International Studies Quarterly

International Studies Quarterly (2015) **59**, 765–775

The Politics of Trade Agreement Design: Revisiting the Depth–Flexibility Nexus

RESEARCH NOTE

Leonardo Baccini
LSE

Andreas Dür
University of Salzburg

and

Manfred Elsig
University of Bern

Existing research indicates the interrelated nature of different dimensions of the design of international institutions. In particular, it shows the greater flexibility of deep agreements. We argue—and demonstrate empirically—that the positive relationship between depth and flexibility holds for preferential trade agreements (PTAs). But we add two qualifications to the conventional wisdom that depth and flexibility go hand in hand. First, we argue that the positive relationship between depth and flexibility proves weaker for democracies than for nondemocracies. Second, when making deep agreements more flexible, countries also add strings to the use of the additional flexibility provisions. An original data set on the design of 587 PTAs allows us to test our arguments. Both descriptive evidence and multivariate statistics support the theoretical expectations. The findings contribute to the literatures on the design of international institutions and the causes and consequences of PTAs.

Existing research on the design of international institutions shows that deep agreements tend to be more flexible than shallow ones are (Downs, Rocke, and Barsoom 1996; Rosendorff and Milner 2001; Kucik and Reinhardt 2008).[1] We argue that this pattern also holds for preferential trade agreements (PTAs). Deep PTAs with many commitments—such as liberalizing services trade while protecting foreign direct investments and also intellectual property rights—should also feature multiple flexibility measures that allow states to temporarily withdraw concessions. We also argue, however, for a more nuanced relationship between depth and flexibility than existing scholarship recognizes. In particular, the positive relationship between depth and flexibility should be less pronounced for democracies than it is for nondemocracies. Moreover, we expect countries to attach strings to the use of the flexibility provisions that they add to deep agreements.

We test these expectations on a new data set on the design of 587 PTAs signed between 1945 and 2009 (Dür, Baccini, and Elsig 2014). Our data set is very comprehensive in terms of both agreements covered and detail of the coding. It contains data on a substantially larger number of PTAs than existing data sets. Moreover, it includes a large number of items that allow us to measure the agreements' depth and flexibility, and the extent of restrictions on the use of flexibility tools. Bivariate and multivariate analyses support our theoretical expectations. The findings prove highly robust to changes in estimation method and the operationalization of variables.

We thus report the first large-n study of the relationship between depth and flexibility with respect to PTAs. In focusing on PTAs, our study nicely complements existing literature on the design of international economic institutions. This literature focuses mainly on the World Trade Organization (WTO) and its forerunner, the General Agreement on Tariffs and Trade (GATT) (Rosendorff and Milner 2001; Rosendorff 2005; Kucik and Reinhardt 2008; Pelc 2009). In the GATT/WTO system,

Leonardo Baccini is assistant professor in International Political Economy at the London School of Economics and Political Science. He is the author of *Cutting the Gordian Knot of Economic Reform: How International Institutions Promote Liberalization* (2015). Further information about his research is available at https://sites.google.com/site/leonardobaccini/.

Andreas Dür is professor of International Politics at the University of Salzburg. He is author of *Protection for Exporters: Power and Discrimination in Transatlantic Trade Relations, 1930–2010* (2010) and co-editor of *Trade Cooperation: The Purpose, Design and Effects of Preferential Trade Agreements* (2015).

Manfred Elsig is associate professor of International Relations at the University of Bern and Director of the Swiss Science Foundation's NCCR Trade Regulation (www.nccr-trade.org). He is the co-editor of *Trade Cooperation: The Purpose, Design and Effects of Preferential Trade Agreements* (2015).

[1] We are grateful to Tobias Böhmelt, Yoram Haftel, Raymond Hicks, Yuch Kono, Yotam Margalit, Helen Milner, Peter Rosendorff, Johannes Urpelainen, the participants at the 2011 IPES conference, 2012 PEIO conference, 2012 and 2013 ECPR Joint Sessions, and a seminar at Columbia University, the anonymous referees, and the editors of *International Studies Quarterly* for helpful comments on earlier versions of this article. We acknowledge support from the Swiss National Science Foundation's NCCR Trade Regulation.

