentives to run such a risk, it turns out that reasonable beliefs eliminate all types smaller than $\bar{\gamma}_0$. Being convinced that it faces a more resolved Foreign government, the Home government is thus better off conceding to the new demand τ' if it is sufficiently close to τ . This in turn makes the deviation to $\tau' < \tau$ profitable for the type- $\bar{\gamma}_0$ Foreign government: a contradiction.

B. Coercive Trade Policy with Full Commitment to the IO

We now turn to the analysis of the consequences of full commitment to the IO on trade coercion outcomes. One of the questions this paper seeks to answer is how international trade institutions, despite their lack of enforcement power, can be effective in settling disputes. We have just shown how the logic of unilateral trade coercion locks the Foreign government into signaling spirals leading to trade wars. Despite being unable to enforce its rulings, can the IO's dispute settlement process do a better job of obtaining concessions from the Home government?

The answer is positive, and the intuition is that full commitment to the IO's dispute settlement process may offer the Foreign government an opportunity to break the spiral of unilateral coercion. To see how this can occur in equilibrium, suppose that $\tau^{io} \geq T(F_0)$. Consider first the stage in which the Home government must decide whether or not to comply with the IO ruling τ^{io} . Failure to comply would trigger a trade war. Therefore, it follows from the analysis of the trade-war stage we have developed in the previous section that it chooses to comply if and only if $\tau^{io} \geq T(F)$, where the cdf F stands for the updated beliefs about the Foreign government's type γ at this stage. This implies that, when confronted with some demand τ from the Foreign government, the Home government's optimal strategy is to concede if and only if $\tau \geq \max\{\tau^{io}, T(F)\}$. As long as $T(F) > \tau^{io}$, the same signaling incentives as under unilateral coercion drive the Foreign government to deviate from successful demands by asking for lower tariffs, which signal high values of γ and reduce the Home government's reservation tariff $T(F)$. When $T(F) \leq \tau^{io}$, however, the Home government's reservation demand becomes equal to τ^{io} and cannot be reduced any further. Believing that the Foreign government's type is high, the Home government can secure the tariff τ^{io} instead of risking a costly trade war. Thus, the presence of the IO mitigates the connection between the Foreign government's type and the trade-war payoffs. In particular, if all types of Foreign government demand τ^{io} , then the Home government's beliefs (correctly derived from Bayes' rule) must be given by F_0 . As $\tau^{io} \geq T(F_0)$, the Home government concedes to this demand and a trade war is avoided. Since $W^*(\tau, \tau_0^*, \gamma)$ strictly decreases with τ for all γ , it follows that no type of Foreign government can profitably deviate: the Home government would always be better off implementing the tariff τ^{io} than conceding to lower demands. We conclude that a trade war can be avoided if $\tau^{io} \geq T(F_0)$.²⁰

²⁰Note that the logic of this argument would remain unaltered if $T(\gamma)$ and $T^*(\gamma)$ were replaced by the reservation-tariff functions $\mathcal{T}(\gamma)$ and $\mathcal{T}^*(\gamma)$, defined in footnote 17 for the case where the Foreign government's retaliatory tariff is constrained by the IO. In particular, as $\mathcal{T}^*(\gamma)$ is a strictly decreasing function—i.e., trade wars are less damaging to the Foreign government as its type increases—signaling incentives and the IO ruling would play the same role as described above.

Our next result shows that the condition $\tau^{io} \geq T(F_0)$ is also necessary for a trade war to be avoided. If τ^{io} is too low then, as in the case of unilateral coercion, the Foreign government's demands spiral down to unacceptable levels leading to a trade war.

PROPOSITION 2: *Suppose the Foreign government is fully committed to the IO—so that coercion must be multilateral. There always exists an equilibrium, and the following is true in any equilibrium:*

- (i) *If $\tau^{io} \geq T(F_0)$, then: either all types of Foreign government obtain the concession τ^{io} , or they all make unsuccessful demands following which the Home government complies with the IO ruling.*
- (ii) *If $\tau^{io} < T(F_0)$, then all types of Foreign government make unsuccessful demands following which the Home government fails to comply with the IO ruling.*

PROOF:

See Appendix B. ■

Combined with Proposition 1, Proposition 2 shows that an IO can affect the outcome of trade coercion and prevent trade wars, even though it has no enforcement power. It also suggests a possible explanation for why trade coercion appears to be more effective in obtaining concessions from target governments when conducted multilaterally. We will elaborate on the empirical and normative implications of the equilibrium analysis in Section III.

C. Coercive Trade Policy with Partial Commitment to the IO

Under partial commitment to the IO, the Foreign government is allowed to choose whether to coerce the Home government unilaterally or multilaterally. Suppose that $\tau^{io} \geq T(F_0)$, and that the Foreign government's type γ satisfies $T(\gamma) < \tau^{io} < T^*(\gamma)$, so that both countries are better off implementing τ^{io} than engaging in a trade war. Our analysis so far reveals that the signaling incentives inherent in unilateral coercion would lead the Foreign government to make inefficient demands to the Home government. To avoid this outcome, the type- γ Foreign government would therefore be expected to adopt multilateral coercion. Some authors argue, however, that taking a trade dispute to an IO signals a lack of resolve—i.e., a low γ —by the sender government (e.g., Reinhardt 2000, Pelc 2010). The next proposition provides a formalization of their argument.

PROPOSITION 3: *Suppose the Foreign government is only partially committed to the IO—so that it can choose between unilateral and multilateral coercion. There exists an equilibrium in which all types of Foreign government coerce unilaterally and fail to obtain a concession. In addition, a trade war arises with probability one in any equilibrium.*

PROOF:

See Appendix C. ■

In other words, partial commitment to the IO yields the same outcome as absence of membership: in both cases, the Foreign government fails to obtain a concession from the Home government, and a trade war ensues.

Note though that the Foreign government's coercive policy has now two components: the demand τ and the method of coercion (unilateral versus multilateral) through which this demand is made. A deviation from multilateral to unilateral coercion in this case conveys the same signal as a deviation to a lower tariff demand in the absence of an IO: the Home government therefore anticipates tougher retaliatory measures in case of a trade war. As in the unilateral-coercion game, such beliefs induce the Home government to concede to lower unilateral demands. This in turn drives the Foreign government to (unilaterally) ask for even lower tariffs until its demands become unacceptable.

These incentives to coerce unilaterally to signal high resolve can only disappear when in equilibrium all types of Foreign government make unsuccessful demands (either unilaterally or multilaterally), thus leading to a trade war. In this case, the Home government interprets any deviation by its foreign counterpart as an attempt to escape this outcome and, consequently, infers that the Foreign government's type γ must be low. It is therefore optimal for the Home government to only accept demands so high that the Foreign government prefers to engage in a trade war.

III. Implications

Our theoretical model provides novel insights on the influence of international trade institutions on coercion outcomes. Importantly, our results are consistent with the stylized facts that have been uncovered in the existing empirical literature. In this section, we briefly review these empirical findings and explain how they relate to our analysis.

