Backtracking Under Scrutiny: Evidence and Regulation in the WTO*

Shing-hon Lam†

1 June 2025

Abstract

This paper investigates how the World Trade Organization shapes the design and enforcement of domestic regulations that affect trade. Under what conditions do governments modify or withdraw trade-related regulations in response to international scrutiny? Although the WTO agreements require that domestic regulations minimize trade restrictiveness, states often justify these measures using public interest objectives such as environmental protection or national security—justifications that can unintentionally or strategically favor domestic firms. Existing literature emphasizes formal dispute settlement and the threat of retaliation as the primary tools of enforcement within the WTO framework. However, newer work in political science and international law points to the importance of transparency, peer review, and evidentiary standards in constraining protectionist behaviour. Building on this insight, the paper theorizes the World Trade Organization as an epistemic regime in which member states assess regulatory legitimacy through the documentation submitted in the notification process. I argue that insufficient documentation may be perceived as weak justification, inviting challenges from other members, while excessive documentation may raise suspicions, signaling overcompensation and triggering further scrutiny. To test this theory, I draw on a dataset of 50,540 notifications of Technical Barriers to Trade and Sanitary and Phytosanitary measures. The main explanatory variable is the number of pages of supporting documentation submitted with each notification, interacted with a measure of protectionist pressure based on the regulated product's trade exposure. The outcome variable is whether the proposed regulation is modified or withdrawn. The results show that countries under protectionist pressure are most likely to alter their proposals when they submit either no supporting evidence or more than ten pages of documentation, suggesting that both under- and over-justification raise red flags among WTO members. These findings challenge dominant theories that prioritize legal retaliation and dispute settlement, highlighting instead the informal yet influential role of evidentiary review in global trade governance. The paper contributes to broader debates on how international institutions constrain protectionism while allowing states to pursue legitimate regulatory goals.

*Paper in preparation for Green Trade Lab Workshop 2025.

†PhD candidate at the Department of Political Science, UCLA. shinghon@ucla.edu.

1 Introduction

Regulations are prerequisites that companies must meet to sell their products within a country. To comply with World Trade Organization (WTO) agreements, governments must design these regulations to minimize trade restrictions while serving legitimate objectives (Barton et al., 2006, 138–9). However, ambitious public policy goals can drive regulations that significantly increase production costs for foreign firms, creating regulatory trade barriers (Sykes, 1999, 3). Under the European Union's proposed deforestation law, for example, exporting a 20-foot container of palm oil would require an estimated 1.2 million documents (Hancock and Ruehl, 2023). Major agricultural exporting countries have all criticised the law as protectionist (Ruehl, Hancock, and Terazono, 2023). Similarly, by 2027, household dishwashers sold in the U.S. must reduce energy use by 27 percent and water use by 34 percent in standard cycles—a shift expected to raise product costs significantly (Lange, 2023). China responded to the U.S. Technical Barriers to Trade (TBT) notification by urging the U.S. to retain existing standards, noting no major technological changes in dishwasher design (AQSIQ, 2021, 560).

These cases raise a broader theoretical question: *under what conditions do governments modify or withdraw regulatory trade barriers in response to international scrutiny?* Research shows the WTO's role in arbitrating a standard of legally permissible trade measures and allowing foreign governments unfairly harmed to seek retaliate (e.g. Rosendorff and Milner, 2001; Rosendorff, 2005). However, the WTO has increasingly recognised trade barriers as an exercise of legitimate regulatory autonomy.[1] States therefore often invoke general exceptions and security exceptions to shield their regulatory trade barriers from international scrutiny (Meyer, 2022, 1337). Other research shows softer tools like peer pressure, monitoring and recommendations from international organisations can influence state behaviour through reputational signalling (e.g. Pevehouse, 2005; Kelley, 2004, 2007; Simmons, 2000; Tomz, 2007). In these accounts, states care about how others perceive their credibility—and they engage in visible acts of compliance or transparency to signal their trustworthiness and alignment with international norms.

However, this literature largely treats disclosure as a binary variable—states are either transparent or opaque—and overlooks how the character and volume of justificatory information itself may function as a reputational signal. International law scholars argue that the scientific basis of a regulation significantly influences whether it is deemed discriminatory (e.g. Rigod, 2015; Lester, 2022), but little attention has been paid to how states present that basis within the notification process. This paper addresses that gap by theorising the WTO as an epistemic regime in

[1] Since the early 2010s, the WTO Appellate Body has justified discriminatory measures using the concept of "legitimate regulatory distinction." It has also limited states' obligations to adhere to international standards created through the participation of both developed and developing countries. Moreover, it has allowed "non-mainstream" opinions to justify Sanitary and Phytosanitary (SPS) measures (Howse and Langille, 2023, 33–34; Lester, 2014; Lester and Stemberg, 2014).

which states signal the legitimacy of their trade barriers through the documentation submitted in the notification process.[2] I hypothesise that insufficient documentation may be perceived as weak justification, inviting challenges from other members, while excessive documentation may raise suspicions, signalling overcompensation and triggering further scrutiny. This account contributes a non-linear model of reputational signalling, where both minimal and maximal disclosure can undermine credibility, and where states must navigate a reputational "Goldilocks zone" of justification.

To test this theory, I analyze a dataset of over 50,000 WTO notifications submitted under the Technical Barriers to Trade (TBT) and Sanitary and Phytosanitary (SPS) agreements. The outcome of interest is whether the proposing country modifies its initially proposed regulation. The main explanatory variable is the volume of supporting documentation submitted with the notification, interacted with the proposing country's protectionist motivation—measured by the import share of the regulated product relative to total trade. The analysis reveals a non-linear relationship: countries facing protectionist pressure are more likely to revise their regulations when they submit either too little or too much documentation, suggesting that both minimal and excessive disclosure can trigger international scrutiny.

These findings have important policy implications. Current approaches that rely on unilateral pressure to eliminate non-tariff barriers—such as those pursued by the Trump administration through retaliatory tariffs and aggressive demands for regulatory change—may overlook the institutional mechanisms through which regulatory revisions actually occur. My analysis suggests that states are more responsive to how their regulatory justifications are perceived within multilateral frameworks like the WTO, where documentation serves not just as procedural compliance but as a reputational signal. As former U.S. ambassador to the EU Anthony Gardner warned, "these things are highly complex and require deep analysis, so if someone like Trump says let's just get rid of them, it's not going to work" (Mackrael, Bade, and Luhnow, 2025). Rather than forcing regulatory change through coercive tactics, more effective strategies would engage governments through established WTO procedures that emphasize transparency, deliberation, and peer review. This underscores the value of multilateral oversight and reputational accountability in shaping the global trade regime, especially when regulatory barriers arise from complex political, cultural, or scientific considerations rather than overt protectionism.

