"A Deepening/Widening Tradeoff? Evidence from the GATT and WTO."

by

David H. Bearce
Professor of Political Science and International Affairs
University of Colorado Boulder
david.bearce@colorado.edu

Max J. Chernoff
Doctoral Student in Political Science
University of Colorado Boulder

Cody D. Eldredge
Assistant Professor of Political Science
Oakland University

3rd draft (February 2021): Prepared for PEIO

Acknowledgements: We thank Brett Bessen, Brendan Connell, Judy Goldstein, Clara Park, Andy Phillips, Adrian Shin, and Jaroslav Tir for helpful comments. We also thank Brandy Jolliff Scott for her research assistance.

Abstract: Due to their consensus practice, this paper proposes that the General Agreement on Tariffs and Trade and its successor, the World Trade Organization, have experienced a deepening/widening tradeoff: as their membership increased (greater width), their effectiveness in promoting trade declined (lesser de facto depth). This hypothesis is tested with multiple samples from different datasets, a variety of statistical specifications, different estimators, and alternative measures of membership/ participation, first separating the GATT from the WTO, which are usually combined into a single variable, and then adding a width interaction variable for each institution. The gravity model results show that both trade institutions were their deepest when they had the fewest member-states and that their effectiveness declined, eventually becoming statistically insignificant, as more countries joined. As a quantitative case study, this paper provides some of the first evidence consistent with a tradeoff between depth and width within international organizations.

Scholars have long posited a potential tradeoff between the depth and width of international organizations (e.g., Oye 1986; Milner 1992; Downs, Rocke, and Barsoom 1996; Stone, Slantchev, and London 2008). The concept of "depth" captures the extent to which international agreements and institutions influence the behavior of member-states based on the idea that deeper regimes should be more effective at promoting cooperation between/among participating countries. The concept of "width" refers to the number of countries within the regime,[1] presumably with decision-making privileges and the ability to exit (von Borzyskowski and Vabulas 2019) if they are dissatisfied with its policies (i.e., formal members). If this tradeoff is empirically valid, then it represents one of the most serious challenges to received international cooperation theory: while regimes may help promote cooperation, this effect diminishes with size. On this basis, the more effective international regimes would be those with fewer member-states, thus limiting the scale of their cooperative influence.

The depth/width tradeoff is often advanced as a static hypothesis comparing regimes at their construction: those formed with more member-states should tend to demand less of these participants. However, the few cross-sectional tests of this hypothesis (e.g., Bernauer et al. 2013; Slapin and Gray 2014) find little support, consistent with Gilligan's (2004) argument that there should be no depth/width tradeoff when member-states are allowed to set their policies at different levels.

[1] It is also important to clarify what the concept of "width" does not include. We are interested in the depth/width tradeoff also known as the "depth versus participation dilemma" (Bernauer et al. 2013). Thus, our concept of width does not include the scope of issues covered by the international regime. Indeed, we would argue that "scope" is more akin to depth, especially for trade regimes (Dür, Baccini, and Elsig 2014) consistent with Koremenos, Lipson, and Snidal's (2001) treatment of membership rules (e.g., who and how many can join) and the scope of issues covered as separate dimensions of institutional design; the former relates to width and the latter to de jure depth.

Indeed, at regime construction, Bearce, Eldredge, and Jolliff (2015, 232) even provide evidence for a positive relationship between de jure depth and width for Koremenos's (2005) sample of international agreements.

However, most international regimes are not intended to be static. Many seek to deepen their cooperation and also to attract more countries, thereby expanding their width. Thus, even if there is no depth/width tradeoff at regime formation, a similar tradeoff could emerge dynamically (Downs, Rocke and Barsoom 1998): as an international regime acquires more member-states, its effectiveness, or ability to promote cooperation, may decline. This related proposition, which considers the effectiveness of an international regime as it evolves, might be better termed as a *deepening versus widening* tradeoff. In this paper, we explore the possibility of this dynamic tradeoff, focusing on the General Agreement on Tariffs and Trade (GATT) and the World Trade Organization (WTO) based on their consensus practice (Steinberg 2002).

We focus on these sequential multilateral trade regimes for three different reasons. First, objectively measuring their "depth" is feasible, unlike for many other international regimes in other issue-areas. Indeed, there is not much empirical evidence bearing on the relationship between depth and width because while the latter is relatively easy to capture (by counting the number of member-states, which may change over time), capturing the former is more challenging. However, the gravity model of bilateral trade offers a way to estimate the *de facto* depth of the GATT and WTO.[2] Included

[2] *De facto* depth is defined by how much the regime actually promotes cooperation among member-states. In contrast, *de jure* depth refers to how much the regime is supposed to promote cooperation based on what member-states agree to do or not do. This latter concept needs to be coded by looking at the written agreements associated with membership, but de facto depth can be estimated based on member-state behavior. Here, we are interested in de facto depth, which may differ greatly from de

on the right-hand side in this specification, the coefficient on a trade regime joint membership variable estimates how much it increases commerce between member-states, or fosters trade cooperation, which is precisely what the concept of "depth" means in this issue-area.

Second, these multilateral trade regimes do not represent "easy" cases for any form of a depth/width tradeoff. As Gilligan (2004, 460) noted about the GATT's creation, "one finds no references to a 'broader versus deeper' debate." Indeed, we accept that when it began operations in 1948, the GATT was arguably the deepest *and* the widest trade regime in existence. However, did the GATT become more or less effective at promoting trade cooperation as more countries joined? Furthermore, after the long Uruguay Round of trade negotiations (1986-1994), the presumably deeper WTO entered into force in 1995 with around 100 member states, *suggesting depth with width*. But how much more did the WTO promote trade compared to the supposedly shallower GATT? Likewise, did the WTO's trade effect change (and in what direction) as it further added member-states into the 21st century?

Third, the GATT and WTO represent substantially important cases since evidence about their de facto depth bears on another major political-economy debate. Concerning the effectiveness of these global trade regimes, a prominent Economics paper (Rose 2004) provided evidence that the GATT/WTO had either no statistically or substantively (when statistically) significant impact on member-state trade, but did not develop any particular argument for why this should be true. Political scientists (Tomz, Goldstein, and Rivers 2007; Goldstein, Rivers, and Tomz 2007) countered that Rose had not properly captured GATT/WTO obligations and, once non-formal participants were included, the coefficient on the global trade regime became statistically significant and substantively large. On

---

jure depth, especially for institutions like the GATT and WTO that potentially suffer from enforcement problems.

this basis, no theory would be needed to explain the shallow impact of the GATT/WTO because its trade effects were, in fact, quite deep.

However, subsequent bilateral trade models (e.g., Eicher and Henn 2011) that include multilateral resistance controls (Anderson and van Wincoop 2003), which are now considered to be a standard part of the gravity specification, show that no matter how the GATT/WTO is measured, either in terms of formal membership or when including non-formal participants, the overall effect of these multilateral trade regimes has been relatively small. We thus still need to explain why the GATT and WTO have had a perhaps surprisingly weak average effect on trade cooperation, requiring a theory that Rose (2004) did not provide.[3] Has their influence varied based on the number of countries within these sequential institutions? And if the impact of the GATT and WTO has differed on this basis, then did they become *less* trade effective as more countries joined, thus weakening their average annual effect? If so, then this would represent evidence consistent with a deepening/widening tradeoff, and we need to explain why such a tradeoff might be expected to emerge within these multilateral trade organizations. We now turn to this task.

1. The Argument

In this section, we discuss the GATT and WTO as separate, but sequential, multilateral trade institutions, proposing that both became less effective, or shallower on a de facto basis, as more states joined due largely to their practice of consensus. Consensus decision-making within these regimes is best understood as a practice, and not as a formal rule, since the 1947 Agreement permitted majority voting. But this practice, which effectively allows every member-state to exercise veto power, quickly

[3] As Rose (2004, 111 *emphasis in original*) concluded: "Of course, the most interesting issue that remains is *why* the GATT/WTO does not seem to have had much of an impact on trade."

became the norm. Indeed, the "principle of consensus decisionmaking is perhaps the most significant and consistent feature of GATT administration" (Stiles 1996, 122).

First, we argue that the consensus practice *in bargaining* effectively created a floor for how much these institutions could deepen on a de jure basis (e.g., negotiate new agreements to restrict other forms of trade protection beyond tariffs and to expand coverage beyond manufactured products into agriculture and services) through their multiple negotiation rounds, although we will discuss the possible exception of the Uruguay Round below. Not only did consensus bargaining create a limit on de jure deepening, but this limit also worked to decrease the de facto depth of these institutions. The latter occurred because it allowed member-states to shift their trade protection from prohibited tariffs towards non-tariff barriers to trade (Kono 2006) and undervalued currencies, which these multilateral institutions failed to prohibit due to the consensus practice in bargaining.