Unilateral versus Multilateral Coercion: The Influence of International Trade Institutions.—Busch and Reinhardt (2000) observes that, during the GATT period, only two-fifths of the rulings in favor of the complainant resulted in full compliance by the defendant—whereas in nearly a third of the cases, defendants failed to comply at all. Even though the establishment of the WTO dispute settlement mechanism improved the situation, as Rossmiller (1994) pointed out, the WTO remains a “court with no bailiff.” These observations prompt the following question: can a multilateral institution influence coercion outcomes despite its lack of enforcement power? Empirical evidence uncovered by Pelc (2010) suggests that this is indeed the case. Focusing on the US experience between 1975 and 2000, he finds that disputes that went through the GATT, rather than relying only on Section 301, are 34 percentage points more likely to result in a concession.

Pelc (2010) suggests that it is the perceived illegitimacy of unilateral coercion and the importance of reputation which decrease the likelihood of a target conceding. While resistance to institutionally constrained demands entails the reputational

cost of being branded a violator, resistance to unilateral threats—regarded as illegitimate by the rest of the world—yields a reputational benefit: it decreases the likelihood of being unilaterally targeted again in the future. Our formal analysis provides an alternative rationale, which focuses on the role played by the sender government's incentives. On the one hand, unilateral coercion creates signaling spirals leading the sender government to make unacceptable demands. On the other, commitment to a multilateral organization can break these spirals and allow the sender government to obtain concessions.

Early Dispute Settlements.—Analyzing evidence on more than 600 GATT-WTO disputes from 1948 through 1999, Busch and Reinhardt (2000) observes that in a majority of cases (about 55 percent), no panel was ever established, and a further 8 percent of them ended prior to the issuance of a panel report. Paraphrasing them, a key question is why should target governments settle early given that they can spurn adverse rulings with impunity. They argue that the source of early concessions lies in the normative power of GATT-WTO rulings and in the pressure to abide by the norm: an adverse ruling may weaken the target government's political position in its own country, as well as its position in ongoing multilateral trade talks. As a result, if the target government is uncertain about the IO ruling, then it may prefer to concede beforehand.

Consistently with the evidence, Proposition 2(i) shows that pre-ruling settlements may occur in equilibrium.²¹ Importantly though, in our setting, the mechanism at work is different: IO rulings do not convey any normative or reputational costs. When the Foreign government anticipates a “high tariff” IO ruling (i.e., when $\tau^{io} \geq T(F_0)$), it expects the Home government to comply with this ruling. This leads the Foreign government to abandon aggressive strategies and to make more accommodating demands to which the Home government is willing to concede.²² Thus, it is mainly the sender government's (rather than target's) incentives which are affected by the prospect of the IO decision.

An Alternative Rationale for International Trade Agreements.—Our model suggests a possible explanation for another empirical puzzle: given that membership in an international trade organization may limit the (coercive) policy discretion of a national government, why would the latter choose to join a supranational body? Most of the existing literature on this topic suggests that states become members of such institutions to solve the coordination problem created by the terms of trade externality from tariffs (e.g., Bagwell and Staiger 1999). Our analysis reveals, however, that another driving force may emanate from informational asymmetries in trade coercion. By helping to explain why demands channeled through the multilateral system may be more successful than unilateral demands, our model provides a new rationale for states' commitment to multilateral institutions.

²¹ The proof of Proposition 2 (Section B in the Appendix) shows that something even stronger is true if $\tau^{io} > T^*(\bar{\gamma})$: in all equilibria, the Foreign government obtains an early concession from the Home government.

²² In fact, Busch and Reinhardt (2000) points out that among those disputes ending prior to a ruling, 67 percent exhibit full or partial concession by the target government.

To see this, suppose that we add an initial stage to the game in which the Foreign government decides whether or not to fully commit to the IO. If $\tau^{io} < T(F_0)$, then it is indifferent between all institutional arrangements: a trade war is inevitable. Suppose instead that $T(F_0) \leq \tau^{io} \leq T^*(\gamma)$. An immediate corollary of Propositions 1–3 is that, in this case, the Foreign government is better off fully committing to the IO.

The Role of Commitment to International Organizations.—Proposition 3 shows that institutions allowing sender governments to choose between unilateralism and multilateralism can reduce the effectiveness of coercion. A leading historical example of the coexistence of these two coercion methods is represented by Section 301 of the US Trade Act of 1974. This provision allowed the United States to take a number of unilateral retaliatory actions against any foreign measures deemed to violate existing agreements or otherwise impeding its interests. At the same time, the United States retained access to the dispute settlement system provided by the GATT-WTO (Pelc 2010).

As argued by Pelc (2010), the availability of unilateral coercion did not deliver the expected results, and in fact, the United States “ultimately found it in its interest to . . . push for greater formal constraints in the Uruguay Round that ultimately raised the costs of unilateralism further.” In our model, if we allowed the Foreign government to choose between full and partial commitment to the IO, then it would strictly prefer the former whenever $T(F_0) \leq \tau^{io} < T^*(\gamma)$. The Foreign government would indeed be better off making a successful demand τ^{io} under full commitment (Proposition 2) than making an unsuccessful demand under partial commitment (Proposition 3). Unlike Pelc’s explanation based on the illegitimacy of unilateral coercion, our result though stems from the Foreign government’s strategic incentives created by the presence of a unilateral option. Even though the Foreign government would be better off if this option were not available, incentives to signal higher levels of resolve to the Home government by deviating from multilateral to unilateral coercion eventually lead the Foreign government to make unacceptable demands (Subsection 3.3). These incentives are reminiscent of Reinhardt’s (2000) observation that taking a dispute to the GATT was a signal of the complainant’s lack of resolve.

IV. Discussion and Concluding Remarks

Summary.—This paper is a first attempt at analyzing the strategic interactions that underlie coercive trade policy. We have studied trade coercion in settings where sender governments may show their resolve by demanding more concessions from target governments. We have seen how the temptation to exaggerate can reduce the likelihood of targets conceding. This problem is especially severe when the sender government is not (fully) committed to a multilateral dispute settlement mechanism. Then, unbound by international commitments, the sender may make excessive demands which are unacceptable to the target. Institutions through which demands are channeled thus matter to coercion outcomes. In accordance with empirical evidence, our results indicate that full commitment to (even weak) multilateral trade

institutions makes trade coercion more effective in obtaining concessions from target governments.

Mixed Strategies and the $T^(\underline{\gamma}) > T(\underline{\gamma})$ Case.*—The intuitions behind Propositions 1 and 2, as well as their proofs, rest on two assumptions: firstly, as we focus on pure strategy equilibria, the Home government is not allowed to randomize between conceding and not conceding to the Foreign government demands; and secondly, we concentrate on cases where $T^*(\underline{\gamma}) < T(\underline{\gamma})$. Though these assumptions ease the exposition, make our results sharper, and allow us to avoid equilibrium existence issues, we show in the supplementary online Appendix that, without them, our main conclusions remain intact. Indeed, allowing the governments to use mixed strategies while still assuming that $T^*(\underline{\gamma}) < T(\underline{\gamma})$ leaves the results unchanged. In the absence of the IO, a trade war is the only possible equilibrium outcome: an equilibrium in which the Foreign government obtains a concession with a positive probability would require some subset of types (including $\underline{\gamma}$) to pool and, by the same logic as in the case of pure strategy equilibria, incentives to signal high levels of resolve would then lead the highest types in that subset to (profitably) deviate by successfully demanding lower tariffs. Such profitable deviations become impossible in the presence of the IO if all types of Foreign government demand $\tau^{io} \geq T(F_0)$, as the Home government can always comply with the IO ruling. The possibility of randomization does not affect the strategic incentives that underlie Propositions 2 and 3, which remain unchanged—except in the knife-edge case where τ^{io} is exactly equal to $T(F_0)$.²³