The remainder of the paper proceeds as follows. Section 2 outlines the WTO's notification process for regulatory measures and the growing concern that such measures may conceal protectionist intent. Section 3 reviews existing literature on non-tariff barriers, focusing on how legal scholars interpret documentation as a signal of regulatory legitimacy. Section 4 develops a theory of strategic disclosure, arguing that—contrary to legal expectations—governments may submit

[2] See Meyer (2021, 55–56) for characterising international organisations as an epistemic regime.

more documentation not to demonstrate justification, but to obscure controversial intent or pre-empt scrutiny. Section 5 describes the original dataset of over 50,000 TBT and SPS notifications. Section 6 presents the regression results.

2 WTO's notification and commenting process for TBT/SPS

The WTO dispute settlement body represents the small tip of a pyramid in which member states scrutinise regulatory trade barriers. Before reaching formal consultations, there are at least three prior stages: *notification*, *commenting*, and *committee-level discussion*. Each stage offers an opportunity for members to examine the protectionist intent behind proposed measures. Through questions, objections, or reputational pressure, these stages can influence the regulating country to withdraw or modify the trade barrier.

At the bottom of the pyramid is the obligation to notify. After drafting and before implementing a measure that may have a "significant effect on trade," countries are obliged to notify other members through an online platform. Notifications disclose the regulation's objectives, its expected trade impact, and the scope of affected products and partners (WTO, 2018). As of June 2025, over 96,000 TBT and SPS notifications had been submitted.[3] The centralisation of this large volume of notifications by the WTO Secretariat—made publicly accessible through its online platform—has significantly enhanced the systematic dissemination and accessibility of information on members' regulatory frameworks. This transparency mechanism gives security and predictability through allowing members and their industries advance insights into the proposed regulatory trade barriers and time to adapt (Karttunen, 2020, 294). Empirically, the strong correlation between the number of TBT notifications submitted and the number of specific trade concerns (STCs) raised in the Committee suggests that members routinely identify issues for scrutiny from the notification process itself (Horn, Mavroidis, and Wijkström, 2013). At the same time, strategic incentives can shape whether and how countries choose to notify. Governments may withhold notifications to avoid exposing potentially non-compliant measures to criticism or dispute. Others may refrain from notifying controversial measures if a more powerful country has adopted similar policies without disclosure, or out of concern that doing so would reveal bargaining positions in ongoing negotiations (Borlini, 2019; Wolfe, 2013, 18).

The second stage is commenting. Member states are obliged to establish national enquiry points to receive and respond to comments from other members and are encouraged to publish their responses. The commenting period will last at least sixty days before a regulation can be finalised (WTO, 2018). Comments may request clarification, raise concerns, or propose amendments. Responses can include the full legislative text, relevant scientific opinions, or measures to maximise flexibility. For example, when Ecuador objected to proposed pesticide residue limits on

[3] Data from https://eping.wto.org/

pineapples, the EU introduced a temporary tolerance level while requesting additional data. In contrast, when Japan raised concerns about the removal of substances from a flavourings register, the EU requested new safety evidence before reconsidering (Downes, 2012). Even when the executive is open to compromise—as in efforts to revise poultry certification language to accommodate U.S. exporters—domestic opposition within the EU can block amendments. Although not all comments result in policy change, the process fosters transparency and dialogue aimed at resolving disputes. It also promotes a culture of critical self-reflection, where regulators assess trade implications to pre-empt scrutiny. Over time, this can institutionalise self-assessment and embed new norms of responsiveness to international concerns into domestic regulatory practice (Downes, 2012).

Finally, in addition to commenting and formal requests for consultation, member states may raise Specific Trade Concerns (STCs) during meetings of WTO technical committees on Sanitary and Phytosanitary (SPS) or Technical Barriers to Trade (TBT) measures. These STCs offer a low-cost, low-friction channel for expressing concern over another member's trade-related regulation—an avenue that faces few procedural barriers and thus tends to be widely utilized by governments seeking dialogue without triggering legal escalation (Busch and Pelc, 2015). Resolutions are often achieved not through adjudication but when the concerned member receives sufficient clarification or when the regulating government amends the measure, either in acknowledgment of the concern's validity or in response to new technical insights shared during the discussion (Lang and Scott, 2009). Even absent legal pressure, the reputational risks of facing difficult or politically charged questions in front of other members can incentivize governments to align domestic policies more closely with WTO norms.[4] Moreover, the peer exchange within these committees promotes policy learning and norm diffusion among national regulators. The information-sharing mandates foster a collaborative atmosphere: semi-formal committee procedures draw officials into deliberative settings where professional norms—particularly those tied to scientific risk assessment—often take precedence over legalistic bargaining. These peer-review dynamics help cultivate cross-national networks that legitimize third-country voices and open space for mutually acceptable regulatory solutions (Downes, 2012; Cho, 2011; Wolfe, 2020).

While the literature acknowledges that pre-litigation mechanisms at the WTO—such as notifications, commenting, and committee discussions—can encourage the modification or withdrawal of regulatory trade barriers, concerns remain about their effectiveness. Members have noted that their comments are often ignored, with limited responses from notifying countries; this appears to be a systemic issue that affects both developed and developing members alike (Karttunen, 2020, 87). In some cases, governments strategically withhold information—such as when national se-

[4] For example, China used an STC to publicly criticize the United States' sweeping invocation of GATT Article XXI security exceptions, labeling it as an "abuse" of the system. Though the U.S. declined to respond, the entire WTO membership was made aware of China's concern through the committee process (Pinchis-Paulsen, 2022, 539).

curity exceptions are invoked without explanation—effectively shutting down legal dialogue and leaving adjudicators to speculate on states' intent (Pinchis-Paulsen, 2022). Additionally, STCs frequently fall short of resolving disputes that touch on sensitive political issues, whether rooted in domestic affairs or complex bilateral relationships. In such cases, disagreement may be too fundamental, or span too many legal instruments, to be addressed meaningfully through committee-level scrutiny alone (Holzer, 2019).

To sum up, the existing literature has not fully explained when and how the WTO's pre-litigation mechanisms facilitate changes in members' trade policies. It has largely overlooked variation in how members behave within these processes, and in particular, how the form and content of notifications—especially the use of scientific evidence—can shape perceptions of a measure's legitimacy or protectionist intent. The next section addresses this gap by theorising the WTO as an epistemic regime, in which documentation operates as a signal of regulatory intent. I argue that the effectiveness of the WTO's pre-litigation tools depends not only on institutional design but also on how member states interpret and respond to these signals during the early stages of scrutiny.