Second, the consensus norm negatively affected the *enforcement* of existing rules, also making these institutions less effective with more member-states, or contracting parties. As described by Kalla (1986, 90 *capitalization in original*), the panel procedure was the GATT's "basic model in dispute settlement." In response to a perceived rule violation, a member-state could ask for a panel to investigate, but the "adoption of [subsequent] panel reports requires a consensus by the CONTRACTING PARTIES, which means the losing country can block a negative report." As further discussed by Lowenfeld (1994, 479-80), the blocking of a panel report "happened not infrequently. Indeed, in some instances a contracting party was able to block, or at least delay, appointment of a panel in the first place."

Given the blocking threat posed by *any* member-state, including but not limited to the defendant, GATT panel reports tended to be weak. As described by Pescatore (1993, 13), these panels "do their best to avoid controversial issues and try to present their arguments as being the expression of the obvious. Reports thus become exercises in understatement and in the expression of legal

platitudes." Indeed, even with weak panel reports, the GATT's enforcement of them was problematic with "only two-fifths of rulings for the complainant result[ing] in full compliance by the defendant. In nearly a third, defendants fail to comply at all, effectively spurning panel rulings (as a result, some of these rulings were not invested with formal legal authority by virtue of the defendant's veto)" (Busch and Reinhardt 2000, 167-8). On this basis, it became more difficult for the GATT to enforce the already prohibited forms of trade protection (e.g., tariffs on manufactured products) with a greater number of countries that could block a panel report, further reducing the institution's effectiveness in promoting trade between member-states.

Despite its consensus practice, it is not hard to understand why the GATT began as a relatively effective, or deep, multilateral trade institution when it started operations in 1948 after concluding the 1947 Geneva I negotiation round. Not only did it have its fewest number of member-states, but "after the Second World War, tariffs were very high" and "there was a strong incentive to make some of these easy cuts to establish a track record" (Martin and Messerlin 2007, 350-1). However, the tariff cuts in the subsequent negotiation rounds (especially Annecy 1949 with 20 members, Torquay 1950-1 with 33, Geneva II 1955-6 with 35, and Dillon 1960-1 with 40 members) were far more modest (ibid, 349) with no real progress to prohibit other forms of trade protection or to expand its coverage beyond manufactured products as the GATT grew wider.

In fact, by the late 1970s, observers like John Jackson (1978, 96) were describing a "near breakdown of the international institutional framework for the trading system" that had "served us remarkably well heretofore". On this point, he (ibid, 97) specifically noted 1) the GATT's inability to negotiate new rules to prohibit the emerging forms of trade protection used as tariff substitutes and 2) the "ease with which parties can evade" the existing rules about tariffs on manufactured products, which are precisely the two problems highlighted above associated with deepening within a widening regime that practices consensus. Indeed, even after the conclusion of the 74-month Tokyo round in

1979, many observers still saw this trade institution as failing due to its growing size. According to Thomas Graham (1983, 128): "[p]reviously, that GATT operates by consensus was a strength" but "the need for consensus led to economic stalemate as GATT's membership diverged." On this basis, he concluded that "the 88-member GATT is [now] less wieldy than when the 24 mostly like-minded governments drew it up in 1947."

In an effort to revive, or deepen, what appears to have become a shallow, or ineffective, global trade regime, the Uruguay round of international negotiations began in 1986. The fact that this bargaining phase took so long to conclude is arguably consistent with an internal depth/width tradeoff, but the fact that it was concluded at all with so many member-states may appear inconsistent. However, as Steinberg (2002, 366) described, it was the "raw use of power" by the United States, joined by European Community representing their member-states in bloc, that finally made it possible to "close the Uruguay Round" after more than eight years of protracted bargaining. Thus, we are not arguing that a wider institution can never deepen. Indeed, deepening with widening may sometimes be possible consistent with what Milner (1992) termed "imposed" cooperation or with what Martin (1994) identified as "coercive" cooperation. Instead, our argument proposes that deepening becomes harder, or rarer, as the institution gets wider with a general tendency in the opposite direction: a consensus institution tends to become less deep at least on a de facto basis as more countries join.

Thus, this exceptional power-based effort to overcome the bargaining problem associated with a consensus practice produced a different institution not only in terms of its name - the World Trade Organization - but also in terms of its structure, adding a General Agreement on Trade in Services (GATS) and an agreement on Trade-Related Property Rights (TRIPS), among many other new features. The WTO's dispute settlement procedures also became more legalized (Rosendorff 2005) with panel reports that could now only be rejected by consensus. By all appearances, the new WTO was a much deeper institution than what it had replaced. As described at the time by Martin and

Winters (1995, 3), the WTO represented "the most fundamental reform of the world trading system since the establishment of the General Agreement on Tariffs and Trade (GATT) in 1947."

But how much deeper was the new WTO? Relative depth cannot be established in a gravity model specification that treats the GATT and WTO as effectively the same institution when they are packaged together into a single right-hand side variable (among other innovations, we will split the standard GATT/WTO variable into separate GATT and WTO variables in the next section). Indeed, other contemporaneous observers pronounced the WTO as "at best a modest strengthening" over the GATT (Stiles 1996, 119). One reason for this skepticism about how much depth the new WTO had achieved and, more importantly, whether it could be sustained as it further widened was that the new multilateral trade institution retained the consensus norm, now codified in Article IX of the WTO Agreement on Decision-Making: "the WTO shall continue the practice of decision-making by consensus followed under the GATT 1947" (Footer 1996/1997, 679).

Indeed, it is not hard to see how the continued consensus bargaining practice, known also as the "single undertaking" within the WTO, has been an obstacle to deepening through the Doha Development Agenda, or to completing the Doha rounds that began in 2001. As described by Hartman (2013, 413), the "Single Undertaking principle makes it particularly challenging for the Doha Rounds to come to a resolution, since there is no agreement unless 'everything is agreed'" to "by a broad cross section" of countries with "widely divergent national interests". And as reported following the 2015 WTO Ministerial Conference in Nairobi, Kenya, the "members of the World Trade Organization have effectively ended the Doha round of negotiations" after 14 years of bargaining stalemate.[4] Thus, since the establishment of the WTO in 1995, there has been no successful

[4]https://www.nytimes.com/2016/01/01/opinion/global-trade-after-the-failure-of-the-doha-round.html .

completion of even a single bargaining round that would deepen this institution at least on a de jure basis. And it has arguably become less deep on a facto basis as member-states continue to substitute non-prohibited forms of trade protection for prohibited tariffs.

Likewise, despite the reform that stopped the ability of a single member-state to reject a panel recommendation, consensus remains a problem for enforcement within the WTO because any member-state can block appointments to its seven-person Appellate Body, which requires at least three judges to hear a case. Indeed, the United States under both the Obama and Trump administrations has exercised this veto option multiple times, leaving the Appellate Body with only one judge as of December 2019. This recent development received much attention, prompting some very negative predictions about the future of this institution. As one observer wrote: "The recent US failure to appoint a judge at the WTO's appellate court has meant that it will no longer be able to intermediate trade disputes. This means that the WTO is [effectively] dead."[5] And while it is the United States that has most recently exercised its veto option, which might also be the case if the WTO operated like the United Nations Security Council in giving a veto to the large founding countries, the WTO's consensus practice means that *any* other member-state could do like the United States, regardless of its size and status within the institution. On this basis, consensus remains a potential problem for WTO enforcement even if the United States under the Biden administration stops its practice of vetoing Appellate Body judges.

While this story suggests a deepening/widening tradeoff first within the GATT and later within the WTO, it remains important to consider how well it matches with the quantitative evidence. Accordingly, we hypothesize that both the GATT and the WTO began as trade effective institutions,

[5] https://www.bruegel.org/2019/12/the-wto-is-dead-long-live-the-wto/ .

increasing commerce between member-states, *but their positive impact weakened as more countries joined.*[6] This hypothesis is notable because it runs contrary to what one might expect based on the understanding that the accession requirements for joining these trade institutions, although they vary by country, have tended to increase over time (Jones 2009, Allee and Scalera 2012, Davis and Wilf 2017). More rigorous accession terms for later entrants should tend to add depth, making these multilateral institutions appear more (not less) effective with additional member-states. Likewise, based on the successive negotiation rounds within the GATT intended to increase its de jure depth, one might expect the institution to become also de facto deeper even as more countries joined. We argue to the contrary.

[6] Our focus on how much the GATT and WTO increased commerce between member-states accords with trade expansion being the primary goal of these multilateral institutions. Consistent with this understanding, the 1994 Marrakesh Declaration (April 15), issued at the end of the Uruguay Round to mark the transition from the GATT to the WTO, opens by saluting "the historic achievement represented by the conclusion of the Round, which… will strengthen the world economy and lead to more trade". Similarly, at its 50th anniversary in 1998, the WTO specifically trumpeted its goal of trade expansion: "Since the General Agreement on Tariffs and Trade began operating from Geneva in 1948, world merchandise trade has increased 16 fold" ranking "among the great international economic achievements of the post-world war era" (Rose 2002, 2). The GATT and WTO certainly have other goals, including decreased trade barriers. But the latter is best understood as a means to the primary end of increasing commerce. Likewise, reducing the volatility of trade flows can be identified as another important goal, but we do not consider this related objective in this paper.