Assuming $T^*(\underline{\gamma}) > T(\underline{\gamma})$ creates equilibrium existence problems: there does not exist a pure strategy equilibrium in the model without IO, and existence of a mixed-strategy equilibrium requires additional conditions on the W and W^* functions (discussed in the supplementary online Appendix S3). Though less extreme than in the $T^*(\underline{\gamma}) < T(\underline{\gamma})$ case, the conclusion from the analysis of mixed-strategy equilibria for the model without the IO is still a negative one: a trade war arises with positive probability in any equilibrium.²⁴ More specifically, signaling incentives again rule out successful demands by pooling types, and therefore, every equilibrium must have the following (partially) separating structure: there exists a threshold $\hat{\gamma} \in (\underline{\gamma}, \bar{\gamma}]$ for the Foreign government's type such that every type $\gamma < \hat{\gamma}$ makes demand $T(\gamma)$ which the (indifferent) Home government concedes to with a probability $\alpha(\gamma)$, where $\alpha(\gamma)$ is a strictly decreasing function, and every type $\gamma > \hat{\gamma}$ ends up in a trade war with probability one. It follows that a trade war cannot be avoided with certainty unless the realization of the Foreign government's type is exactly equal to $\underline{\gamma}$, which is a probability-zero event.²⁵ The analysis of the equilibria for the model with the IO generates the same conclusion as in the $T^*(\underline{\gamma}) < T(\underline{\gamma})$ case: if $\tau^{io} > T(F_0)$, then τ^{io} is the only possible outcome and a trade war never

²³If $\tau^{io} = T(F_0)$, then there may also be equilibria in which the Home government, indifferent between conceding to τ^{io} and a trade war, chooses a trade war with positive probability.

²⁴Of course, there cannot be an equilibrium in which all types of Foreign government end up in a trade war if $T^*(\underline{\gamma}) > T(\underline{\gamma})$, as type $\underline{\gamma}$ could profitably deviate by making a successful demand in $(T(\underline{\gamma}), T^*(\underline{\gamma}))$.

²⁵In the knife-edge case where $T^*(\underline{\gamma}) = T(\underline{\gamma})$, there can also be a pure strategy equilibrium in which type $\underline{\gamma}$ is the only type of Foreign government that obtains a concession with positive probability.

occurs in equilibrium. The basic intuition behind this result is that: (i) demanding lower tariffs to signal high levels of resolve is no longer a profitable deviation from a pooling equilibrium in which all types of Foreign government demand $\tau^{io} \geq T(F_0)$, as the Home government is better off complying with the IO ruling rather than conceding to lower tariff; and (ii) equilibria with a partially separating structure as above cannot exist in the presence of the IO, as the (partially) informed Home government prefers complying to its ruling $\tau^{io} > T(F_0)$ rather than engaging in a trade war with the highest types of Foreign government. If the Foreign government can choose between unilateral and multilateral coercion, then we obtain a counterpart of Proposition 3: in any equilibrium, a trade war occurs with positive probability whenever the Foreign government's type exceeds $\underline{\gamma}$ (which is itself a probability-one event). We thus conclude that, as in the $T^*(\underline{\gamma}) < T(\underline{\gamma})$ case, the IO can be effective in preventing trade wars despite its lack of enforcement power, and in particular, the Foreign government is more likely to obtain concessions from the Home government by using multilateral coercion rather than unilateral coercion. Moreover, allowing the Foreign government to choose between these two modes of coercion can make coercion less effective.

Beyond Trade Coercion.—Proposition 1 is reminiscent of the unraveling result obtained in disclosure games—e.g., Grossman (1981) and Milgrom (1981).²⁶ Interestingly, however, while the unraveling phenomenon in disclosure games results from the privately-informed agents' ability to reveal their types directly, the sort of unraveling of pooling equilibria in our unilateral-coercion game follows from the Home government's *inferences* from the Foreign government's actions. As we saw in Section III, criterion D1 implies that, in a putative pooling equilibrium, the Home government interprets deviations from a successful demand as certainly coming from the highest pooling type, thus allowing the latter to effectively disclose its private information (and profitably deviate) by making more aggressive demands. Complete breakdowns have been shown to occur in other bargaining processes with adverse selection, but the agents' failure to come to agreement (irrespective of their types) in those alternative settings stems from different mechanisms—e.g., rational expectations in Milgrom and Stokey (1982) or sequential offers in Vincent (1989).

Note that the logic of Proposition 1 can be extended to any bargaining or contracting environments beyond trade dispute settlement that possess the following key properties: (i) a privately-informed party (the “sender”) makes a take-it-or-leave-it offer τ to another party (the “receiver”), which determines how the surplus generated by an agreement would be split between the two parties; (ii) the receiver's benefit from any agreement is monotonic, say increasing, in the sender's type, γ ; and (iii) the relative cost to the sender of deviating from a successful offer τ to a more demanding offer τ' —measured by the ratio

$$\frac{[\text{Sender's payoff under agreement } \tau] - [\text{Sender's payoff if no agreement}]}{[\text{Sender's payoff under agreement } \tau'] - [\text{Sender's payoff if no agreement}]}$$

²⁶ We thank an anonymous referee for drawing our attention to the analogy between the two results.

—is decreasing in her type. It is the latter condition, coupled with criterion D1, that leads to the “unraveling” of the pooling PBEs in which offers would be successful: as the highest pooling type has the lowest relative cost of deviating from τ to τ' , the receiver is convinced that a deviation to τ' must come from this type and, therefore, accepts τ' . This makes the deviation profitable to the highest pooling type. Property (iii) typically holds in bargaining situations (lemons markets, international conflict resolution, ...) where the proposer’s outside option improves with its type.

Research Avenues.—There are a number of research avenues opened up by our results, three of which we will briefly discuss. First, our positive theory of the impact of multilateral institutions on trade coercion outcomes naturally prompts a normative question: what would an optimal dispute settlement mechanism be in the presence of informational asymmetries?²⁷ Answering this question would require a richer framework, i.e., one that would further our understanding of the effects of settlement mechanisms both on membership in international trade institutions and on target governments’ policy choices that are likely to trigger coercive responses.

As we noted at the outset (see footnote 12), our analysis focused on coercion itself and not on its ultimate origin. It would be interesting to investigate why do dissatisfied governments use coercion instead of potentially more efficient bargaining approaches. Trade coercion typically involves two policy instruments: the target’s trade policy, which is the source of the sender’s discontent, and the sender’s policy, which is only used as a retaliation instrument. By focusing its demand on the former instrument, the sender government leaves out mutually advantageous agreements, which would be available if its demand would involve instead a combination of both instruments.

Finally, the mechanisms of trade coercion put forward in this paper can be captured with our simple one-period model, in which sender and target governments do not have to anticipate further dealings with each other. Consider a dynamic extension of this model in which, whenever the target would either concede to the sender’s demand or comply to the IO ruling, the same one-period game would be repeated. If the sender’s type were persistent across periods, as in Bagwell (2009), then governments would have to anticipate the future consequences of their choices. In particular, this would exacerbate the signaling incentives that arise in the one-period model, especially in cases where there is no IO or the sender is only partially committed to the IO: the possibility of future interactions would increase the sender’s expected benefits from signaling high levels of resolve with more aggressive (unilateral) demands. In addition, repeated interactions would create reputational incentives for the target that could also lead to trade wars. Indeed, if the target’s type were also private information (and persistent over time), then it would have incentives to reject the sender’s demands, so as to be perceived as “tough” in future periods.²⁸ But we leave this as a topic for future research.