3 The role of scientific evidence in regulatory trade barriers

The requirement that WTO members justify their regulatory trade barriers with scientific evidence is widely viewed as a safeguard against both overt and covert protectionism (Herwig, 2008). Scholars argue that this evidentiary threshold promotes transparency, constrains arbitrary decision-making, and enhances the credibility of national regulations. Scientific reasoning acts as a legitimating device: it disciplines risk regulation by exposing false claims or prejudice, while still permitting governments to adopt stricter standards for political or societal reasons if justified (Howse, 2000). Because science does not dictate a unique regulatory outcome and remains open to interpretation, it functions as a shared meta-norm for resolving jurisdictional conflicts between exporting and importing countries. By rendering decisions intelligible across borders, scientific evidence facilitates mutual understanding even where members disagree on acceptable levels of risk (Joerges, 2005). In this way, scientifically grounded measures face less scrutiny from trading partners and are less likely to be challenged or modified. The more robust and transparent the scientific justification, the more legitimate the measure appears—and the more likely it is to withstand external pressure.

Yet while scientific evidence may bolster the legitimacy of regulatory trade measures, it cannot alone determine whether a measure is protectionist. As Peel observes after surveying fifteen years of WTO case law, scientific reasoning under the SPS Agreement has not delivered clear or consistent standards for distinguishing between legitimate regulation and disguised protectionism (Bonneuil and Levidow, 2012). Legal scholars and science studies researchers emphasize that scientific assessments are themselves shaped by normative assumptions and procedural context. Walker

(2003, pp. 166–197), for instance, argues that scientific input into legal decision-making necessarily involves judgments about uncertainty and acceptable risk—decisions that cannot be made by science alone. Similarly, Busch et al. (2004) argue that risk assessment is not a value-neutral science but a heterogeneous set of methods conditioned by institutional and political factors. On this view, what appears as 'scientific evidence' in WTO dispute settlement is not a pure reflection of objective knowledge but the product of institutional co-production between legal and scientific domains (Jasanoff, 2004). Panels do not merely receive expert advice; they actively shape what counts as relevant expertise, select which experts to consult, and determine how that knowledge is used (Jasanoff, 2008). In this sense, science in WTO adjudication is not a neutral arbiter of truth, but a reconstructed image crafted within legal procedures.

As a result, while greater scientific justification may reduce the likelihood of challenge, it does not eliminate it entirely. This nuance informs the observable implications of my signaling model: although regulatory measures accompanied by more extensive evidence are less frequently scrutinized, challenge rates remain strictly positive even at high levels of evidentiary disclosure. This reflects the legal and epistemic ambiguity surrounding what counts as 'legitimate' risk regulation, and aligns with the idea that suspicion never fully vanishes—particularly when over-signaling may itself be interpreted as strategic behaviour.

4 A continuous-evidence signaling game with endogenous modification

4.1 Setup

We consider a signaling game between a Proposing Country (Player 1) and an Observer Country (Player 2). Player 1 can be one of two types: protectionist (P) or legitimate regulator (L). The type $\theta \in \{P, L\}$ is private information to Player 1. Player 1 first chooses whether to enact a measure for legitimate or protectionist purposes. If the L-type measure is enacted, Player 1 receives payoff l, and Player 2 suffers a loss of l. If the P-type measure is enacted, Player 1 receives payoff $p > l$, and Player 2 suffers a loss of p. After choosing a type, Player 1 selects an evidence level $E \in [0, +\infty)$. This evidence is observable by Player 2.

Player 2 observes E and decides whether to challenge the measure. If Player 2 does not challenge, the game ends. The resulting payoffs are $(l-E, -l)$ if the measure is L-type, and $(p-E, -p)$ if P-type. If Player 2 chooses to challenge, she incurs a fixed cost a, and Player 1 must then decide whether to modify the regulation. If Player 1 modifies, the game ends with payoffs $(-E, -a)$ for L-type and $(-E, -a)$ for P-type. If Player 1 refuses to modify, Player 2 may choose to escalate the matter to court. If the case proceeds to court, both players incur cost c, and the court ensures the regulation is struck down with no benefit to Player 1 or harm to Player 2 beyond sunk costs. In this case, the payoffs are $(-E-c, -a-c)$for L-type and $(-E-c, -a-c)$ for P-type. If Player 2 declines to go to court, the measure remains in place. The payoffs then depend on the type: if

the regulation is L-type, payoffs are $(l - E, -a - l)$; if P-type, $(p - E, -a - p)$.

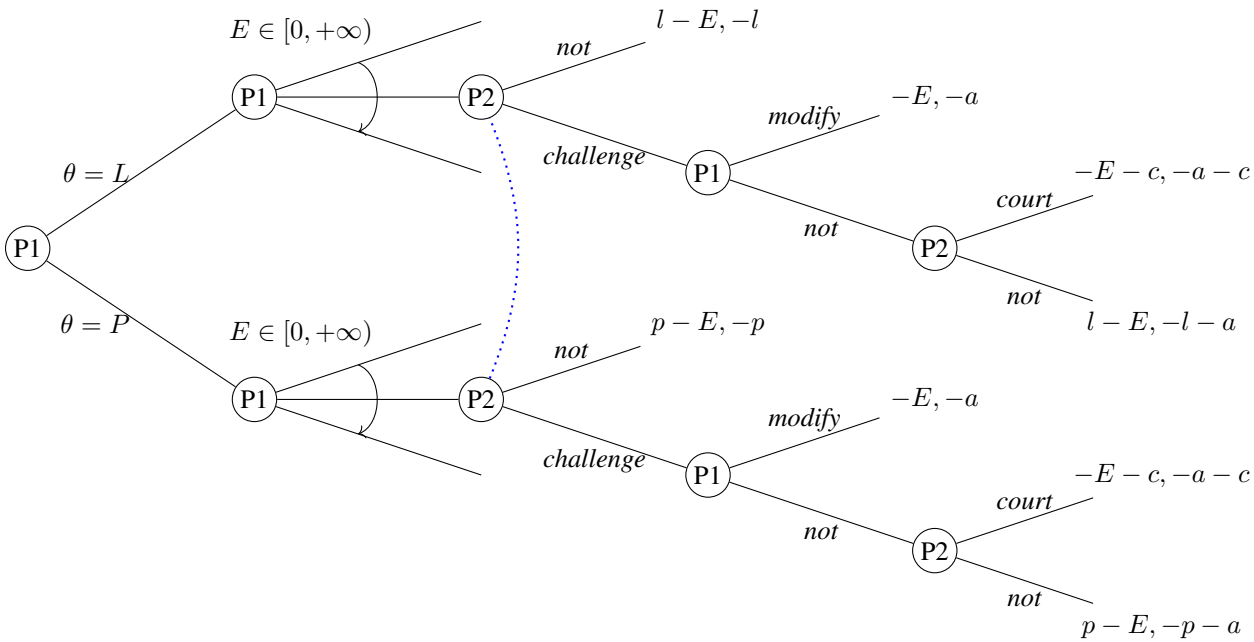

Figure 1: Two-player signalling game between a Proposing Country (Sender) **P1** and a WTO Observer (Receiver) **P2**. Dashed arc illustrates continuous evidence choice.