Baccini, Leonardo, Andreas Dür, and Manfred Elsig. (2015) The Politics of Trade Agreement Design: Revisiting the Depth–Flexibility Nexus. *International Studies Quarterly*, doi: 10.1111/isqu.12188

negotiations take place against the background of a more or less fixed institutional structure, which in most cases leaves only a few elements up for discussion. In negotiating a PTA, by contrast, countries make many choices regarding institutional design at the same time. Moreover, accumulating evidence suggests that variation in PTA design matters for key economic outcomes such as trade flows and foreign direct investments.[2] These considerations point to the importance of understanding the relationship between depth and flexibility in PTAs.

This article also contributes to the broader literature on the design of international institutions (see Koremenos, Lipson, and Snidal 2001). In particular, our study provides good reasons to believe that the depth–flexibility relationship is more nuanced than currently recognized also for international institutions other than PTAs. Both our argument about regime type and the concept of flexibility strings should apply to other institutional settings. We expand on these implications in the conclusion.

The Relationship between Depth and Flexibility in PTAs

International agreements vary in terms of both depth of cooperation and their flexibility. "Depth" refers to the extent to which an agreement constrains state behavior.[3] In the particular case of trade agreements, deep agreements liberalize trade more than hollow agreements. An important determinant of the depth of a trade agreement is the average tariff cuts it requires from states. Trade agreements, however, can also contribute to liberalizing trade relations between states by opening some service sectors to foreign competition or by allowing foreign companies to bid for tenders for government procurement contracts. Similarly, the protection of foreign direct investment can substantially enhance market access for exporters, as foreign investments can be a precondition for trade. A trade agreement can also go beyond protecting investments against arbitrary decisions and explicitly liberalize foreign investments (World Trade Organization 2011:137–140).

Other behind-the-border obstacles that a deep trade agreement may remove include burdensome technical standards, discriminatory food safety and animal and plant health measures, inadequate protection of intellectual property rights, and competition rules that discriminate against foreign traders. Some PTAs, for example, foresee the mutual recognition of international product standards or the harmonization of technical regulations (Piermartini and Budetta 2009). This reduces transaction costs and increases market integration. The strengthening of intellectual property rights also enables exporters to pursue a long-term market penetration strategy. Absence of patent or trademark protection directly translates into less exports and less technology transfer in the case of investments directed at foreign markets (Maskus and Penubarti 1995). Finally, cooperation on competition policy addresses unfair business behavior by state enterprises and private firms. Competition provisions thus facilitate exports and foreign investments. The depth of a trade agreement, therefore, is a function of tariff cuts and provisions concerning services, government procurement, investments, standards, intellectual property rights, and competition.

Flexibility provisions are devices included in an agreement that allow states to anticipate and respond to domestic contingencies or to adjust their policies for other purposes without violating the terms of an agreement. They provide for legally accepted opt-outs without leading to a *de jure* breach of an agreement and encompass exit options, duration and renegotiation clauses, reservations, escape clauses, and withdrawal clauses (see, for example, Koremenos et al. 2001; Rosendorff and Milner 2001; Helfer 2005, 2013; Koremenos 2005; Neumayer 2007).

In trade agreements, we often find flexibility instruments such as special and general safeguard provisions, antidumping clauses, balance of payments (BoP) exceptions, and tariff overhang (namely a gap between applied and bound tariffs). The Treaty of Rome (1957), for example, allowed member states of the European Economic Community to suspend part of their liberalization in the case of BoP difficulties (Art. 109). Similarly, the EU–South Korea PTA contains a special safeguard related to cars (Elsig and Dupont 2012). In addition, safety valves exist to address anticompetitive behavior by other states (excessive use of subsidies to boost exports) or by firms (dumping practices). These exist in the form of antidumping and countervailing duty provisions.

While depth and flexibility capture two distinct dimensions of institutional design, several studies have argued and/or shown that depth and flexibility are positively related (Rosendorff and Milner 2001; Kucik and Reinhardt 2008; Johns 2014). Two different causal pathways may explain this relationship. For one, flexibility may facilitate the negotiation of deeper agreements (Kucik and Reinhardt 2008). This argument assumes that states value the long-term benefits of cooperation, but also face short-term domestic pressure to violate an agreement, which varies over time. When the pressure is strong, governments have an incentive to violate agreements that do not contain flexibility provisions (Downs et al. 1996). Violation, in turn, may prompt retaliation, causing a breakdown of cooperation. States, therefore, only negotiate deep agreements if they can include flexibility in the agreement that ensures the long-term viability of cooperation.