2. The Evidence

The Gravity Model

Our hypothesis is tested using a gravity model of bilateral trade that first separates GATT from WTO membership to estimate the depth of each institution and then adds variables capturing their respective widths. With the directed dyad/year as the unit of analysis, the gravity model takes country *i*'s *Imports* from country *j* in year *t* as its dependent variable. As independent variables, our analysis focuses on dichotomous indicators for joint formal membership in the *GATT* and *WTO* with the coefficient on these variables providing the average effect of the institution on annual trade between member-states. Thus, the *de facto* depth of a trade regime is estimated. However, the width of the trade regime must be coded into separate independent variables that count the number of member-states in the GATT and WTO, which varies by year.

Using information from the World Trade Organization,[7] we calculate the *effective* number of countries in these two sequential institutions in each year beginning in 1948. Since a country could join at any point during the year, we code it as becoming an effective member in the first year when it was a member for at least six months, or for a majority of the year.[8] We would obtain very similar

[7] See: https://www.wto.org/english/thewto_e/gattmem_e.htm and https://www.wto.org/english/thewto_e/whatis_e/tif_e/org6_e.htm .

[8] For example, it is often noted that 23 countries participated in the 1947 Geneva I negotiations to form the GATT, which began operations in 1948. But only ten of these countries were effective member-states in 1948 by officially joining during the first half of that year: Australia, Belgium, Canada, Cuba, France, Luxembourg, Netherlands, South Africa, the United Kingdom, and the United States. Brazil, Sri Lanka, India, New Zealand, Norway, Pakistan, Zimbabwe (then South Rhodesia) joined in the latter half of 1948, with Chile joining in 1949. Burma, China, Czechoslovakia, Lebanon,

results based on the number that joined at any point during the year, but the former appears to offer a more valid operationalization of "width" as it was actually experienced within the institution during that calendar year. Consistent with our argument that a deepening versus widening tradeoff emerged within the GATT and WTO *due to their consensus practice*, our measure of width does not weight (by GDP or some other measure of market size or power) the number of member-states because the consensus-based veto options related to bargaining and enforcement could be exercised by any country within the institution regardless of its size or status.

Figure 1: Effective Number of Member-States in the GATT (1948-1994) and WTO (1995-2020).

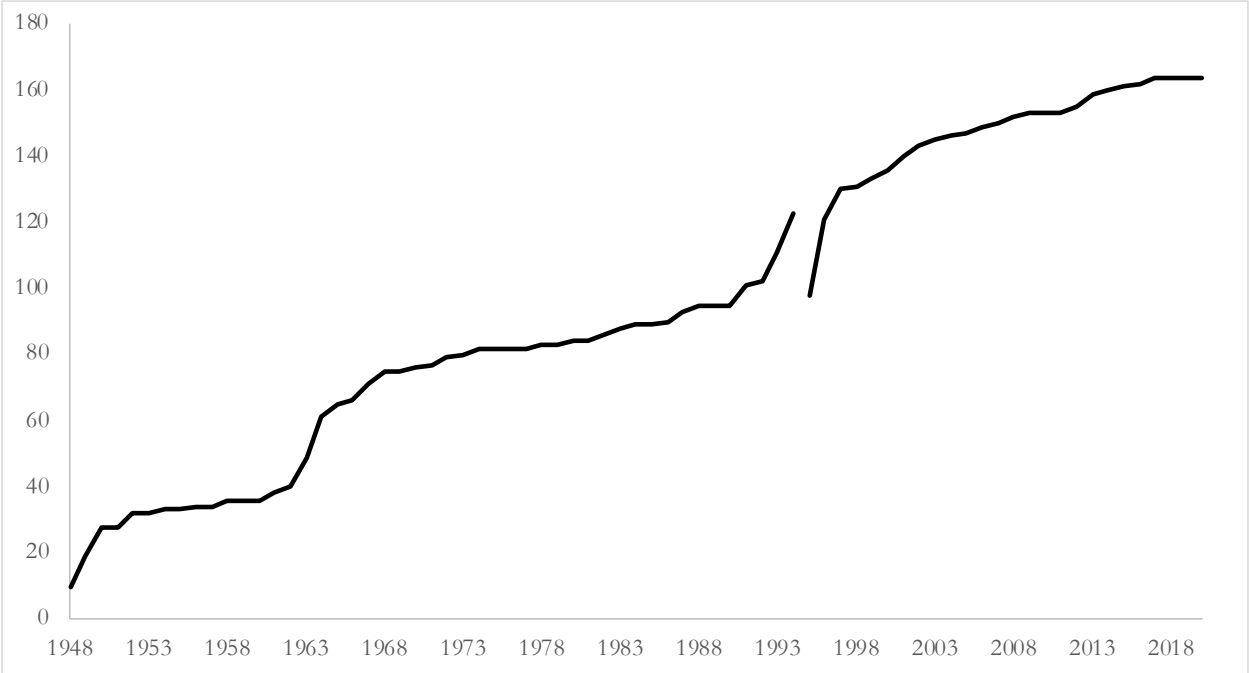


Figure 1 plots this unweighted count of the effective number of member-states for the GATT and WTO by year. Not surprisingly, there appears to be a structural break in 1995 when transitioning from the GATT to the WTO based on ratification delay within certain countries that were GATT

and Syria also participated in the 1947 negotiations, but did not officially join the GATT, according to the World Trade Organization (see the links above).

members in 1994 but did not officially join the WTO immediately. Thus, while the width of these multilateral institutions tends to increase over time, this increase is not monotonic with a decrease in the effective number of member-states in 1995 that is not restored to at least the 1994 value until 1997.

Using the information from Figure 1, our gravity models follow the basic structure of equation (1), where *Joint Membership* represents a dichotomous variable indicating whether countries *i* and *j* are both members of the same multilateral institution, either the *GATT* or *WTO*, in year t. These membership dummies are then interacted with their *Width* measure, coding the effective number of members in the respective institution in year *t* if both *i* and *j* are members. If *i* and *j* are not joint members of the multilateral trade institution, then their *Width* is equal to *0* for any year. This coding of *Width* provides both cross-sectional variation (between joint members and other dyads) and temporal variation (for the joint member dyads).

$$Imports_{ijt} = B_1*Joint\ Membership_{ijt} + B_2*(Joint\ Membership_{ijt}*Width_{ijt}) + B_3*Width_{ijt} + Controls_{ijt} + e_{ijt} \quad (1)$$

Given the interaction variable, the coefficient on the *Joint Membership* constitutive variable (B_1) estimates the de facto depth of the institution in question when its *Width* is equal to 0. Obviously, this is an out-of-sample value since the *GATT* and *WTO* always had a non-zero number of members. But B_1 combined with B_2, (the coefficient on the interaction variable) with the latter multiplied by the effective number of member-states provides the marginal effect of *Joint Membership* given its institutional *Width*. Our hypothesis that the GATT and WTO began as trade effective institutions, but with a weakening impact as more countries joined, predicts that B_1 should be positive, while B_2 should be negative.

Equation (1) is written to also include a *Width* constitutive variable with its coefficient capturing the effect of institution's width when *Joint Membership* is equal to *0*, or for non-joint member

dyads. However, not only is B_3 theoretically expected to be zero (i.e., the institution's width should have no direct effect on trade involving non-members), this constitutive variable must drop in our estimation of equation (1) because it is perfectly collinear with the interaction variable. This collinearity stems from the fact that *Width* is nested within *Joint Membership*, or only has a non-zero value when *Joint Membership* is equal to 1. This nesting means that *Joint Membership*Width* has the same non-zero value as *Width* when *Joint Membership* is equal to 1 and always has a zero value when *Joint Membership* is equal to 0. Thus, either *Joint Membership*Width* or *Width* must drop from the specification, and we choose to retain the interaction variable.[9]

We estimate several different versions of equation (1) using existing gravity model datasets (Bearce, Eldredge, and Jolliff 2016; Goldstein, Rivers, and Tomz 2007), that should be familiar to many readers, thus allowing one to more easily compare our conditional estimates of the GATT's and WTO's impact on bilateral trade with the unconditional estimates that already appear in the literature. Using these datasets, we present three parallel sets of statistical results, also varying the gravity specification (e.g., different sets of fixed effects) and estimation (e.g., log linear OLS and poisson pseudo-maximum likelihood) to better assess the robustness of results.

[9] Alternatively, we could retain the *Width* constitutive term, dropping the interaction term, to estimate an additive linear model where the marginal effect of a trade institution comes from the *Joint Membership* coefficient plus the coefficient on *Width* multiplied by the number of member-states. As one can observe, this is the same math for calculating the marginal effect with the interaction variable and would provide an identical result.

OLS with 3-Way Fixed Effects

First making use of the dataset from Bearce, Eldredge and Jolliff (2016) with annual coverage from 1948 to 2012 (N=681,086), we estimate OLS models of the logged value of country *i*'s *Imports* from country *j* in year *t* (measured in constant 2009 US dollars).[10] The gravity model is traditionally estimated in a log linear form so that the coefficients provide the percentage change in *Imports*. Consequently, the dependent variable becomes the log of *Imports*, and the independent variables enter either as dichotomous indicators (e.g., GATT WTO joint membership) or as the logged value of a continuous variable (e.g., GATT WTO width).