²⁷ Maggi and Staiger (2011) answers a similar question, but in a complete-information setting where states of the world are “vague” and subject to interpretation by contracting governments.

²⁸ Pelc (2010) informally discusses such incentives.

MATHEMATICAL APPENDIX

A. Proof of Proposition 1

To prove Proposition 1, we must show that: (i) in any equilibrium of the game without the IO, the Home government never concedes to the Foreign government's demands; and (ii) there exists an equilibrium in which the Home government never concedes to the Foreign government's demands.

CLAIM 1: *Suppose that the Foreign government can only coerce unilaterally. In any equilibrium, the Home government never concedes to its demands.*

PROOF:

First of all, observe that only one demand can successfully be made in equilibrium. To see this, suppose that two different demands τ_1 and τ_2 are made successfully in equilibrium by types γ_1 and γ_2 , respectively. Assume without loss of generality that $\tau_1 < \tau_2$. By definition of an equilibrium, type γ_2 must find it profitable to make successful demand τ_2 ; hence, $T^*(\gamma_2) \geq \tau_2 > \tau_1$. But this implies that type γ_2 could profitably deviate by making claim τ_1 : $W^*(\tau_1, \tau_0^*, \gamma_2) > W^*(\tau_2, \tau_0^*, \gamma_2)$.

Now, we establish the claim in two steps: (i) we first show that if a demand is successful in equilibrium, then it must emanate from a single type; and (ii) we then show that this is impossible in equilibrium:

- (i) We proceed by contradiction. Suppose that multiple types make a successful demand, say τ , in some equilibrium. From our initial observation above, all the other equilibrium demands are unsuccessful. Let $\Gamma_\tau \subseteq [\underline{\gamma}, \bar{\gamma}]$ be the set of types that demand τ , and let $\gamma^{\text{sup}} \equiv \sup \Gamma_\tau$ (observe that, by assumption, $\gamma^{\text{sup}} \in (\underline{\gamma}, \bar{\gamma}]$). By definition of a PBE, we must have $\tau \leq T^*(\gamma)$ for all $\gamma \in \Gamma_\tau$ —otherwise, some type in Γ_τ could profitably deviate by making an unacceptable demand—and, therefore, $\tau \leq T^*(\gamma^{\text{sup}})$. As $T^*(\gamma)$ is a strictly decreasing function, this implies that $\tau < T^*(\gamma)$ for all $\gamma < \gamma^{\text{sup}}$; so that all types $\gamma < \gamma^{\text{sup}}$ strictly prefer τ to a trade war. Hence, in equilibrium, all types $\gamma < \gamma^{\text{sup}}$ must make the unique successful demand τ . Furthermore, by definition of a PBE, all types $\gamma > \gamma^{\text{sup}}$ must prefer a trade war to τ : $\tau > T^*(\gamma)$ for every $\gamma > \gamma^{\text{sup}}$ (recall that indifferent types choose to avoid a trade war). By continuity of $T^*(\cdot)$, therefore, we must have $T^*(\gamma^{\text{sup}}) = \tau$. Being indifferent between τ and a trade war, the type- γ^{sup} Foreign government chooses τ . We have thus established that $\Gamma_\tau = [\underline{\gamma}, \gamma^{\text{sup}}]$.

Confronted with demand τ , the Home government—whose updated beliefs F_τ assign a probability of one to the event “ $\gamma \in [\underline{\gamma}, \gamma^{\text{sup}}]$ ”—optimally chooses to concede in the equilibrium under consideration. As the distribution of types has full support on $[\underline{\gamma}, \gamma^{\text{max}}]$, this implies that $\tau \geq T(F_\tau) > T(\gamma^{\text{sup}})$.

Now take any tariff $\tau' \in (T(\gamma^{\text{sup}}), \tau)$, and observe that no type of Foreign government demands τ' in equilibrium. Indeed, by definition, all types in Γ_τ demand $\tau \neq \tau'$. As for types γ outside Γ_τ , they must be greater than γ^{sup} .

Therefore, if type $\gamma > \gamma^{\text{sup}}$ demanded $\tau' > T(\gamma^{\text{sup}}) > T(\gamma)$, then the Home government would concede, thus contradicting our previous result that only one demand can be successful in equilibrium. All the premises of Lemma 1 in the supplementary online Appendix are thus satisfied: when confronted with demand τ' , the Home government believes that the Foreign government's type is lower than γ^{sup} with probability zero. As $\tau' > T(\gamma^{\text{sup}})$, the Home government concedes to demand τ' (off the equilibrium path). As $T^*(\gamma^{\text{sup}}) > \tau'$, this implies that demanding τ' is a profitable deviation for the type- γ^{sup} Foreign government, giving the desired contradiction. As a consequence, Γ_τ is either a singleton or an empty set.

- (ii) Suppose γ_τ is the unique type that makes a successful demand τ in some equilibrium. Bayesian updating implies that demand τ fully reveals the type of the Foreign government. Therefore, $T(\gamma_\tau) \leq \tau \leq T^*(\gamma_\tau)$ —otherwise either the Home government or the type- γ_τ Foreign government could profitably deviate from their equilibrium strategies. From our assumption that $T^*(\underline{\gamma}) < T(\underline{\gamma})$, this in turn implies that $\gamma_\tau \neq \underline{\gamma}$. Now take any type $\gamma < \gamma_\tau$. By assumption, a trade war occurs when the Foreign government is of type γ (γ_τ is the only type that makes a successful demand). As $T^*(\gamma) > T^*(\gamma_\tau) \geq \tau$, however, the type- γ Foreign government strictly prefers τ to a trade war. It could therefore profitably deviate by making the successful demand τ . Combined with (i), this proves that in any equilibrium all types of Foreign government make unsuccessful demands. ■

CLAIM 2: There exists an equilibrium of the game without the IO, in which the Home government never concedes to the Foreign government's demands.

PROOF:

Let k be a strictly positive number and consider the following strategy profile and beliefs: the type- γ Foreign government demands a tariff $\tau_k(\gamma) \equiv T(\gamma) - k$; the Home government's strategy when confronted with a demand τ is to concede if and only if $\tau \geq T(\underline{\gamma})$; it believes that the Foreign government is of type γ with probability one when confronted with demand $\tau_k(\gamma)$, for all $\gamma \in [\underline{\gamma}, \bar{\gamma}]$, and that it is of type $\underline{\gamma}$ when confronted with any other demand.