Alternatively, deep agreements may lead to greater domestic demand for flexibility. This argument starts with the insight that international agreements have distributional effects. In the case of trade, agreements benefit exporters that gain from better foreign market access and multinational companies that (plan to) invest in partner countries. At the same time, they hurt import-competing interests that face greater import-competition. Importantly, a trade agreement's distributional effects depend on its depth. Recent research shows that deeper agreements have a greater positive trade flow effect than shallow ones (Baier and Bergstrand 2007; Dür et al. 2014; Egger and Nigai 2015). They also encourage more foreign direct investments (Büthe and Milner 2014).

Given these distributional effects, deep agreements can be expected to receive support from exporters and multinational companies. At the same time, plans for a deep agreement should lead to particularly intensive lobbying from import-competing groups that demand protection. The reason is that the number of economic sectors that potentially experience negative effects from a PTA

[2] For the effect of trade agreement design on trade flows, see Dür et al. (2014) and Egger and Nigai (2015), and for foreign direct investments, see Büthe and Milner (2014).

[3] This is similar to Downs et al. (1996:383) who define depth as "the extent to which [an agreement] requires states to depart from what they would have done in its absence."

increases together with the depth of an agreement. Service providers, for example, do not have to fear foreign competition if an agreement only liberalizes trade in goods. Similarly, sectors protected via discriminatory technical standards do not need to bother about tariff cuts, unless the trade agreement also includes a provision that harmonizes technical standards. If import-competitors cannot block an agreement, they will push for a minimum objective, namely ensuring that the agreement is highly flexible. For them, flexibility provisions soften the impact of an agreement and serve as a form of protection (Grossman and Helpman 1994). In the face of strong lobbying by import-competitors, governments can be expected to make deep agreements flexible.

Both causal arguments thus lead to the expectation that depth and flexibility are positively related. So far, most studies test this relationship between depth and flexibility for the case of trade agreements negotiated in the framework of the GATT/WTO. Kucik and Reinhardt (2008), for example, show that countries that have an antidumping instrument agree to lower tariff bindings upon entry into the GATT/WTO system (and sustain them as members) than other countries. In PTAs, we expect the causal mechanism to work in the other direction. PTA negotiators tend to resolve the question of depth first and only then decide on the amount of flexibility included in the agreement. In fact, as soon as countries signal their intention to negotiate a PTA, they specify how ambitious the future agreement should be, long before they negotiate specific flexibility provisions. The causal link thus most probably goes from depth to flexibility. As in the empirical analysis below we do not test for the causal direction, we remain agnostic about the direction and simply posit a positive relationship between depth and flexibility:

Hypothesis 1: *The deeper a PTA, the more flexible it is.*

For two reasons, we expect that the conventional wisdom that motivates Hypothesis 1 needs to be qualified. First, regime type conditions the relationship between depth and flexibility, because democracies rely more strongly on nontariff barriers and trade remedies for protection than nondemocracies (Kono 2006:374). In democracies, voters punish politicians that impose high tariffs, which is a transparent tax on consumers. So that they still can satisfy interest groups that demand protection, democratic governments rely on devices about which voters have little information, namely nontariff barriers and trade remedies.

For the design of PTAs, this "optimal obfuscation" argument implies that democracies will include flexibility provisions even in shallow agreements, as they need them to respond to protectionist demands in a manner that avoids punishment by voters. In the absence of flexibility, they would be in breach of the agreement when, for example, imposing antidumping or countervailing duties. Nondemocracies, by contrast, see little advantage from flexibility in shallow agreements. By definition, a shallow agreement does not hinder them selectively to use trade barriers to placate import-competitors, and they do not need flexibility provisions for that purpose as they have no incentive to disguise their reliance on protectionism for reasons related to voters' concerns.

As depth increases, we expect to see only a small increase in flexibility for democracies, which already include much flexibility in shallow agreements. The causal argument presented above about the relationship between depth and flexibility thus should mainly apply to nondemocracies. They see their ability to respond to protectionist demands via traditional trade barriers restricted in deep agreements and then add flexibility provisions to deep agreements that satisfy import-competing interests. Table 1 summarizes this argument. Importantly, we only expect regime type to condition the relationship between depth and flexibility for classical escape instruments. For flexibility measures related to tariffs (such as tariff cut phase-ins, as discussed below), we do not expect such a conditional relationship. Even if democracies rely less on tariffs for protectionist purposes, they have little to gain from a very rapid reduction of existing tariff rates. We thus hypothesize:

Hypothesis 2: *The positive relationship between depth and flexibility is weaker for democracies than for nondemocracies.*

Second, when governments add more flexibility to deep agreements, they likely attach strings to the use of these additional flexibility provisions. The causal logic underlying this qualification starts with the insight that flexibility poses risks for both governments and exporters alike. Countries may use flexibility to ease adjustment costs and thus reduce temporarily high costs of compliance. As this ensures the long-term viability of cooperation, this is the intended use of flexibility. Some countries, however, may abuse flexibility provisions that offer them some discretion to impose temporary trade barriers with the aim of giving rents to particular domestic constituencies. Overuse of the opt-outs, in turn, jeopardizes the overall benefits of the agreement (and its stability) in the long run. It may even nullify the benefits of market access. Exporters that value market access thus oppose lax rules on flexibility which would allow foreign governments or importers to demand too much protection (Kucik 2012:98).