Our control variables address potential sources of unobserved heterogeneity when trying to estimate the impact of the GATT and WTO. First, we include a dummy variable for joint membership within a preferential trading arrangement (*PTA*), most of which are nested within GATT and/or WTO dyad/years. Thus, the omission of a PTA control would bias our estimate of *GATT* and *WTO* effects, likely in a positive direction. Including the *PTA* dummy also allows us to compare the estimated effect of narrower PTAs with the wider GATT and WTO, which bears on the depth versus width tradeoff across different types of trade regimes. Second, we include directed dyad fixed effects to proxy the trade heterogeneity based on attributes related to the pair of states (e.g., trade barriers between them and their common history/experience).

Third, to address "multilateral trade resistance" (Anderson and van Wincoop 2003),[11] we also add fixed effects for the importer/year and the exporter/year. These two additional sets of country-

[10] Following Subramanian and Wei (2007), we drop the observations associated with trade flows of less than $500,000.

[11] The concept of "multilateral trade resistance" comes from the understanding that the amount of trade between two countries is determined not only by the barriers between them but also by the

specific year fixed effects also control for any time trends (on a country-specific basis) that might confound our estimation of GATT and WTO width effects. Without these fixed effects, one could argue that any observed width effect is simply some temporal influence (e.g., technological change) unrelated to the number of countries within these trade institutions. However, with country-specific year fixed effects, the temporal trends have been removed from the data. Indeed, we cannot include the standard set of year fixed effects (common to all countries and often used to proxy time trends) because they drop due to perfect collinearity with the importer/year and exporter/year fixed effects.

Of course, including these three sets of fixed effects cause most of the standard gravity model control variables to fall from our specification. For example, the inclusion of directed dyad fixed effects eliminates variables such as the distance between the countries, whether or not they share a language or border, their combined land area, and the number of landlocked or island countries within the dyad. Likewise, the inclusion of importer/year and exporter/year fixed effects eliminates any variable measuring the properties of a single country within the dyad, including those that change over time such as their economic size, population, and level of economic development. These country-specific year fixed effects also cause the control variable measuring if only one member of the dyad was a member of the GATT and/or WTO to drop due to collinearity. We note this exchange of specific control variables (whose coefficients are irrelevant for our hypothesis) for three-way fixed effects because while our specification may seem sparse in terms of the familiar gravity model controls, it is, in fact, quite rich with more than 45,000 fixed effects: approximately 25,000 directed dyads, 10,000 importer/years, and another 10,000 exporter/years.

---

barriers that each country faces with other countries, which changes over time. This concept can be captured with time dummies for each country within the dyad (i.e., importer/year and exporter/year fixed effects).

Table 1: OLS Models of *Imports* (logged) with Three-Way Fixed Effects.

Model:	1.1	1.2	1.3	1.4	1.5	1.6	1.7
GATTWTO	0.08 (0.02)						
GATT		0.07 (0.03)	0.09 (0.03)		0.33 (0.19)	0.48 (0.19)	
*GATT*GATT Members* (logged)					-0.06 (0.05)	-0.10 (0.05)	
WTO		0.14 (0.05)		0.19 (0.04)	7.50 (2.34)		7.27 (2.10)
*WTO*WTO Members* (logged)					-1.49 (0.48)		-1.44 (0.43)
PTA	0.41 (0.02)	0.41 (0.02)	0.28 (0.03)	0.22 (0.02)	0.41 (0.02)	0.28 (0.03)	0.23 (0.02)
Constant	16.65 (0.01)	16.64 (0.01)	16.48 (0.01)	16.88 (0.03)	16.66 (0.02)	16.49 (0.01)	16.93 (0.03)
Directed Dyad Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer/Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter/Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.87	0.87	0.87	0.92	0.87	0.87	0.92
N	681,086	681,086	399,946	279,380	681,086	399,946	279,380

Unit of analysis is the directed dyad/year.
OLS coefficients with robust standard errors clustered on the directed dyad in parentheses.
GATT Members and *WTO Members* constitutive variables drop due to collinearity with their respective interaction variables.

Table 1 presents a sequence of OLS gravity models with *Imports* (logged) as the dependent variable. To better link our conditional estimates of the GATT and WTO based on their width variation, we begin with the unconditional estimates that are standard in the literature. Model 1.1 is the traditional model with GATT and WTO joint membership combined into a single dummy variable (*GATTWTO*) with its effect estimated next to the *PTA* dummy. Both enter with positive coefficients that are statistically significant at conventional levels, but the *PTA* coefficient is also significantly larger than that of the *GATTWTO*. While results like these appear in the existing literature, they have not been discussed in terms of the depth/width tradeoff. PTAs are clearly narrower institutions than the

GATT and WTO, and our results also show them, at least on average, to be significantly deeper, consistent with this proposed tradeoff at least when comparing the two different types of trade institutions: bilateral/regional versus multilateral/global.

In model 1.2, we depart from the standard specification that combines the GATT and WTO by including separate variables for each institution to compare their average annual effect on trade between member-states. Consistent with the argument that the Uruguay Round negotiations worked to deepen the multilateral trade institution (Martin and Winters 1995), the *WTO*'s coefficient is twice as large as the *GATT*'s, although this difference is not quite significant at conventional levels.[12] However, our result showing that the WTO's average trade effectiveness is not weaker than the GATT's bears some discussion. Others have argued that one should observe a weaker effect for the WTO because "after the Uruguay Round", marking the start of the WTO, "only a handful of countries remained outside the regime: a few Middle Eastern nations, formerly Communist countries, and microstates" (Goldstein, Rivers, and Tomz 2007, 56). Indeed, this membership variation may produce problems in a specification *without* importer/year and exporter/year fixed effects, but our inclusion of these country time trends absorbs the effect of any country-specific attribute, including those that vary over time (e.g., being a petro-state, having a former Communist government, and the country's size). And with this specification, we observe results more consistent with the standard expectation that the WTO was deeper on average than the regime that it replaced.

It is also important to note that there may be hidden model misspecification when combining the GATT years with those that include the WTO. Not only is there no overlapping coverage between these two variables to require that their years be combined into the same sample, but the combined sample also forces all of the control variables (captured through fixed effects) to have the same effect

[12] The F test for $GATT=WTO$ is 2.03 (p=0.15).

across both the GATT and the WTO eras, which may be invalid. We thus split the sample into the GATT era and WTO era in models 1.3 and 1.4, respectively. In model 1.3 the *GATT* coefficient enlarges by about 30 percent compared to the same in model 1.2, while in model 1.4 the *WTO* coefficient expands by about 40 percent compared to model 1.2. Indeed, the *WTO* coefficient in model 1.4 is significantly larger than the *GATT* coefficient in model 1.3,[13] consistent with the understanding that the WTO has been, at least on average, a deeper trade institution than was the GATT.

Having provided some validity for our decision to separate the GATT from the WTO, we now test our hypothesis by interacting these joint membership variables with their respective width measures: *GATT Members* (logged) and *WTO Members* (logged). As stated earlier, the deepening versus widening hypothesis predicts that the *GATT* and *WTO* constitutive terms should enter with positive signs and their interaction terms with negative signs to indicate declining de facto depth as these multilateral trade institutions became wider. The full sample results in model 1.5 accord with these expectations with the coefficients only enlarging in the expected direction when looking the split sample results in models 1.6 and 1.7.

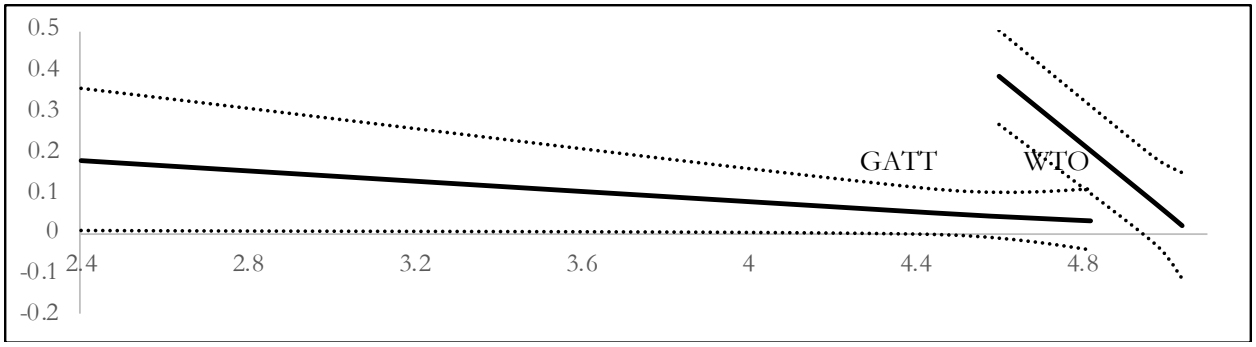
More important than the individual coefficients in these models are their combined effect at different width values. With a separate panel for each model, Figure 2 graphs the marginal effects of the *GATT* and/or the *WTO* (on the y-axis) for different values of their width (on the x-axis). It should be noted that since the marginal effect of the *GATT* and *WTO* provides their de facto depth, Figure 2 effectively plots *depth over width* with a negative slope showing a tradeoff between these two dimensions. As illustrated in the first panel (from model 1.5), the GATT was its most trade effective

[13] The F test for *WTO*=.09 in model 1.4 is 6.01 (p=0.01). Correspondingly, the F test for GATT=.19 in model 1.3 is 15.12 (p=0.00).