It is readily checked that the Home government's beliefs satisfy Bayes' rule whenever possible. By Lemma 2 in the supplementary online Appendix, they also satisfy Criterion D1. It is also readily checked that, given these beliefs, the Home government's strategy is a best response to the Foreign government's: given its beliefs, accepting any offer below [resp. above] $T(\underline{\gamma})$ would make the Home government strictly worsen off [resp. better off] than triggering a retaliatory trade war. Finally, as the Home government rejects any demand below $T(\underline{\gamma})$, the only possible deviation for the Foreign government would be to make a demand $\tau \geq T(\underline{\gamma})$. But, as $T(\underline{\gamma}) > T^*(\underline{\gamma}) \geq T^*(\gamma)$ for all types $\gamma \in [\underline{\gamma}, \bar{\gamma}]$, such a deviation would not be profitable. ■

B. Proof of Proposition 2

We prove Proposition 2 in five steps. Steps 1 and 2 show that, in any equilibrium, either all types of Foreign government successfully demand τ^{io} or they all make unsuccessful demands. Step 3 shows that all types successfully demand τ^{io} in equilibrium if and only if $\tau^{io} \geq T(F_0)$. Finally, Step 4 shows that a trade war never arises in equilibrium when $\tau^{io} \geq T(F_0)$, and that all types obtain concession τ^{io} when $\tau^{io} > T^*(\bar{\gamma})$. Finally, Step 5 proves existence of and characterizes equilibria when $\tau^{io} < T(F_0)$, showing that: all types of Foreign government fail to obtain a concession from the Home government, and the latter never complies with the IO ruling—thus completing the proof of the proposition.

Step 1: If the Foreign government makes a successful demand in equilibrium, then this demand must be τ^{io} .

Consider an equilibrium in which a demand τ is successfully made by a non-empty set of Foreign-government types Γ_τ . Let F be the Home government's updated beliefs after receiving this demand. Obviously, τ is the only successful proposal made in equilibrium—otherwise all types making the highest demands could profitably deviate by making the lowest demand. As it is optimal for the Home government to concede to τ , we must have $\tau \geq \tau^{io}$.

Now suppose by contradiction that $\tau > \tau^{io}$. As T^* is a decreasing function and indifferent players prefer agreements over disagreements, the set of types demanding τ must be of the form $[\underline{\gamma}, \hat{\gamma}]$, with $\hat{\gamma} \in [\underline{\gamma}, \bar{\gamma}]$. We distinguish between two different cases:

Case 1: $\hat{\gamma} > \underline{\gamma}$. In this case, $\tau \geq \max\{T(F), \tau^{io}\} \geq T(F) > T(\hat{\gamma})$. As any other equilibrium demand $\hat{\tau} \neq \tau$ is unsuccessful, we must have $\hat{\tau} < T(\hat{F})$, where \hat{F} represents the Home government's updated beliefs after observing $\hat{\tau}$. Moreover, as any such demand emanates from types $\gamma > \hat{\gamma}$ (and T is strictly decreasing), we must also have $T(\hat{F}) \leq T(\hat{\gamma})$. It follows that no demand in $(T(\hat{\gamma}), \tau)$ is made on the equilibrium path. Now consider a deviation from τ to $\tau' \in (\max\{T(\hat{\gamma}), \tau^{io}\}, \tau)$. By Lemma 3 in the supplementary online Appendix, reasonable beliefs F' must assign zero probability to the event $\{\gamma < \hat{\gamma}\}$ following demand τ' . This implies that $T(F') \leq T(\hat{\gamma}) < \tau'$, which in turn implies that demand $\tau' < \tau$ would be successful. By definition of a PBE, this is impossible: all types in $[\underline{\gamma}, \hat{\gamma}]$ can profitably deviate.

Case 2: $\hat{\gamma} = \underline{\gamma}$. In this case, demand τ reveals that the Foreign government's type is $\underline{\gamma}$. As it is optimal for the Home government to concede to τ , we must have $\tau \geq \max\{T(\underline{\gamma}), \tau^{io}\} \geq T(\underline{\gamma}) > T^*(\underline{\gamma})$. But this implies that the type- $\underline{\gamma}$ Foreign government could profitably deviate by making an unacceptable demand $\tau' < \tau^{io}$ (whether this leads to compliance with τ^{io} or with a trade war, it ends up strictly better off).

Step 2: In any equilibrium, either all types of Foreign government successfully demand τ^{io} or they all make unsuccessful demands.

From Step 1, to prove this statement, it suffices to show that, in any equilibrium, if some type successfully demands τ^{io} then all types do. We proceed by contradiction: suppose that a non-empty subset of types $\Gamma^{io} \neq [\underline{\gamma}, \bar{\gamma}]$ make the only successful demand τ^{io} in some equilibrium. As $T^*(\gamma)$ is a strictly decreasing function (and indifferent types are assumed to prefer a successful over an unsuccessful demand), Γ^{io} must be of the form $[\underline{\gamma}, \hat{\gamma}]$ with $\hat{\gamma} \geq \underline{\gamma}$.

Let F represent the Home government's beliefs when it receives demand τ^{io} . As it concedes to τ^{io} in equilibrium, $\tau^{io} \geq \max\{T(F), \tau^{io}\} \geq T(F)$. From our initial assumption, there must be a type γ' outside $[\underline{\gamma}, \hat{\gamma}]$ which makes an unsuccessful demand, say τ' , in equilibrium. Bayesian updating implies that the Home government's beliefs assign zero probability to the event $\{\gamma \leq \hat{\gamma}\}$ following demand τ' . As $T(\cdot)$ is strictly decreasing in γ , this in turn implies that $\tau^{io} \geq T(F) \geq T(\hat{\gamma}) \geq T(F')$ where F' represents the Home government's beliefs following demand τ' . Hence, the Home government complies with the IO ruling after rejecting demand τ' in this equilibrium, leaving the type- γ' Foreign government indifferent between its unsuccessful equilibrium demand τ' and the successful demand τ^{io} . According to our indifference condition, it should then demand τ^{io} instead of τ' .

Step 3: There is an equilibrium in which all types make a successful demand if and only if $T(F_0) \leq \tau^{io}$.

Necessity.—If all types of Foreign government demand τ^{io} in equilibrium, then the Home government's beliefs when receiving this demand are given by F_0 . As a consequence, we must have $\tau^{io} \geq \max\{T(F_0), \tau^{io}\} \geq T(F_0)$.

Sufficiency.—Suppose that $T(F_0) \leq \tau^{io}$, and consider the following strategy profile and beliefs: all types of Foreign government demand τ^{io} ; the Home government concedes to (multilateral) demand τ if and only if $\tau \geq \tau^{io}$ and always accepts the IO's ruling; it maintains its initial beliefs F_0 if it receives demand τ^{io} , and believes that the Foreign government is of type $\bar{\gamma}$ otherwise.

As $\tau^{io} \geq T(F_0) > T(\bar{\gamma})$, the Home government's beliefs ensure that it is always optimal for it to comply with the IO ruling and to concede to demand $\tau \geq \tau^{io}$ from the Foreign government. Anticipating that it will get payoff $W(\tau^{io}, \tau_0^*)$ if it does not concede to the Foreign government's demand, it is also optimal for the Home government not to concede to any $\tau < \tau^{io}$.

Given the Home government's strategy, the Foreign government has two options: (i) to make a successful demand $\tau \geq \tau^{io}$ and thus get a payoff of $W^*(\tau, \tau_0^*, \gamma)$; or (ii) to make an unsuccessful demand and thus get a payoff of $W^*(\tau^{io}, \tau_0^*, \gamma)$. As $W^*(\cdot, \tau_0^*, \gamma)$ is a strictly decreasing function for all $\gamma \in \Gamma$, demanding τ^{io} is the best strategy for any type of Foreign government.

Finally, it is readily checked that the Home government's beliefs satisfy Bayes' rule whenever possible. Moreover, by Lemma 4 in the supplementary online Appendix, they are reasonable.

Step 4: If $\tau^{io} \geq T(F_0)$, then a trade war never arises in equilibrium. In addition, if $\tau^{io} > T^*(\bar{\gamma})$ then all types of Foreign government make successful demands in any equilibrium.