We analyze a signaling game in which a sender, Player 1 (P1), has a private type $\theta \in \{L, P\}$, where L denotes a legitimate regulatory motivation and P a protectionist one. The prior probability that $\theta = P$ is π, and $1 - \pi$ for type L. P1 selects a level of evidence $E \in \mathbb{R}_{\geq 0}$, which is observed by a receiver, Player 2 (P2). P2 then updates her belief about P1's type and chooses whether to challenge the regulatory measure.

The structure of payoffs is as follows. If no challenge occurs, P1 obtains a payoff equal to the regulatory value minus the cost of producing evidence, $u_1 = v_\theta - E$, with $v_P = p$ and $v_L = l$, while P2 bears the corresponding loss $u_2 = -v_\theta$. If P2 challenges, she pays a litigation cost $a > 0$, and P1 is assumed—at this first stage of the model—to modify the measure, resulting in utility $u_1 = -E$, $u_2 = -a$. The signal cost E is always sunk and unavoidable.

4.2 Equilibrium analysis

We introduce an endogenous challenge strategy $c(E) \in [0, 1]$, representing the probability that P2 challenges after observing evidence level E. Her goal is to minimize her expected loss, which arises from either incorrectly allowing harmful regulations to stand or from incurring unnecessary challenge costs.

Receiver's problem. Given the probability density functions over evidence chosen by types P and L, $r_P(E)$ and $r_L(E)$, P2 chooses a challenge strategy $c(E)$ to minimize her expected loss:

$$\min_{c(\cdot)} \mathbb{E}[L_2] = \int r_P(E) \left[\underbrace{c(E)a}_{\text{loss if challenge}} + \underbrace{(1-c(E))p}_{\text{loss if no challenge}} \right] dE + \int r_L(E) \left[\underbrace{c(E)a}_{\text{same challenge cost}} + \underbrace{(1-c(E))l}_{\text{smaller loss if no challenge}} \right] dE,$$

where a, p, and l represent the respective losses from challenging or not challenging. This expression integrates over all evidence levels E, weighting the loss from each action by the probability that a P-type or L-type chooses that signal. P2's problem is to determine, for each E, the challenge probability $c(E)$ that minimizes this overall expected loss. Later, we show that this optimization leads to a threshold rule where P2 challenges if and only if her posterior belief that P1 is protectionist exceeds a cutoff.

Sender's Problem. Each type-θ sender chooses a probability density function $r_\theta(E)$ over evidence levels $E \geq 0$ to maximize their expected payoff:

$$\max_{r_\theta(E)} \int r_\theta(E) \left[(1-c(E))v_\theta - E \right] dE,$$

subject to:

- $r_\theta(E) \geq 0$ for all $E \geq 0$ (non-negativity),
- $\int r_\theta(E)\, dE = 1$ (total probability integrates to one).

This is an infinite-dimensional optimization problem where the sender chooses a distribution over signals. Let us interpret the components of the objective:

- $r_\theta(E)$ is the density over signal levels. That is, $r_\theta(E)\, dE$ represents the probability that a sender of type $\theta \in \{P, L\}$ chooses a signal in the infinitesimal range $[E, E + dE]$.
- $(1 - c(E))v_\theta$ is the expected benefit from the regulation surviving scrutiny:
  - v_θ is the value the sender receives if the regulation is not challenged (equal to p for type P and l for type L),
  - $1 - c(E)$ is the probability that the receiver does not challenge at signal E.
- E is the cost of producing signal E.

The sender integrates this net payoff across all signal levels, weighted by how often each E is chosen. Thus, the optimization problem seeks the best density $r_\theta(E)$ that maximizes average expected benefit minus signaling cost, given the receiver's challenge strategy $c(E)$.

Incentive Compatibility. In a separating equilibrium, each sender type $\theta \in \{L, P\}$ selects a probability density function $r_\theta(E)$ to maximize expected utility, subject to the constraint that the

density integrates to one. Formally, each type solves:

$$\max_{r_\theta(\cdot)} \int r_\theta(E)\left[(1-c(E))v_\theta - E\right] dE \quad \text{subject to} \quad \int r_\theta(E)\, dE = 1,$$

where v_θ denotes the benefit of regulation to type θ, E is the cost of signaling, and $c(E)$ is the probability of being challenged after sending signal E.

The associated Lagrangian is:

$$\mathcal{L}_\theta = \int r_\theta(E)\left[(1-c(E))v_\theta - E\right] dE - \lambda_\theta \left(\int r_\theta(E)\, dE - 1 \right),$$

with λ_θ the Lagrange multiplier on the density constraint. Taking the first-order condition with respect to $r_\theta(E)$ yields:

$$(1-c(E))v_\theta - E = \lambda_\theta \quad \text{for all } E \text{ with } r_\theta(E) > 0.$$

Now consider both types at the same signal level E. Each type's marginal utility from signal E is given by:

$$\begin{aligned} \text{P-type:} \quad & (1-c(E))v_P - E, \\ \text{L-type:} \quad & (1-c(E))v_L - E. \end{aligned}$$

So if a signal is, for example, twice as valuable to a P-type as to an L-type, we expect P-types to use that signal more frequently. This leads to the proportionality rule:

$$\frac{r_L(E)}{r_P(E)} \propto \frac{\text{marginal value for L}}{\text{marginal value for P}}.$$

While it may be intuitive to think that a type with higher marginal value would use a signal more often, the logic in equilibrium is the opposite. The key idea is that a type who benefits more from avoiding challenge is willing to pay more to do so. As a result, that type (e.g., a protectionist) is willing to bear the cost of sending higher evidence levels that would be too expensive for the other type (e.g., a legitimate sender). This implies that the high-value type's density declines more slowly in E: they continue signaling at higher levels, while the low-value type drops out sooner.