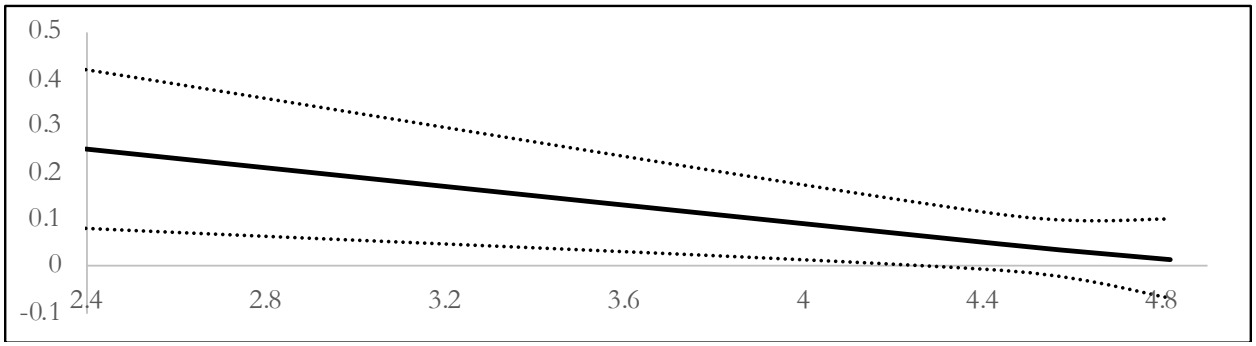
when it had the fewest members (10 effective members corresponding to a logged value of 2.4) with a marginal effect of 0.18, which is more than two times larger than its average annual effect in model 1.2. However, the marginal effect of *GATT* becomes statistically *in*significant when *GATT Members* (logged) passes 4.0, corresponding to roughly 60 members, which was the width reached by the GATT around 1964 when it started the Kennedy round of negotiations. The GATT era results from model 1.6 show a similar pattern with an even steeper tradeoff between de facto depth and width in the second panel of Figure 2.

Figure 2: The Marginal Effect of *GATT* and *WTO* Conditioned on Width: *_Members* (logged).

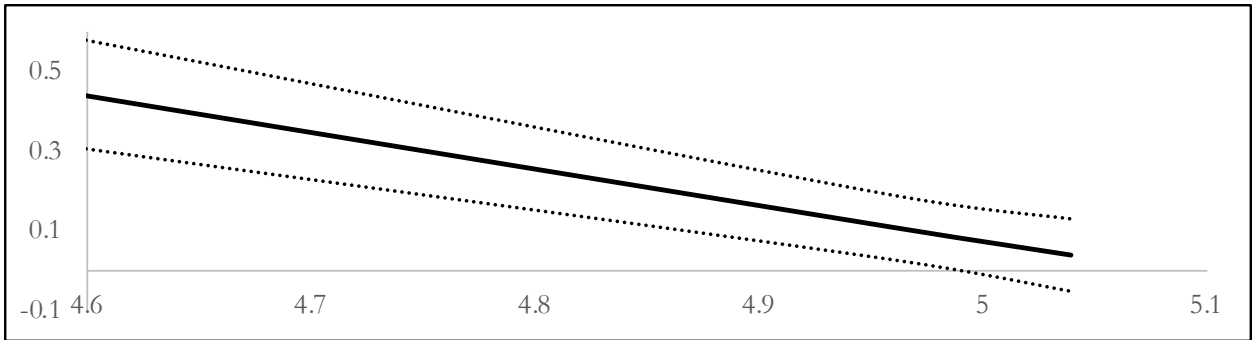
Model 1.5: GATT and WTO



Model 1.6: GATT Era



Model 1.7: WTO Era



The x-axis provides the marginal effect, and the y-axis provides the number of *_Members* (logged). The solid line indicates the point estimate with the dashed lines indicating 95 percent confidence intervals.

The deepening/widening tradeoff is even starker for the WTO. From model 1.5, this institution was its most trade effective in 1995 with its fewest number of members (98 effective members corresponding to a logged value of 4.6). At this width, the marginal effect of the WTO was 0.39, which is almost three time greater than its average effect from model 1.2. However, its de facto depth declined quickly, becoming statistically insignificant as *WTO Members* (logged) approached 5, corresponding to about 148 members, which was the width reached by 2005 after the launch of the Doha round. The third panel of Figure 2 from model 1.7 reveals similar results for the WTO in closer detail.

PPML with 3-Way Fixed Effects

The gravity model is now sometime estimated not as a log linear OLS but instead using poisson pseudo-maximum likelihood (PPML) with three-way fixed effects (as employed above) *and* with the standard errors clustered on all three of these dimensions. This PPML gravity specification has proven to be highly restrictive with scholars finding very weak effects not only for the GATT WTO but also for narrower and deeper regional arrangements such as Economic and Monetary Union in Western Europe (Larch et al. 2019). Indeed, using a PPML specification, Esteve-Pérez, Gil-Pareja, and Llorca-Vivero (2020) report that the wider multilateral institutions do not increase trade at all with their

combined GATT WTO dummy tending to enter with a *negative* sign that is sometime statistically significant (although they offer no explanation for this unexpected result).

While this is a very restrictive gravity specification, sometimes delivering what one might interpret to be implausible results, it is nonetheless useful to see if our hypothesis can be supported using PPML estimation with three-way fixed effects and three-way clustering of the standard errors. While poisson regressions can be described as log linear models, the PPML estimation takes the unlogged values of *Imports* as its dependent variable with exponentiation applied to the right-hand side variables, allowing one to retain the observations associated with zero values of trade. Thus, while we continue to use the gravity model dataset from Bearce, Eldredge and Jolliff (2016), our sample for the PPML regressions now expands to over a million directed dyad/year observations (N=1,039,979), dropping no observations associated with small values of trade.

Following the same sequence of models as in Table 1, Table 2 reports our PPML results with this larger sample, three-way fixed effects, and three-way clustering of the standard errors. Like Esteve-Pérez, Gil-Pareja, and Llorca-Vivero (2020), the combined *GATTWTO* joint membership variable is statistically insignificant, even entering with a negative sign in model 2.1. And consistent with the results reported by Larch et al. (2019), the *PTA* coefficient is also strongly attenuated compared to the same in model 1.1, but it remains positively signed and statistically significant. In model 2.2, we measure *GATT* joint membership separately from the *WTO*. While the positive *WTO* coefficient is not statistically different from zero, it is nonetheless significantly greater than the *GATT*'s also insignificant but negative coefficient.[14]

[14] The F test for *WTO*=*GATT* in model 2.2 is 4.36 (p=0.04).

Table 2: PPML Models of *Imports* with Three-Way Fixed Effects.

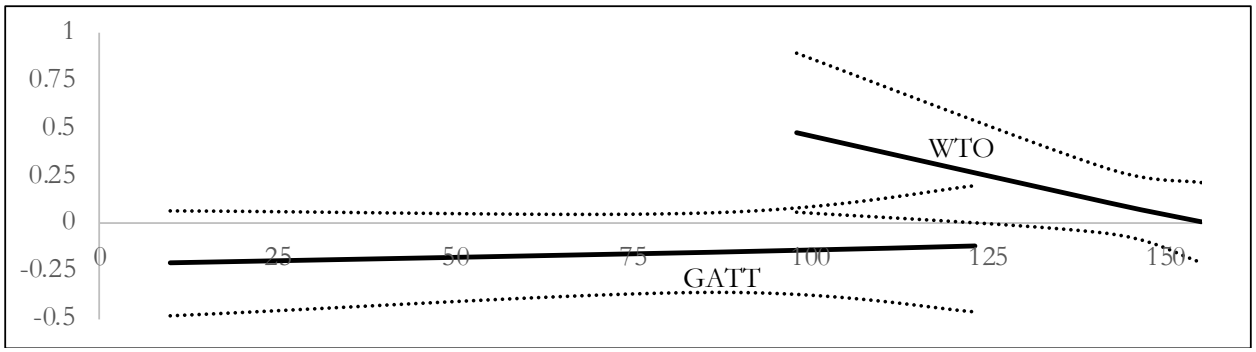
Model:	2.1	2.2	2.3	2.4	2.5	2.6	2.7
GATTWTO	-0.04 (0.07)						
GATT		-0.17 (0.10)	-0.12 (0.06)		-0.21 (0.16)	-0.02 (0.14)	
*GATT** *GATT Members*					0.001 (0.002)	-0.002 (0.002)	
WTO		0.09 (0.09)		0.05 (0.05)	2.15 (1.15)		2.96 (1.00)
*WTO** *WTO Members*					-0.014 (0.007)		-0.020 (0.007)
PTA	0.17 (0.04)	0.16 (0.04)	0.27 (0.03)	0.01 (0.05)	0.17 (0.04)	0.27 (0.03)	0.01 (0.05)
Constant	23.14 (0.06)	23.14 (0.06)	22.51 (0.05)	23.51 (0.05)	23.09 (0.06)	22.54 (0.06)	23.49 (0.05)
Directed Dyad Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer/Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter/Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99
N	1,039,979	1,039,979	569,173	468,997	1,039,979	569,173	468,997

Unit of analysis is the directed dyad/year.
PPML coefficients with robust standard errors clustered on the directed dyad, importer/year and exporter/year in parentheses.
GATT Members and *WTO Members* constitutive variables drop due to collinearity with their respective interaction variables.