Suppose that $\tau^{io} \geq T(F_0)$. To prove the statements above, we must first show that the Home government complies with the IO ruling whenever it rejects a demand from the Foreign government on the equilibrium path. To this end, consider an equilibrium—say σ —in which some type of Foreign government makes an unsuccessful demand. From Step 2, this implies that all types make unsuccessful demands. Let T^σ be the set of demands made by all types of Foreign government in σ , and let $\{\Gamma_\tau\}_{\tau \in T^\sigma}$ be a partition of $[\underline{\gamma}, \bar{\gamma}]$ such that all types in Γ_τ demand τ in equilibrium. Suppose first that the Home government rejects the IO ruling after rejecting any demand $\tau \in T^\sigma$. Letting F_τ denote the Home government's beliefs following demand τ , this would imply that $\tau^{io} < T(F_\tau)$ for all $\tau \in T^\sigma$; contradicting our assumption that $\tau^{io} \geq T(F_0)$.

Suppose now that T^σ can be partitioned into two non-empty, disjoint subsets T_1 and T_2 such that the Home government always concedes [resp. does not always concede] to the IO ruling after rejecting any $\tau \in T_1$ [resp. any $\tau \in T_2$]. In particular, observe that if a type γ prefers the IO ruling τ^{io} to a trade war, then so do all types $\gamma' < \gamma$ (recall that T^* is a strictly decreasing function). As σ is an equilibrium, no type that makes a demand in T_1 can profitably deviate by mimicking a type that makes a demand in T_2 , and vice versa. This implies that there exists a threshold type $\hat{\gamma}$ such that all types smaller [resp. larger] than $\hat{\gamma}$ belong to T_1 [resp. to T_2]. This in turn implies that the Home government learns that the Foreign government's type is lower [resp. greater] than $\hat{\gamma}$ when it receives a demand $\tau_1 \in T_1$ [resp. a demand $\tau_2 \in T_2$]. As $\tilde{W}(\gamma)$ is a strictly decreasing function, Bayesian updating then implies that $E_{F_{\tau_1}}[\tilde{W}(\gamma)] \geq \tilde{W}(\hat{\gamma}) \geq E_{F_{\tau_2}}[\tilde{W}(\gamma)]$ for all $\tau_1 \in T_1$ and all $\tau_2 \in T_2$. However, in equilibrium, the Home government prefers τ^{io} to a trade war after rejecting τ_1 and (strictly) prefers a trade war to τ^{io} after rejecting τ_2 ; that is

$$(A1) \quad W(\tau^{io}, \tau_0^*) \geq E_{F_{\tau_1}}[\tilde{W}(\gamma)] \geq E_{F_{\tau_2}}[\tilde{W}(\gamma)] > W(\tau^{io}, \tau_0^*),$$

which is of course impossible. This establishes that, in equilibrium σ , the Home government always complies with the IO ruling after rejecting any $\tau \in T^\sigma$.

Now suppose that $\tau^{io} \geq T(F_0)$ and $\tau^{io} > T^*(\bar{\gamma})$, and that there is an equilibrium in which some (and therefore all) types make unsuccessful demands. From Step 3 and the argument in the previous paragraph, we know that all equilibrium demands lead to the implementation of τ^{io} . This implies that demand τ^{io} must be unsuccessful in equilibrium; otherwise all Foreign government's types would be indifferent between their equilibrium unsuccessful demands and τ^{io} (and would therefore choose to demand τ^{io}). This in turn implies that demand τ^{io} is followed by a

trade war; otherwise the Home government would be indifferent between conceding and not conceding to τ^{io} and, therefore, would choose to concede. By definition of an equilibrium, no type of Foreign government can profitably deviate by demanding τ^{io} (thus triggering a trade war); that is: $W^*(\tau^{io}, \tau_0^*, \gamma) \geq \tilde{W}^*(\gamma)$ or, equivalently, $\tau^{io} \leq T^*(\gamma)$ for all $\gamma \in [\underline{\gamma}, \bar{\gamma}]$. As $T^*(\cdot)$ is a strictly decreasing function, this is equivalent to $\tau^{io} \leq T^*(\bar{\gamma})$, thus contradicting the assumption that $\tau^{io} > T^*(\bar{\gamma})$.

Step 5: If $\tau^{io} < T(F_0)$, then: (i) all types of Foreign government fail to obtain a concession from the Home government; and (ii) the latter never complies with the IO ruling. Such an equilibrium exists.

Suppose that $\tau^{io} < T(F_0)$. Part (i) is an immediate consequence of Steps 2 and 3. To prove part (ii), suppose toward a contradiction that there is an equilibrium in which a non-empty set of types of Foreign government, say Γ^{io} , make unsuccessful demands followed by compliance with τ^{io} . Observe that $\Gamma^{io} \neq [\underline{\gamma}, \bar{\gamma}]$, i.e., a non-empty subset of types must make unsuccessful demands followed by trade wars. To see this, suppose instead that all types' demands lead the Home government to comply with τ^{io} . Letting F_τ denote the Home government's beliefs following demand τ , this would imply that $\tau^{io} \geq T(F_\tau)$ for all on-the-equilibrium-path demands τ and, therefore, that $\tau^{io} \geq T(F_0)$; thus contradicting $\tau^{io} < T(F_0)$.

By definition of an equilibrium, $\gamma \in \Gamma^{io}$ if and only if $T^*(\gamma) \geq \tau^{io}$ (otherwise γ could profitably deviate by mimicking a type outside Γ^{io}). As T^* is a strictly decreasing function, there exists a threshold type $\hat{\gamma} < \bar{\gamma}$ such that $\Gamma^{io} = [\underline{\gamma}, \hat{\gamma}]$. This implies that, when the Home government receives a demand τ' from a type outside Γ^{io} , its updated beliefs F' assign a zero probability to the event $\{\gamma \leq \hat{\gamma}\}$. Hence, $T(F') \leq T(F_\tau) \leq \tau^{io}$ for any demand τ made by a type in Γ^{io} . But this implies that the Home government should comply with τ^{io} after rejecting demand τ' , yielding the desired contradiction.

We now have to prove that such an equilibrium exists. We argue that the following strategy profile and system of beliefs constitute an equilibrium: all types of foreign government demand τ^{io} ; the Home government concedes to demand τ if and only if $\tau \geq T(\underline{\gamma})$; it never complies with the IO ruling; and it believes that the Foreign government's type is $\underline{\gamma}$ if the latter demands $\tau \neq \tau^{io}$, and maintains its initial beliefs F_0 otherwise.

To see that the Foreign government does not have a profitable deviation, observe that it could only change the equilibrium outcome (i.e., a trade war) by making a demand $\tau \geq T(\underline{\gamma})$. As $T^*(\gamma) \leq T^*(\underline{\gamma}) < T(\underline{\gamma})$, this would be unprofitable to all Foreign government's types $\gamma \in [\underline{\gamma}, \bar{\gamma}]$.

As the Home government's beliefs are F_0 when it receives demand τ^{io} and $\tau^{io} < T(F_0)$, it is optimal for it not comply with ruling τ^{io} after rejecting demand τ^{io} . This in turn implies that it is also optimal to reject demand τ^{io} . When it receives a demand $\tau \neq \tau^{io}$, the Home government believes that the Foreign government is of type $\underline{\gamma}$. As $T(\underline{\gamma}) > T(F_0) > \tau^{io}$, it is optimal for the Home government to

trigger a trade war by rejecting the IO ruling. This in turn implies that it is a best response to concede to demand τ if and only if $\tau \geq T(\underline{\gamma})$.