Thus, even though the protectionist type values the regulation more, they are not more likely to choose any given signal E; instead, their signal distribution has thicker tails. That is why the

correct expression places the higher-value type in the numerator of the density ratio:

$$\frac{r_L(E)}{r_P(E)} = \frac{(1 - c(E))v_P - E}{(1 - c(E))v_L - E}.$$

Here:

- The left-hand side represents the relative likelihood that evidence level E is selected by a legitimate rather than protectionist sender.
- The right-hand side reflects the ratio of marginal incentives to use signal E across types.

Since protectionist types derive greater value from maintaining the regulation ($p > l$), they are more willing to bear high signaling costs. As a result, $r_P(E)$ decays more slowly than $r_L(E)$ in E. In the special case where challenges never occur ($c(E) = 0$), the expression simplifies to:

$$\frac{r_L(E)}{r_P(E)} = \frac{p - E}{l - E}.$$

At high values of E, the cost of signaling approaches or exceeds l, causing legitimate types to exit. P-types, who still benefit from high p, continue to produce strong evidence. As a result, very high levels of evidence become increasingly associated with protectionist intent—a key feature of belief dynamics in the model.

From here, P2 updates her belief using Bayes' Rule:

$$\mu(E) = \frac{\pi r_P(E)}{\pi r_P(E) + (1 - \pi) r_L(E)} = \left[1 + \frac{1 - \pi}{\pi} \cdot \frac{r_L(E)}{r_P(E)}\right]^{-1}.$$

A challenge is optimal for P2 if the expected regulatory harm exceeds the cost of litigating:

$$\mu(E)p + (1 - \mu(E))l > a \quad \Rightarrow \quad \mu(E) > \bar{\mu} := \frac{a - l}{p - l}.$$

This gives the sharp cutoff rule:

$$c(E) = \begin{cases} 1 & \text{if } \mu(E) > \bar{\mu}, \\ 0 & \text{otherwise.} \end{cases}$$

Post-Challenge Decision: Modify or Resist. To model the full dynamics of retreat, we introduce an additional endogenous decision: after being challenged, the sender (P1) may choose whether to *modify* the regulation or *resist* and risk further escalation. If P1 refuses to modify, the receiver (P2) may escalate the matter to court at a cost $c > 0$. This introduces a richer enforcement structure that reflects institutional path-dependence in regulatory disputes.

After being challenged, P1 compares two options:

- **Modify:** The sender concedes and removes the regulation. She loses the signaling cost E and receives no regulatory benefit:

$$u_1^{\text{modify}} = -E$$

- **Do Not Modify:** The sender resists. Let $\delta(E)$ denote the probability that P2 escalates to court. P1's expected payoff becomes:

$$u_1^{\text{not modify}} = (1 - \delta(E))(v_\theta - E) + \delta(E)(-E - c)$$

  where v_θ is the value of the regulation to type $\theta \in \{L, P\}$.

P1 will choose to modify when doing so yields a higher payoff than resisting. The indifference condition is:

$$-E = (1 - \delta(E))(v_\theta - E) + \delta(E)(-E - c)$$

Solving this equation yields the threshold at which P1 is just willing to concede:

$$\begin{aligned} -E &= (1 - \delta)(v_\theta - E) + \delta(-E - c) \\ &= v_\theta - E - \delta v_\theta - \delta c \\ 0 &= v_\theta - \delta v_\theta - \delta c \\ \delta(v_\theta + c) &= v_\theta \\ \delta(E) &= \frac{v_\theta}{v_\theta + c} \end{aligned}$$

Therefore, P1 modifies the regulation if and only if:

$$\delta(E) \geq \frac{v_\theta}{v_\theta + c}$$

This threshold is increasing in v_θ and decreasing in c, meaning that types who place higher value on the regulation (e.g., protectionists) are more willing to resist challenge, while larger litigation penalties make all types more likely to retreat early.

Escalation and Modification Thresholds. On the receiver's side, P2 considers whether to escalate the dispute to court after challenging P1. Escalation incurs a fixed cost $c > 0$, so she will proceed only if the expected regulatory harm exceeds this cost.

P2 updates her belief after observing evidence E:

$$\mu(E) := \Pr(\theta = P \mid E)$$

This is the posterior probability that the sender is a protectionist type. The expected harm from allowing the regulation to remain is:

$$\mu(E) \cdot p + (1 - \mu(E)) \cdot l$$

where p and l represent the loss from harmful and legitimate regulations, respectively. Escalation is optimal if this expected harm exceeds the litigation cost c:

$$\begin{aligned} \mu(E)p + (1 - \mu(E))l &> c \\ \mu(E)(p - l) + l &> c \\ \mu(E)(p - l) &> c - l \\ \mu(E) &> \bar{\mu}_{\text{court}} := \frac{c - l}{p - l} \end{aligned}$$

Hence, P2 escalates only when her belief $\mu(E)$ exceeds this threshold. As the cost of court proceedings c increases, P2 becomes more selective about which cases to pursue.

From the sender's perspective, P1 anticipates this decision and compares whether to modify the regulation preemptively. As derived earlier, P1 is indifferent between modifying and resisting when the escalation probability $\delta(E)$ crosses the threshold:

$$\delta(E) \geq \frac{v_\theta}{v_\theta + c}$$

where $v_\theta \in \{l, p\}$ is the value of the regulation to type-θ. If escalation becomes likely, modifying becomes the less costly option.

We now ask: what value of the threshold μ^* would generate the desired challenge probability $\delta^* \in (0, 1)$? To answer this, observe that since $\delta(E) = \mathbf{1}_{\mu(E) > \mu^*}$, it follows that

$$\Pr(\mu(E) > \mu^*) = \delta^* \quad \Rightarrow \quad \Pr(\mu(E) \leq \mu^*) = 1 - \delta^* \quad \Rightarrow \quad \mu^* = \text{the } (1-\delta^*)\text{-quantile of } \mu(E).$$

Accordingly, if the distribution of $\mu(E)$ is uniform on the unit interval, we obtain:

$$\mu^* = 1 - \delta^*.$$

Hence, the following relationship holds:

$$\delta(E) = 1 \iff \mu(E) > \mu^* \quad \Rightarrow \quad \mu^* = 1 - \delta^*$$

Given the observer's strategy is to challenge if and only if the posterior belief exceeds a threshold μ^*, we previously derived:

$$\mu^* = 1 - \delta^* = 1 - \frac{v_\theta}{v_\theta + c} = \frac{c}{v_\theta + c}.$$

In addition, suppose the observer only escalates (e.g., proceeds to court) if their posterior belief also exceeds an institutional threshold $\bar{\mu}_{\text{court}}$. Then escalation occurs only if both conditions are satisfied. That is, Player 2 challenges if:

$$\mu(E) > \max\left(\bar{\mu}_{\text{court}}, \frac{c}{v_\theta + c}\right).$$

This expression represents the most stringent of the two thresholds—either the observer's internal cost-based cutoff or an external limit imposed by institutional credibility constraints.