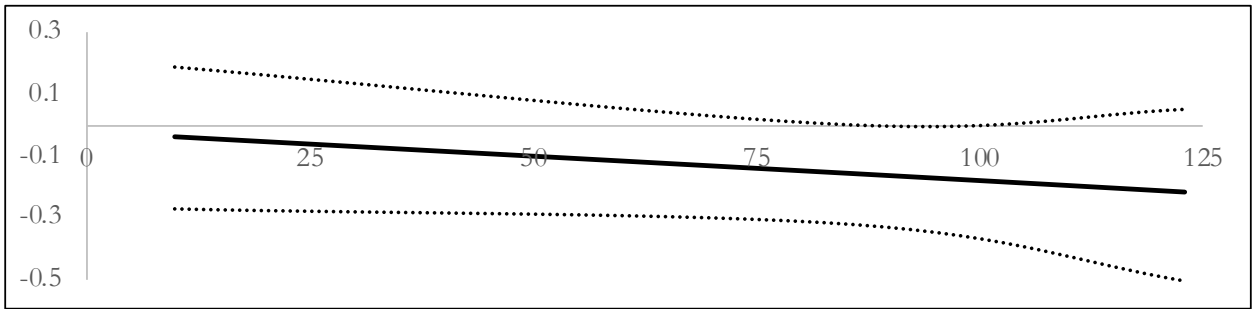
With split samples for each era, it is interesting to observe that *GATT* takes on a statistically significant negative coefficient in model 2.3, while *WTO* remains positive but statistically insignificant in model 2.4. We have no theory to explain why the GATT should have *decreased* bilateral trade between its member-states (as already observed by Esteve-Pérez, Gil-Pareja, and Llorca-Vivero 2020), but the fact that the WTO has been *on average* relatively ineffective at increasing bilateral trade could be explained by our proposed deepening/widening tradeoff: this generally deeper trade institution was more effective when it had fewer member-states and became less effective as more countries joined, leading to an average effect that is positive but insignificant.

Figure 3: The Marginal Effect of *GATT* and *WTO* Conditioned on Width: *_Members*.

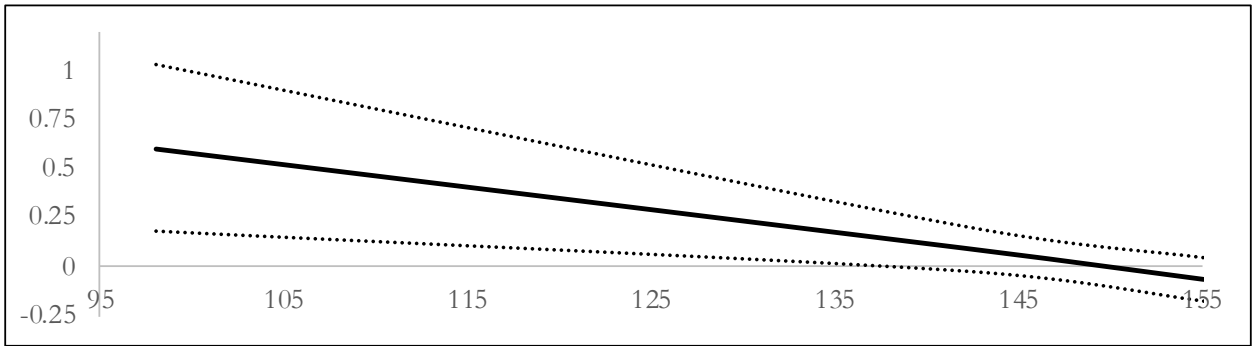
Model 2.5: GATT and WTO



Model 2.6: GATT Era



Model 2.7: WTO Era



The x-axis provides the marginal effect, and the y-axis provides the number of *_Members*. The solid line indicates the point estimate with the dashed lines indicating 95 percent confidence intervals.

We explore this possibility in model 2.5 using the full sample, model 2.6 for the GATT era, and model 2.7 for the WTO era. At least for the GATT, there is no evidence of a deepening/widening tradeoff, but also no contrary evidence when using the PPML specification. The GATT appears as ineffective with fewer member-states and also ineffective with more, although we note that the implausible result that the GATT significantly decreased trade between members (model 2.3) disappears in models 2.5 and 2.6 with the addition of the width interaction variable. This result is

illustrated in the first and second panels of Figure 2, showing the GATT to be ineffective (but not significantly associated with decreased trade) regardless of its width.

However, even when using the restrictive PPML specification, there is support for a deepening/widening tradeoff within the WTO. The interaction results, illustrated in both the first and third panels of Figure 3 show the *WTO* to be significantly associated with increased *Imports* when its width was less than 125 countries in the full sample (model 2.5) and less than 135 members in WTO era split sample (model 2.7), becoming trade *in*effective as more countries joined.

OLS with 2-Way Fixed Effects

If PPML with three-way fixed effects and three-way clustered standard errors represents a difficult specification for finding any depth for the multilateral trade institutions, then it is also useful to consider a more generous specification. Perhaps counterintuitively, when using a model associated with a stronger impact for the GATT and WTO, it may be harder to find evidence that this impact also attenuates as these institutions grew wider. We thus make use of the gravity dataset from Goldstein, Rivers, and Tomz (2007) with annual coverage from 1946 to 2004 (N=381,656).[15] Estimated log linear OLS, we follow their gravity specification with one change reported below.

First, Goldstein, Rivers, and Tomz operationalize the GATT and WTO to include both formal *and non-formal participants*. Second, their specification does not include importer/year and exporter/year fixed effects; it does, however, include both directed dyad and year fixed effects. With two-way fixed effects, more of the specific gravity controls (e.g., *One GATTWTO* and *Log Product Real GDP*) are retained in the specification. Third, we combine their separate measures for reciprocal and non-

[15] Their dataset drops the directed dyad/year observations associated with *Imports* less than $100,000 (in 1967 US dollars).

reciprocal PTAs into a single *PTA* dummy so that we can more easily compare the impact of these narrower regional arrangements with the wider GATT and WTO.

Using this second dataset, Table 3 presents the same basic sequence of models as shown earlier. Model 3.1 takes the specification that combines joint participation in the GATT with that in the WTO, and *Both GATTWTO* enters with a positive coefficient (unlike in model 2.1) that is more than four times larger than the same in model 1.1. It is also notable that the impact of the global trade regime is significantly larger than that for *PTA*, although this result is less due to the broader measure of GATT WTO *participation* (compared to formal membership) and more due to the use of year fixed effects, replacing the two sets of country/year fixed effects.[16] We next split *Both GATTWTO* into separate measures for the two institutions: *Both GATT* and *Both WTO*. These results show a larger positive coefficient for the GATT than for the WTO in both the full sample (model 3.2) and in the split samples (models 3.3 and 3.4). However, these new results (compared to those in Tables 1 and 2) again stem more from replacing the importer/year and exporter/year fixed effects with year dummies than with measuring the GATT and WTO based on participation rather than on formal membership.[17]

[16] As evidence on this point, if we re-estimate model 3.1 with three-way fixed effects, then the marginal effects of *Both GATTWTO* and *PTA* are, respectively, 0.07 (0.03) and 0.24 (0.02), consistent with the results in Table 1.

[17] For example, if we re-estimate model 3.2 with three-way fixed effects, then the marginal effect of *Both GATT* and *Both WTO* are, respectively, 0.06 (0.03) and 0.14 (0.06), also consistent with the results in Table 1.

Table 3 Models of *Imports* (logged) with Two-Way Fixed Effects.