Finally, it is readily checked that the Home government's beliefs satisfy Bayes' rule whenever possible. Moreover, Lemma 5 in the supplementary online Appendix shows that they also satisfy criterion D1. ■

C. Proof of Proposition 3

We prove Proposition 3 in two steps:

Step 1: There exists an equilibrium in which all types of Foreign government coerce unilaterally and fail to obtain a concession.

Let κ be a strictly positive number and consider the following strategy profile and beliefs: the type- γ Foreign government makes unilateral demand $\tau_\kappa(\gamma) \equiv T(\gamma) - \kappa$; the Home government concedes to a unilateral demand τ if and only if $\tau \geq T(\underline{\gamma})$; it concedes to a multilateral demand τ if and only if $\tau \geq \max\{T(\underline{\gamma}), \tau^{io}\}$; it complies with the IO ruling if and only if $\tau^{io} \geq T(\underline{\gamma})$; it believes that the Foreign government is of type γ when it is confronted with unilateral demand $\tau_\kappa(\gamma)$, for all $\gamma \in [\underline{\gamma}, \bar{\gamma}]$, and that it is of type $\underline{\gamma}$ when confronted with any other demand.

To see that these strategy profile and system of beliefs constitute an equilibrium, note first that the Home government's beliefs are consistent with Bayes' rule whenever possible. Moreover, Lemma 6 in the supplementary online Appendix shows that they are reasonable. The Foreign government can only change the outcome by making either a unilateral demand $\tau \geq T(\underline{\gamma})$ or a multilateral demand $\tau \geq \max\{T(\underline{\gamma}), \tau^{io}\}$. As $T^*(\gamma) \leq T^*(\underline{\gamma}) < T(\underline{\gamma}) \leq \max\{T(\underline{\gamma}), \tau^{io}\}$, however, such deviations can only make it worse off. Finally, it is readily checked that, given its beliefs, the Home government's strategy is a best response to the Foreign government's.

Step 2: In any equilibrium, a trade war arises with a probability of one.

To prove this statement, we will establish in turn that in equilibrium: (i) if all types of Foreign government make unilateral demands, then the Home government never concedes; (ii) if all types make multilateral demands, then the Home government never concedes to those demands and never complies with the IO ruling; and (iii) if some types coerce unilaterally and others multilaterally, then all their demands are unsuccessful and lead to a trade war:

- (i) If all types coerce unilaterally in equilibrium, then by the same argument as in Proposition 1 they all fail to obtain a concession (all deviations available in the game without IO are still available). Hence, a trade war ensues for all possible realizations of the Foreign government's type.
- (ii) Consider an equilibrium in which all types of Foreign government coerce multilaterally, and suppose (by contradiction) that some type's demand does

not lead to a trade war. By the same argument as in Proposition 2, this implies that $\tau^{io} \geq T(F_0)$ and that all types' demands lead to the implementation of τ^{io} . This in turn implies that $\tau^{io} \leq T^*(\bar{\gamma})$ —otherwise the type- $\bar{\gamma}$ Foreign government could profitably deviate by making an unacceptable unilateral demand $\tau' < T(\bar{\gamma})$. Lemma 7 in the supplementary online Appendix shows that, in such a case, reasonable beliefs must assign a probability of one to type $\bar{\gamma}$ following any (off-the-equilibrium-path) unilateral demand $\tau' < \tau^{io}$.

Now consider a deviation to unilateral demand $\tau' \in (T(\bar{\gamma}), \tau^{io})$ (observe that $T(\bar{\gamma}) < T(F_0) \leq \tau^{io}$). As the Home government believes that this demand emanates from the type- $\bar{\gamma}$ government, it should concede to it. This makes the deviation profitable for all types of Foreign government.

- (iii) Consider an equilibrium in which $[\underline{\gamma}, \bar{\gamma}]$ can be partitioned into two non-empty subsets Γ_1 and Γ_2 such that all types in Γ_1 [resp. Γ_2] coerce multilaterally [resp. unilaterally]. Proceeding by contradiction, assume that in this equilibrium, a trade war is avoided for some realization of the Foreign government's type. By the same argument as in Proposition 1, all types in Γ_2 fail to obtain a concession; so that a trade war occurs if $\gamma \in \Gamma_2$. Therefore, the types avoiding a trade war must be in Γ_1 . By the same argument as in Proposition 2, tariff τ^{io} must then be implemented whenever the Foreign government's type is in Γ_1 .

By definition of an equilibrium, types in Γ_1 cannot profitably deviate by mimicking types in Γ_2 , and vice versa. As $T^*(\gamma)$ is strictly decreasing in γ , this implies that there must be a type $\hat{\gamma} \in (\underline{\gamma}, \bar{\gamma})$ such that $\hat{\gamma} = (T^*)^{-1}(\tau^{io})$ and $\Gamma_1 = [\underline{\gamma}, \hat{\gamma}]$. (Note that $\hat{\gamma} > \underline{\gamma}$ because $T^*(\underline{\gamma}) < T(\underline{\gamma})$; and $\hat{\gamma} < \bar{\gamma}$ because by assumption $\Gamma_2 \neq \emptyset$.) We distinguish between two different cases:

- (a) If $T^*(\hat{\gamma}) \leq T(\hat{\gamma})$, then we have $\tau^{io} = T^*(\hat{\gamma}) \leq T(\hat{\gamma}) < T(\gamma)$ for all $\gamma < \hat{\gamma}$. This implies that there must be a demand τ emanating from some type, or some subset of types, in Γ_1 such that the Home government's updated beliefs F_τ satisfy $T(F_\tau) > \tau^{io}$. This in turn implies that it is optimal for the Home government to reject both demand τ and ruling τ^{io} in order to trigger a trade war; a contradiction.
- (b) If $T^*(\hat{\gamma}) > T(\hat{\gamma})$, then $\tau^{io} > T(\hat{\gamma})$. Consider a unilateral demand $\tau' \in (T(\hat{\gamma}), \tau^{io})$. Observe that this demand is only made off the equilibrium path: types $\gamma \leq \hat{\gamma}$ make multilateral demands, and types $\gamma > \hat{\gamma}$ make unsuccessful demands (as $T(\gamma) < T(\hat{\gamma}) < \tau'$ for all $\gamma > \hat{\gamma}$, the Home government would concede to τ' if it emanated from types $\gamma > \hat{\gamma}$ in equilibrium). Furthermore, Lemma 8 (in the supplementary online Appendix) shows that the Home government's beliefs F' when it receives unilateral demand τ' must assign zero probability to the event $\{\gamma < \hat{\gamma}\}$; so that $T(F') \leq T(\hat{\gamma}) < \tau'$ (recall that $T(\gamma)$ is strictly decreasing in γ). This implies that if some type of Foreign government deviated to unilateral demand τ' , then the Home government would

concede. As $\tau' < \tau^{io} = T^*(\hat{\gamma})$, this deviation is profitable to all types in Γ_1 —this is again a contradiction. ■