It follows that Player 1 will choose to modify the measure only when they expect to be challenged and the challenge would likely succeed in court. Hence, the sender modifies if and only if:

$$\mu(E) > \max\left(\bar{\mu}_{\text{court}}, \frac{c}{v_\theta + c}\right).$$

We now formalize how the posterior belief $\mu(E)$—the probability that Player 1 is a protectionist type—can increase again at high evidence levels, even though it typically declines with moderate increases in E. This non-monotonicity is driven by the endogenous response of type-specific densities to the rising cost of signaling.

Let $R(E) := \frac{r_L(E)}{r_P(E)}$ denote the likelihood ratio of legitimate to protectionist densities. Then, by Bayes' rule, Player 2's posterior belief is:

$$\mu(E) = \left[1 + \frac{1-\pi}{\pi} R(E)\right]^{-1}, \quad \text{so} \quad \mu'(E) > 0 \iff R'(E) < 0.$$

In a separating region—where types choose different evidence levels and incentive compatibility binds—the relative densities are determined by marginal indifference conditions. Suppose legitimate and protectionist types have valuations $l < p$, and face linear signaling costs E. Then,

$$\frac{r_L(E)}{r_P(E)} = \frac{p - E}{l - E} \quad \Rightarrow \quad \mu(E) = \left[1 + \frac{1-\pi}{\pi} \cdot \frac{p-E}{l-E}\right]^{-1}.$$

This expression is strictly increasing in E over the interior of the separating region, which implies that $\mu(E)$ declines monotonically with increasing evidence. Intuitively, higher E is more likely to come from the higher-value (legitimate) type, leading to lower posterior suspicion.

However, this pattern can reverse when signaling becomes prohibitively costly for legitimate types. Suppose there exists a threshold $E = \bar{E}$ such that:

$$\lim_{E \to \bar{E}^+} r_L(E) = 0,$$

but for some $\delta > 0$, we have:

$$r_P(E) > \epsilon > 0 \quad \text{for all } E \in [\bar{E}, \bar{E} + \delta].$$

Then,

$$\lim_{E \to \bar{E}^+} \frac{r_L(E)}{r_P(E)} = 0 \quad \Rightarrow \quad \mu(E) \to 1.$$

That is, at very high evidence levels, legitimate types drop out of the distribution entirely because the signaling cost exceeds their valuation l. In contrast, some protectionist types—who derive greater benefit p—continue to send high signals in order to feign legitimacy and avoid being challenged. As a result, observing a very high evidence level becomes suspicious rather than reassuring: $\mu(E)$ rises again as $E \to \bar{E}^+$.

This reversal marks the breakdown of the separating region and introduces a pooling behavior where only protectionist types persist, creating a "return of suspicion at high evidence."

4.3 Return of suspicion at high evidence

We now formalize how the posterior belief $\mu(E)$—the probability that Player 1 is a protectionist type—can increase again at high evidence levels, even though it typically declines with moderate increases in E. This reversal is driven by the endogenous response of type-specific densities to the cost of signaling.

Let $R(E) := \frac{r_L(E)}{r_P(E)}$ denote the likelihood ratio of legitimate to protectionist density. Then Player 2's posterior belief is given by:

$$\mu(E) = \left[1 + \frac{1 - \pi}{\pi} R(E)\right]^{-1}, \quad \text{so} \quad \mu'(E) > 0 \iff R'(E) < 0.$$

In the separating region—where types choose different evidence levels and incentive compati-

bility binds—we have:

$$\frac{r_L(E)}{r_P(E)} = \frac{p-E}{l-E} \quad \Rightarrow \quad \mu(E) = \left[1 + \frac{1-\pi}{\pi} \cdot \frac{p-E}{l-E}\right]^{-1}.$$

This expression is strictly increasing in E, meaning that $\mu(E)$ decreases monotonically as evidence rises, so long as both densities behave according to the separating equilibrium.

However, this pattern can reverse when signaling costs become prohibitively high for legitimate types. Suppose there exists a threshold $E = \bar{E}$ such that:

$$\lim_{E \to \bar{E}^+} r_L(E) = 0,$$

but for some $\delta > 0$, we have:

$$r_P(E) > \epsilon > 0 \quad \text{for all } E \in [\bar{E}, \bar{E} + \delta].$$

Then:

$$\lim_{E \to \bar{E}^+} \frac{r_L(E)}{r_P(E)} = 0 \quad \Rightarrow \quad \mu(E) \to 1.$$

At very high evidence levels, L-types stop signaling entirely because the cost E exceeds their benefit l, while some P-types with higher payoff p continue choosing high E to feign legitimacy and avoid being challenged. As a result, observing such high E becomes suspicious rather than reassuring.

Lemma (Return of Suspicion). If $r_L(E), r_P(E)$ are continuously differentiable at some E_0, and

$$\frac{d}{dE}\left(\frac{r_L(E)}{r_P(E)}\right)\Big|_{E=E_0} < 0,$$

then

$$\mu'(E_0) > 0.$$

Proof. Differentiating the belief expression:

$$\mu(E) = \left[1 + \frac{1-\pi}{\pi} R(E)\right]^{-1} \quad \Rightarrow \quad \mu'(E) = -\left[1 + \frac{1-\pi}{\pi} R(E)\right]^{-2} \cdot \frac{1-\pi}{\pi} \cdot R'(E).$$

Therefore, $\mu'(E) > 0$ if and only if $R'(E) < 0$. ■

4.4 Observable implications

This mechanism yields several predictions that can be tested in observable data:

1. **Non-monotonic modification behaviour for protectionist types.** Among actors with strong regulatory interests (high v_θ), the probability of backing down after challenge will be highest at very low and very high evidence levels, and lower at intermediate evidence. This is because:

$$\begin{aligned} \mu(E) \text{ high at low } E &\Rightarrow \text{credible court threat} \Rightarrow \text{modification}, \\ \mu(E) \text{ low at moderate } E &\Rightarrow \text{no court threat} \Rightarrow \text{resistance}, \\ \mu(E) \text{ high again at large } E &\Rightarrow \text{overcompensation backfires}. \end{aligned}$$

2. **Flat or monotonic behaviour for legitimate types.** Because the legitimate type receives a lower payoff l from retaining the measure, even mild suspicion is sufficient to induce modification. As a result, their behaviour is less sensitive to E and more likely to display a smooth or flat pattern.
3. **Disproportionate court escalation at extreme evidence levels.** Since $\mu(E)$ is high at both ends of the evidence spectrum, Player 2 is most likely to escalate to court either when evidence is too weak (suggesting concealment) or too strong (suggesting overcompensation). Intermediate E yields the lowest rate of escalation.