	3.1	3.2	3.3	3.4	3.5	3.6	3.7
Both GATTWTO	0.38 (0.03						
Both GATT		0.40 (0.03)	0.38 (0.04)		1.41 (0.13)	1.56 (0.13)	
Both GATT GATT Members* (logged)					-0.25 (0.03)	-0.31 (0.03)	
Both WTO		0.18 (0.05)		0.13 (0.05)	0.60 (0.36)		1.76 (0.34)
Both WTO WTO Members* (logged)					-0.10 (0.08)		-0.34 (0.07)
One GATTWTO	0.21 (0.03)						
One GATT		0.21 (0.03)	0.22 (0.03)		0.17 (0.03)	0.15 (0.03)	
One WTO		0.08 (0.05)		0.00 (0.04)	0.04 (0.05)		-0.02 (0.04)
PTA	0.21 (0.02)	0.21 (0.02)	0.19 (0.02)	0.14 (0.03)	0.22 (0.02)	0.21 (0.02)	0.14 (0.03)
GSP	-0.17 (0.02)	-0.17 (0.02)	-0.15 (0.02)	0.07 (0.03)	-0.17 (0.02)	-0.14 (0.02)	0.06 (0.03)
Currency Union	0.54 (0.09)	0.53 (0.09)	0.64 (0.11)	-0.02 (0.03)	0.52 (0.09)	0.63 (0.11)	-0.01 (0.03)
Colonial Orbit	0.87 (0.08)	0.87 (0.08)	0.74 (0.08)	-0.05 (0.12)	0.80 (0.08)	0.66 (0.08)	-0.07 (0.12)
Log Product Real GDP	0.66 (0.01)	0.66 (0.01)	0.65 (0.01)	0.39 (0.02)	0.66 (0.01)	0.65 (0.01)	0.39 (0.02)
R^2	0.83	0.83	0.83	0.94	0.83	0.83	0.94
N	381,656	381,656	284,744	96,912	381,656	284,744	96,912

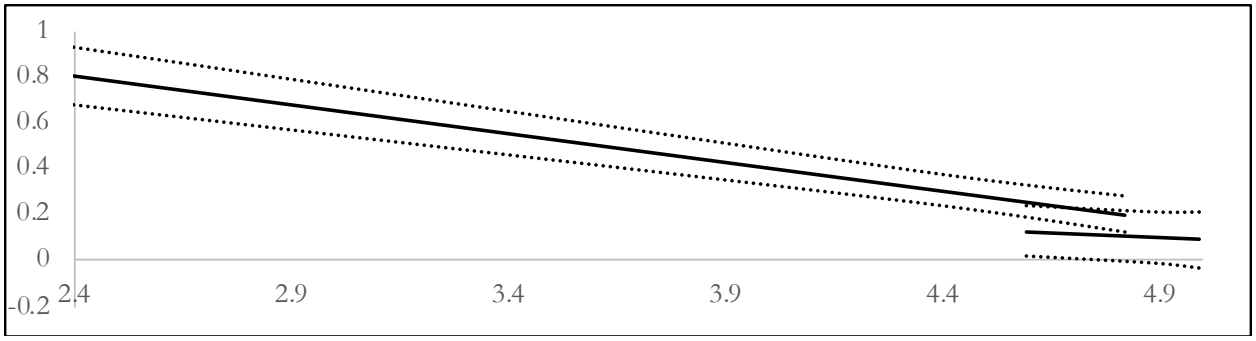
Unit of analysis is the directed dyad/year.
OLS coefficients with robust standard errors clustered on the dyad in parentheses.
Directed dyad and year fixed effects included.
GATT Members (logged) and *WTO Members* (logged) constitutive variables drop due to collinearity with their respective interaction variables.

Thus, the question becomes with a different dataset and a different specification producing stronger results especially for the GATT, can one still observe evidence that their positive association with bilateral trade declines as they take on more member-states? To answer this question, we add the width interaction variables in the full sample (model 3.5) and then in the split samples (models 3.6 and 3.7), and the results continue to accord with our hypothesis, especially for the GATT. As

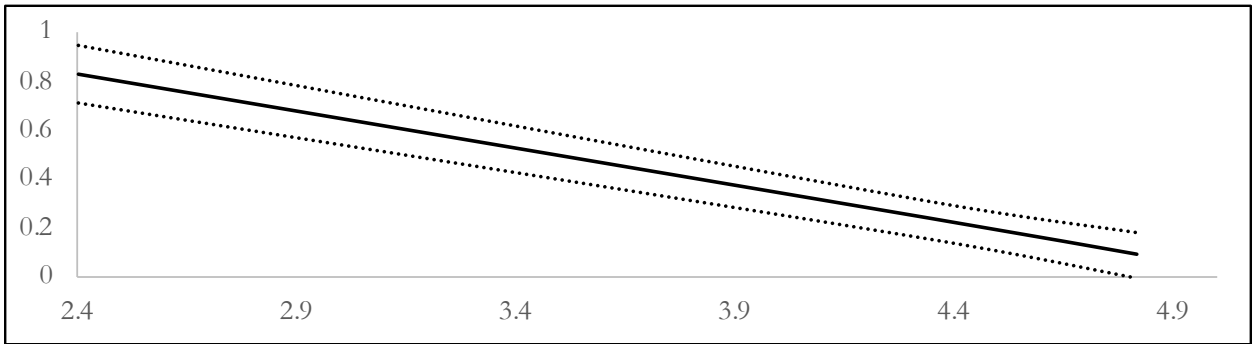
illustrated in Figure 4, the GATT was highly effective with the fewest number of member-states, producing a marginal effect of 0.81 (0.06) with 2.4 members (logged). And using this specification with two-way fixed effects, the GATT never lost its effectiveness, although it declined significantly with more member-states. The first and third panels of Figure 4 show that WTO was also trade effective when it began in 1995 with its narrowest membership at 4.59 (logged) countries. But unlike the GATT, the WTO did lose its trade effectiveness with greater width, becoming statistically insignificant as the number of members (logged) reached 4.9.

Figure 4: The Marginal Effect of *GATT* and *WTO* Conditioned on Width: _ *Members* (logged).

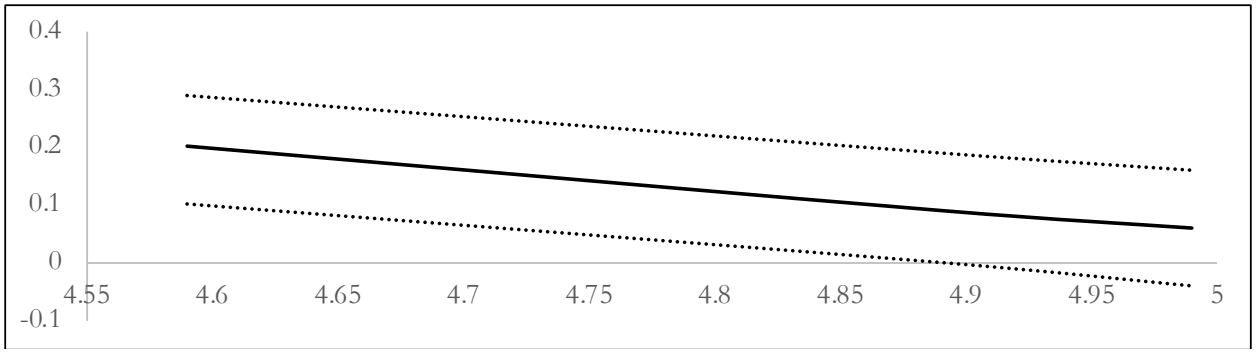
Model 3.5: GATT and WTO



Model 3.6: GATT Era



Model 3.7: WTO Era



The x-axis provides the marginal effect, and the y-axis provides the number of *_Members* (logged). The solid line indicates the point estimate with the dashed lines indicating 95 percent confidence intervals.

3. Discussion

Using multiple samples from different datasets, a variety of statistical specifications (e.g., two-way and three-way fixed effects), different estimators (OLS and PPML), and alternative measures of GATT WTO membership or participation, this paper has provided robust evidence to support our hypothesis that there should have been a deepening versus widening tradeoff within these multilateral trade institutions based on their consensus practice in bargaining and enforcement. In addition to providing some of the first quantitative evidence consistent with a depth/width tradeoff within international organizations, these results also bear on some perhaps surprising findings in the literature on trade institutions.

Rose (2004), in his set of empirical exercises to show that the GATT/WTO has not had a significant effect in increasing trade, allowed its effect to vary over time and by negotiation round (ibid, 105-6, Tables 2 and 3), finding declining effects by year and by negotiation round. But he offered no particular explanation for why this might be the case, other than to point out that these results were generally consistent with the weak average impact of these multilateral trade regimes. Likewise, in their response paper, Tomz, Goldstein and Rivers (2007, 2015, Table 3) performed the latter

exercise, also finding declining effects by negotiation round but emphasizing that the effects were positive and statistically significant in most rounds.

However, why should the trade impact of these multilateral institutions fall with successive negotiation rounds *when the purpose of each round was to increase de jure depth*, or to expand their coverage in terms of products and forms of trade protection? Indeed, if de jure depth simply translates into de facto depth, then we should expect to see these multilateral institutions becoming more trade effective with each negotiation round. But we observed just the opposite, and the argument/evidence presented in this paper can explain why: over time and concurrent with each negotiation round, the GATT/WTO also became wider with a negative impact on its de facto depth even while its de jure depth increased (Davis 2004).