REFERENCES

- Bagwell, Kyle.** 2009. "Self-Enforcing Trade Agreements and Private Information." NBER Working Paper 14812.
- Bagwell, Kyle, Petros C. Mavroidis, and Robert W. Staiger.** 2007. "Auctioning Countermeasures in the WTO." *Journal of International Economics* 73 (2): 309–32.
- Bagwell, Kyle, and Robert W. Staiger.** 1999. "An Economic Theory of GATT." *American Economic Review* 89 (1): 215–48.
- Bagwell, Kyle, and Robert W. Staiger.** 2005. "Enforcement, Private Political Pressure, and the General Agreement on Tariffs and Trade/World Trade Organization Escape Clause." *Journal of Legal Studies* 34 (2): 471–525.
- Beshkar, Mostafa.** 2010. "Trade Skirmishes and Safeguards: A Theory of the WTO Dispute Settlement Process." *Journal of International Economics* 82 (1): 35–48.
- Beshkar, Mostafa.** 2016. "Arbitration and Renegotiation in Trade Agreements." *Journal of Law, Economics, and Organization* 32 (3): 586–619.
- Beshkar, Mostafa, and Eric W. Bond.** 2017. "Cap and Escape in Trade Agreements." *American Economic Journal: Microeconomics* 9 (4): 171–202.
- Bown, Chad P.** 2002. "The Economics of Trade Disputes, the GATT's Article XXIII, and the WTO's Dispute Settlement Understanding." *Economics and Politics* 14 (3): 283–323.
- Busch, Marc L.** 2000. "Accommodating Unilateralism? U.S. Section 301 and GATT/WTO Dispute Settlement." Typescript. Queen's School of Business, Kingston, Ontario.
- Busch, Marc L., and Eric Reinhardt.** 2000. "Bargaining in the Shadow of the Law: Early Settlement in GATT/WTO Disputes." *Fordham International Law Journal* 24 (1–2): 158–72.
- Cho, In-Koo, and David M. Kreps.** 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102 (2): 179–222.
- Conconi, Paola, Giovanni Facchini, and Maurizio Zanardi.** 2012. "Fast-Track Authority and International Trade Negotiations." *American Economic Journal: Economic Policy* 4 (3): 146–89.
- Eaton, Jonathan, and Maxim Engers.** 1992. "Sanctions." *Journal of Political Economy* 100 (5): 899–928.
- Eaton, Jonathan, and Maxim Engers.** 1999. "Sanctions: Some Simple Analytics." *American Economic Review* 89 (2): 409–14.
- Gilligan, Michael, Leslie Johns, and B. Peter Rosendorff.** 2010. "Strengthening International Courts and the Early Settlement of Disputes." *Journal of Conflict Resolution* 54 (1): 5–38.
- Grossman, Gene M., and Elhanan Helpman.** 1994. "Protection for Sale." *American Economic Review* 84 (4): 833–50.
- Grossman, Gene M., and Elhanan Helpman.** 2005. "A Protectionist Bias in Majoritarian Politics." *Quarterly Journal of Economics* 120 (4): 1239–82.
- Grossman, Sanford J.** 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics* 24 (3): 461–83.
- Horn, Henrik, Giovanni Maggi, and Robert W. Staiger.** 2010. "Trade Agreements as Endogenously Incomplete Contracts." *American Economic Review* 100 (1): 394–419.
- Hörner, Johannes, Massimo Morelli, and Francesco Squintani.** 2015. "Mediation and Peace." *Review of Economic Studies* 82 (4): 1483–1501.
- Hungerford, Thomas L.** 1991. "GATT: A Cooperative Equilibrium in a Noncooperative Trading Regime?" *Journal of International Economics* 31 (3–4): 357–69.
- Kerr, William A., and Jill E. Hobbs.** 2002. "The North American–European Union Dispute over Beef Produced Using Growth Hormones: A Major Test for the New International Trade Regime." *World Economy* 25 (2): 283–96.
- Klimenko, Mikhail, Garey Ramey, and Joel Watson.** 2008. "Recurrent Trade Agreements and the Value of External Enforcement." *Journal of International Economics* 74 (2): 475–99.
- Levy, Philip I.** 1999. "Lobbying and International Cooperation in Tariff Setting." *Journal of International Economics* 47 (2): 345–70.
- Limão, Nuno, and Kamal Saggi.** 2008. "Tariff Retaliation versus Financial Compensation in the Enforcement of International Trade Agreements." *Journal of International Economics* 76 (1): 48–60.
- Ludema, Rodney D.** 2001. "Optimal International Trade Agreements and Dispute Settlement Procedures." *European Journal of Political Economy* 17 (2): 355–76.

- Maggi, Giovanni.** 1999. "The Role of Multilateral Institutions in International Trade Cooperation." *American Economic Review* 89 (1): 190–214.
- Maggi, Giovanni, and Robert W. Staiger.** 2011. "The Role of Dispute Settlement Procedures in International Trade Agreements." *Quarterly Journal of Economics* 126 (1): 475–515.
- Martin, Alberto, and Wouter Vergote.** 2008. "On the Role of Retaliation in Trade Agreements." *Journal of International Economics* 76 (1): 61–77.
- Mayer, Wolfgang.** 1981. "Theoretical Considerations on Negotiated Tariff Adjustments." *Oxford Economic Papers* 33 (1): 135–53.
- Milgrom, Paul R.** 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics* 12 (2): 380–91.
- Milgrom, Paul, and Nancy Stokey.** 1982. "Information, Trade and Common Knowledge." *Journal of Economic Theory* 26 (1): 17–27.
- Park, Jee Hyeong.** 2011. "Enforcing International Trade Agreements with Imperfect Private Monitoring." *Review of Economic Studies* 78 (3): 1102–34.
- Pelc, Krzysztof J.** 2010. "Constraining Coercion? Legitimacy and Its Role in U.S. Trade Policy 1975–2000." *International Organization* 64 (1): 65–96.
- Puckett, A. Lynne, and William L. Reynolds.** 1996. "Rules, Sanctions and Enforcement under Section 301: At Odds with the WTO?" *American Journal of International Law* 90 (4): 675–89.
- Reinhardt, Eric.** 2000. "Aggressive Multilateralism: The Determinants of GATT/WTO Dispute Initiation, 1948–1998." https://www.iatp.org/sites/default/files/Aggressive_Multilateralism_The_Determinants_of.pdf.
- Riezman, Raymond.** 1991. "Dynamic Tariffs with Asymmetric Information." *Journal of International Economics* 30 (3–4): 267–83.
- Rosendorff, B. Peter.** 2005. "Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure." *American Political Science Review* 99 (3): 389–400.
- Rossmiller, George E.** 1994. "Discussion." In *Agricultural Trade Conflicts and GATT: New Dimensions in U.S.-European Agricultural Trade Relations*, edited by Giovanni Anania, Colin A. Carter, and Alex F. McCalla, 262–63. Boulder, CO: Westview Press.
- Schoppa, Leonard J.** 1999. "The Social Context in Coercive International Bargaining." *International Organization* 53 (2): 307–42.
- Vincent, Daniel R.** 1989. "Bargaining with Common Values." *Journal of Economic Theory* 48 (1): 47–62.
- World Trade Organization.** 2002. *The Legal Texts: The Results of the Uruguay Round of Multilateral Trade Negotiations*. Geneva: World Trade Organization.
- Zangl, Bernhard.** 2008. "Judicialization Matters! A Comparison of Dispute Settlement under GATT and the WTO." *International Studies Quarterly* 52 (4): 825–54.