5 Data

To test the observable implications, I obtain 66,231 notifications from the online WTO TBT/SPS `epingalert.org` database between January 2010 and August 2024. I first classify the regulations using the document symbol — with or without "Add." or "Corr.". Members will initially submit the notifications with their initially proposed regulatory trade barriers, for example `G/TBT/N/UGA/680`. They may submit subsequent notifications, for example `G/TBT/N/UGA/680/Add.1` to inform the status of the regulation. The statuses may include notifying the adoption, modification or withdrawal of the regulation, or they may provide additional information or amend the commenting period for the regulation. I read the documents with symbol "Add." or "Corr." to determine if they are notifying the adoption. My variable of interest is whether the member will adopt the initial regulatory trade barriers straight away. I therefore use each regulation as unit of analysis: if the regulation does not have subsequent addendum or corrigendum, or if the addendum only states that the regulation has been adopted, it is classified as not modifying the proposed regulation; otherwise, it is classified as modifying the proposed regulation. Figure 2 shows the descriptive data. The average percentage of modifying the proposed regulation is 14.1%.

My main explanatory variable is the amount of evidence presented by the proposing country.

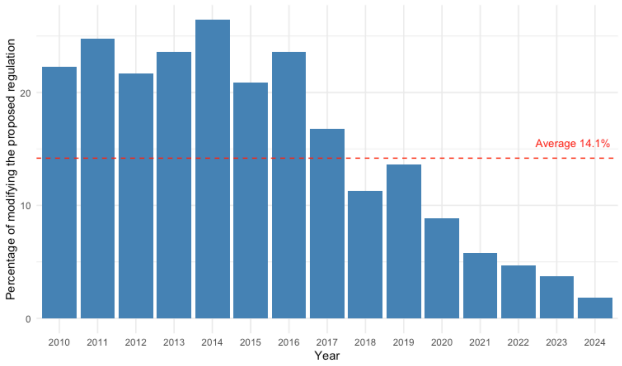

Figure 2: Descriptive data

As a proxy, I use the number of pages in the attached document that the proposing country provides with the notification. To determine whether the type of the regulation θ is legitimate or protectionist, I leverage the import penetration data. I map the product listed on the notification with the import export data from BACI and calculate the import penetration as product import divided by the sum of product import and product export. Low values (e.g. 0.1–0.3) may signal domestic self-sufficiency or limited international competitiveness, while high values (e.g. 0.6–0.9) suggest openness and reliance on foreign supply. A persistent level around 0.5, however, may indicate a policy-induced ceiling on imports—enough to avoid autarky but still consistent with protectionist intent.

To explore how the complexity of regulatory proposals interacts with protectionist motives, I estimate a regression using `pages_factor` as a categorical variable for document length. The variable groups proposals into three bins: 0 for proposals with zero pages (17,883 observations), 1 for those with 1–10 pages (13,008 observations), and 2 for those exceeding 10 pages (14,729 observations). These categories capture meaningful variation in the informational content of proposals. By interacting `pages_factor` with both the protectionist score (`protect`) and its square (`protect^2`), the model allows the marginal effect of protectionism on modification likelihood to vary across different levels of proposal complexity—testing whether longer or more detailed proposals reduce scrutiny or help mask protectionist intent.

6 Results

The results illustrate a non-monotonic relationship between import penetration and the predicted probability of regulatory modification, with a pronounced peak around a protection score of 0.45. This suggests that scrutiny is highest when foreign competition is neither negligible nor dominant—precisely the ambiguous middle ground where protectionist motives are most suspect. At this level, the amount of supporting documentation makes a substantial difference: providing over 10 pages of evidence reduces the probability of modification by more than 4 percentage points

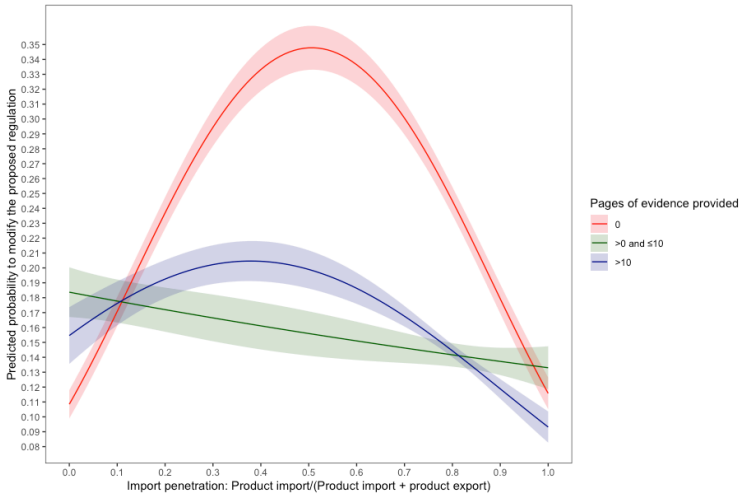

Figure 3: Results

compared to submitting only 1–10 pages. This gap highlights how detailed justification can meaningfully lower perceived protectionism when challenge risk is at its peak.

7 Conclusion

This paper has argued that the WTO functions not only as a legal enforcement mechanism but also as an epistemic regime in which states signal the legitimacy of trade-related regulations through the volume of supporting documentation. Using a novel dataset of over 50,000 TBT and SPS notifications, the analysis demonstrates that both under- and over-disclosure raise the likelihood of regulatory modification, especially under conditions of moderate import penetration where protectionist intent is most ambiguous. These findings challenge prevailing views that prioritize legal retaliation and instead highlight the informal but powerful role of evidentiary signaling in global trade governance. As regulatory conflicts increasingly revolve around politically sensitive issues like environmental and health standards, understanding how states use documentation to navigate international scrutiny becomes essential to designing more credible and responsive multilateral oversight mechanisms.

References

AQSIQ. 2021. "2020 niandu Zhongguo dui WTO qita chengyuan TBT tongbao de pingyi yijian huizong (Summary of China's Comments on the TBT Notifications of Other WTO Members in 2020)." *General Administration of Quality Supervision, Inspection and Quarantine of the PRC*, October.