This understanding also provides an alternative explanation for the results in Mansfield (1998) and Mansfield and Reinhardt (2003), showing that states are more likely to form a PTA during a GATT negotiation round based on the logic that doing so might "enhance their bargaining leverage in these negotiations" (Mansfield 1998, 535-6). We do not dispute that this logic may be true, but their result can also be explained by the unmet demand for freer trade during and even after the conclusion of a negotiation round. To the extent that many states demand deeper trade that the multilateral/global regime could not meet due its growing width with a consensus practice, then these states might be expected to satisfy their demand for depth through narrower PTAs, which our results generally show to be deeper on a de facto basis compared to the GATT/WTO. Thus, rather than identifying the proliferation of PTAs as a cause for the GATT/WTO's weakness (e.g., Bhagwati 2008), our argument/evidence identifies PTA proliferation *as the result* of the GATT/WTO's weakness, or de facto shallowness.

Finally, our evidence of a deepening/widening tradeoff within the GATT and WTO has implications beyond international trade. We argued that this tradeoff stemmed in large part from their

consensus practice, and these multilateral trade institutions are not the only international regimes that operate with such. As described by Steinberg (2002, 340-1), various regimes, including "the Association of Southeast Asian Nations (ASEAN), Conference on Security and Cooperation in Europe (CSCE), the Executive Committee of the International Monetary Fund (IMF), the GATT/WTO, Common Market of the South, Mercado Comun del Sur (MERCOSUR), North Atlantic Treaty Organization (NATO), Organization for Economic Cooperation and Development (OECD), and many specialized agencies of the United Nations (UN), including the UN Development Program (UNDP) and the Executive Committee of the UN High Commission on Refugees (UNHCR), usually have taken decisions only with the consensus or unanimous support of member states."

On this basis, one might also expect to observe a de facto deepening/widening tradeoff within these international regimes. Furthermore, the research design used here offers a way to consider this possibility. While there may not be a "gravity model" equivalent in most issue-areas, any model that seeks to test how much an international agreement or institution has influenced the prescribed behavior of its member-states is effectively estimating its *de facto depth*. Thus, the addition of a width independent variable interacted with a regime membership variable can tell us whether the latter shows a lesser (greater) effect on state behavior as the number of member-states grows (declines). On the basis of the evidence presented here, we posit that a similar tradeoff may exist in other international regimes covering different issue-areas.

References

Allee, Todd L., and Jamie E. Scalera. 2012. "The divergent effects of joining international organizations: Trade gains and the rigors of WTO accession." *International Organization* (66 (2): 243-276.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with gravitas: A solution to the border puzzle." *American Economic Review* 93 (1): 170-192.

Bearce, David H., Cody D. Eldredge, and Brandy J. Jolliff. 2016. "Does institutional design matter? A study of trade effectiveness and PTA flexibility/rigidity." *International Studies Quarterly* 60 (2): 307-316.

Bearce, David H., Cody D. Eldredge, and Brandy J. Jolliff. 2015. "Do finite duration provisions reduce international bargaining delay?" *International Organization* 69 (1): 219-239.

Bernauer, Thomas, Anna Kalbhenn, Vally Koubi, and Gabriele Spilker. 2013. "Is there a 'Depth versus Participation' dilemma in international cooperation?" *Review of International Organizations* 8 (4): 477-497.

Bhagwati, Jagdish. 2008. *Termites in the trading system: How preferential agreements undermine free trade*. Oxford University Press.

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): 158-72.

Davis, Christina L. 2004. "International institutions and issue linkage: Building support for agricultural trade liberalization." *American Political Science Review* 98 (1): 153-169.

Davis, Christina L., and Meredith Wilf. 2017. "Joining the Club: Accession to the GATT/WTO." *Journal of Politics* 79 (3): 964-978.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1998. "Managing the evolution of multilateralism." *International Organization* 52 (2): 397-419.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. "Is the good news about compliance good news about cooperation?" *International Organization* 50 (3): 379-406.

Dür, Andreas, Leonardo Baccini, and Manfred Elsig. 2014. "The design of international trade agreements: Introducing a new dataset." *Review of International Organizations* 9 (3): 353-375.

Eicher, Theo S., and Christian Henn. 2011. "In search of WTO trade effects: Preferential trade agreements promote trade strongly, but unevenly." *Journal of International Economics* 83 (2): 137-153.

Esteve-Pérez, Silviano, Salvador Gil-Pareja, and Rafael Llorca-Vivero. 2020. "Does the GATT/WTO promote trade? After all, Rose was right." *Review of World Economics* (156): 377-405.

Footer, Mary E. 1996/1997. "Role of Consensus in GATT/WTO Decision-making." *Northwestern Journal of International Law and Business* 17: 653-680.

Gilligan, Michael J. 2004. "Is there a broader-deeper tradeoff in multilateral international agreements?" *International Organization* 58 (3): 459-84.

Goldstein, Judith L., Douglas Rivers, and Michael Tomz. 2007. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* 61 (1): 37-67.

Graham, Thomas R. 1983. "Global Trade: War & Peace." *Foreign Policy* 50 (Spring): 124-137.

Hartman, Stephen W. 2013. "The WTO, the Doha Round Impasse, PTAs, and FTAs/RTAs." *International Trade Journal*, 27 (5): 411-430

Jackson, John H. 1978. "The Crumbling Institutions of the Liberal Trade System." *Journal of World Trade* 12 (2): 93-106.

Jones, Kent. 2009. "The political economy of WTO accession: the unfinished business of universal membership." *World Trade Review* 8 (2): 279-314.

Kalla, Patricia. 1986. "The GATT Dispute Settlement Procedure in the 1980s: Where Do We Go from Here?" *Dickinson Journal of International Law* 5 (1): 82-101.

Kono, Daniel Y. 2006. "Optimal obfuscation: Democracy and trade policy transparency." *American Political Science Review* 100 (3): 369-384.

Koremenos, Barbara. 2005. "Contracting around international uncertainty." *American Political Science Review* 99 (4): 549-565.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. "The rational design of international institutions." *International Organization* 55 (4): 761-799.

Larch, Mario, Joschka Wanner, Yoto V. Yotov, and Thomas Zylkin. 2019. "Currency unions and trade: A PPML re-assessment with high-dimensional fixed effects." *Oxford Bulletin of Economics and Statistics* 81 (3): 487-510.

Lowenfeld, Andreas F. 1994. "Remedies along with Rights: Institutional Reform in the new GATT." *American Journal of International Law* 88 (3): 477-488.

Mansfield, Edward D. and Eric Reinhardt. 2003. "Multilateral Determinants of Regionalism: The Effects of GATT/WTO on the Formation of Preferential Trading Arrangements." *International Organization* 57 (4): 829-862.

Mansfield, Edward D. 1998. "The proliferation of preferential trading arrangements." *Journal of Conflict Resolution* 42 (5): 523-543.

Martin, Lisa L. 1994. *Coercive cooperation: Explaining multilateral economic sanctions*. Princeton University Press.

Martin, Will, and Patrick Messerlin. 2007. "Why is it so difficult? Trade liberalization under the Doha Agenda." *Oxford Review of Economic Policy* 23 (3): 347-366.

Martin, Will, and L. Alan Winters. 1995. *The Uruguay Round: Widening and Deepening the World Trading System*. Washington, DC: The World Bank.

Milner, Helen. 1992. "International theories of cooperation among nations: Strengths and weaknesses." *World Politics* 44 (3): 466-496.

Oye, Kenneth A., ed. 1986. *Cooperation under anarchy*. Princeton University Press.

Pescatore, Pierre. 1993. "The GATT Dispute Settlement Mechanism: Its Present Situation and Its Prospects." *Journal of World Trade* 27 (1): 5-20.

Rose, Andrew K. 2004. "Do we really know that the WTO increases trade?" *American Economic Review* 94 (1): 98-114.

Rose, Andrew K. 2002. "Do we really know that the WTO increases trade?" National Bureau of Economic Research (NBER) Working Paper 9273 (October).

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.

Slapin, Jonathan B., and Julia Gray. 2014. "Depth, ambition and width in regional economic organizations." *Journal of European Public Policy* 21 (5): 730-745.

Steinberg, Richard H. 2002. "In the shadow of law or power? Consensus-based bargaining and outcomes in the GATT/WTO." *International Organization* 56 (2): 339-374.

Stiles, Kendall. 1996. "Negotiating Institutional Reform: The Uruguay Round, the GATT, and the WTO." *Global Governance* 2 (1): 119-148.

Stone, Randall W., Branislav L. Slantchev, and Tamar R. London. 2008. "Choosing how to cooperate: A repeated public-goods model of international relations." *International Studies Quarterly* 52 (2): 335-362.

Subramanian, Arvind, and Shang-Jin Wei. 2007. "The WTO promotes trade, strongly but unevenly." *Journal of International Economics* 72 (1): 151-175.

Tomz, Michael, Judith L. Goldstein, and Douglas Rivers. 2007. "Do we really know that the WTO increases trade? Comment." *American Economic Review* 97 (5): 2005-2018.

Von Borzyskowski, Inken, and Felicity Vabulas. 2019. "Hello, goodbye: When do states withdraw from international organizations?" *Review of International Organizations* 14 (2): 335-366.