Barton, John H, Judith L Goldstein, Timothy E Josling, and Richard H Steinberg. 2006. *The Evolution of the Trade Regime: Politics, Law, and Economics of the GATT and the WTO*. Princeton University Press.

Bonneuil, Christophe, and Les Levidow. 2012. "How does the World Trade Organization know?

The mobilization and staging of scientific expertise in the GMO trade dispute." *Social studies of science* 42(1): 75–100.

Borlini, Leonardo S. 2019. "A crisis looming in the dark: Some remarks on the reform proposals on notifications and transparency." *Questions of International Law* 63.

Busch, Lawrence, Robin Grove-White, Sheila Jasanoff, David Winickoff, and Brian Wynne. 2004. "Amicus Curiae Brief Submitted to the Dispute Settlement Panel of the WTO in the Case of EC-Biotech (WT/DS291, 292 and 293).". Available at: http://www.ecolomics-international.org/biosa_ec_biotech_amicus_academic2_ieppp_lancasteru_coord_0404.pdf (accessed 23 September 2011).

Busch, Marc L, and Krzysztof J Pelc. 2015. "Dispute Settlement in the WTO.".

Cho, Sungjoon. 2011. "From control to communication: Science, philosophy, and world trade law." *Cornell Int'l LJ* 44: 249.

Downes, Chris. 2012. "The Impact of WTO Transparency Rules: Is the 10,000 th SPS Notification a Cause for Celebration?—A Case Study of EU Practice." *Journal of International Economic Law* 15(2): 503–524.

Hancock, Alice, and Mercedes Ruehl. 2023. "Indonesia and Malaysia freeze trade talks with EU over palm oil." *Financial Times*, May 31.

Herwig, Alexia. 2008. "Whither Science in WTO Dispute Settlement?" *Leiden Journal of International Law* 21(4): 823–846.

Holzer, Kateryna. 2019. "Addressing tensions and avoiding disputes: Specific trade concerns in the TBT Committee." *Global Trade and Customs Journal* 14(3).

Horn, Henrik, Petros C Mavroidis, and Erik N Wijkström. 2013. "In the shadow of the DSU: Addressing specific trade concerns in the WTO SPS and TBT committees." *Journal of World Trade* 47(4).

Howse, Robert. 2000. "Democracy, science, and free trade: risk regulation on trial at the World Trade Organization." *Michigan Law Review* 98(7): 2329–2357.

Howse, Robert, and Joanna Langille. 2023. "Continuity and Change in the World Trade Organization: Pluralism Past, Present, and Future." *American Journal of International Law* 117(1): 1–47.

Jasanoff, Sheila. 2004. "Ordering knowledge, ordering society." In *States of knowledge*. Routledge , 13–45.

Jasanoff, Sheila. 2008. "30 Making Order: Law and Science in Action.".

Joerges, Christian. 2005. "Juridification patterns for social regulation and the WTO: A theoretical framework.".

Karttunen, Marianna B. 2020. *Transparency in the WTO SPS and TBT agreements: the Real Jewel in the Crown*. Cambridge University Press.

Kelley, Judith. 2004. "Who Keeps International Commitments and Why? The International Criminal Court and Bilateral Nonsurrender Agreements." *American Political Science Review* 99(3): 573–589.

Kelley, Judith. 2007. "International actors on the domestic scene: Membership conditionality and socialization by international institutions." *International Organization* 58(3): 425–457.

Lang, Andrew, and Joanne Scott. 2009. "The hidden world of WTO governance." *European Journal of International Law* 20(3): 575–614.

Lange, Jeva. 2023. "The Culture War Comes For Dishwashers." *Heatmap News*, May 10. https://heatmap.news/politics/dishwashers-biden-conservatives-culture-war.

Lester, Simon. 2014. "Finding the boundaries of international economic law." *Journal of International Economic Law* 17(1): 3–9.

Lester, Simon. 2022. "Where Did "Scientific Basis" Come From in Trade Rules?" *International Economic Law and Policy Blog*, October 2.

Lester, Simon, and William Stemberg. 2014. "The GATT origins of TBT Agreement Articles 2.1 and 2.2." *Journal of International Economic Law* 17(1): 215–232.

Mackrael, Kim, Gavin Bade, and David Luhnow. 2025. "Trump Wants to Rewire Every Other Economy With Trade Onslaught ; The president wants to end so-called non-tariff barriers, including EU regulations that keep out U.S. poultry." *Wall Street Journal*, April 9. https://www.wsj.com/economy/trade/trump-trade-war-economy-goal-c9210091.

Meyer, Timothy. 2021. "Cooperating without Sanctions: Epistemic Institutions versus Credible Commitments Regimes in International Law." In *International Law as Behavior*, eds. Harlan Grant Cohen, and Timothy Meyer. Cambridge University Press , 45–73.

Meyer, Timothy. 2022. "The Political Economy of WTO Exceptions." *Washington University Law Review* 99: 1299–1369.

Pevehouse, Jon C. 2005. *Democracy from Above: Regional Organizations and Democratization*. Cambridge: Cambridge University Press.

Pinchis-Paulsen, Mona. 2022. "Let's Agree to Disagree: A Strategy for Trade-Security." *Journal of International Economic Law* 25(4): 527–547.

Rigod, Boris. 2015. *Optimal regulation and the law of international trade: the interface between the right to regulate and WTO law*. Number 18 Cambridge University Press.

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.

Rosendorff, B Peter, and Helen V Milner. 2001. "The optimal design of international trade institutions: Uncertainty and escape." *International Organization* 55(4): 829–857.

Ruehl, Mercedes, Alice Hancock, and Emiko Terazono. 2023. "EU deforestation law triggers ire of its trading partners." *Financial Times*, February 5.

Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94(4): 819–835.

Sykes, Alan O. 1999. "Regulatory protectionism and the law of international trade.".

Tomz, Michael. 2007. *Reputation and International Cooperation: Sovereign Debt across Three Centuries*. Princeton, NJ: Princeton University Press.

Walker, Vern R. 2003. "The myth of science as a neutral arbiter for triggering precautions." *BC Int'l & Comp. L. Rev.* 26: 197.

Wolfe, Robert. 2013. "Letting the sun shine in at the WTO: How transparency brings the trading system to life." *World Trade Organization, Staff Working Paper* ERSD-2013-03.

Wolfe, Robert. 2020. "Reforming WTO conflict management: Why and how to improve the use of 'specific trade concerns'." *Journal of International Economic Law* 23(4): 817–839.

WTO. 2018. "WTO TBT Enquiry Point Guide: Making Transparency Work.".