With the aim of avoiding abuse of flexibility provisions with a discretionary element, governments then have a strong incentive ex-ante to define limits to the use of flexibility instruments. They can be expected to agree on procedural constraints to control the application of escape tools as a sort of hand-tying approach. Various possibilities exist for restricting the use of flexibility measures. In terms of safeguards and antidumping, states may limit the duration of an antidumping duty, restrict the upper level of the trade remedy imposed, or make reference to the GATT/WTO legal framework that prescribes a number of procedural and substantive obligations that shield against abuse. In the case of the South Korea–US negotiations, for example, the Korean car industry lobbied hard for changes in US trade remedy laws that would make the imposition of antidumping duties more difficult. As regards subsidies, treaties may push governments to decrease or eliminate them or demand active cooperation among authorities to address competition-related negative spillovers. For instance, the trade agreement between Japan and Switzerland foresees that while parties to the agreement can maintain high levels of domestic support, no export subsidies shall be introduced or maintained on agricultural products for which they agreed on tariff liberalization.[4] We call these measures "flexibility strings."

[4] In the WTO, export subsidies in agricultural products are generally still allowed under specific conditions. The abolition of these subsidies forms part of the single negotiation package in the context of the current negotiations in the framework of the WTO.

Table 1. Regime Type, Depth, and Flexibility

	Shallow	*Deep*
Democracies	Need flexibility to respond to protectionist demands	Need flexibility to respond to protectionist demands
Nondemocracies	Less need for flexibility; protectionist demands are satisfied with other trade barriers	Need for flexibility, as ability to use other measures constrained

The second qualification of the conventional wisdom that we derive from this reasoning is that countries will attach strings to the use of flexibility provisions. Governments will seek an optimal degree of flexibility that allows for temporary breach and adaptation if necessary, but restrict this flexibility through a set of tools. We thus formulate the following hypothesis:

Hypothesis 3: *PTAs that are more flexible contain more flexibility strings.*

Before empirically examining our hypotheses, we address two potential objections to our argument. On the one hand, some evidence suggests that governments do not start from scratch in designing PTAs, but rely on templates. Even if they do so, however, this should not affect our analysis, as the PTA template likely includes all of the aspects of the design of a PTA that we capture in our hypotheses: depth, flexibility, and flexibility strings. To the extent that copying takes place, we expect governments to copy provisions from all three areas, meaning that the postulated relationships should remain intact.

On the other hand, our discussion so far has ignored the fact that two or more countries contribute to the design of a PTA. In a highly asymmetric relationship, the larger country (for example, the United States) may not bother about including flexibility provisions into a formally deep agreement, if the smaller country (for example, Oman) does not pose a threat for its import-competing sectors. The smaller country, in turn, may not have the power to insist on the inclusion of the flexibility provisions or to resist the imposition of flexibility strings. In such a case, the postulated relationships between depth, flexibility, and flexibility strings would disappear. We expect, however, that import-competitors in a large country worry that the agreement could set a precedent, and thus defend the inclusion of flexibility provisions even in agreements with smaller trading partners. Although we think that power asymmetry should not affect our results, below we carry out an empirical test that directly tackles this issue.

Operationalization and Descriptive Evidence

For the empirical analysis of the relationship between depth, flexibility, and flexibility strings, we rely on an original data set on the design of 587 PTAs signed between 1945 and 2009. The list of agreements contains 358 bilateral and 229 plurilateral agreements and substantially goes beyond the list of agreements registered with the WTO (60% of them feature in the WTO list). It results from the integration of several existing lists (including those maintained by the Organization of American States' Foreign Trade Information System and the World Bank) and the systematic search of government web pages. The average agreement in the data set is relatively young (no fewer than 73% of the agreements have been signed since 1990), was signed by developing countries (67% are of a South–South type), and is a full free trade agreement (60% of the data set).

We coded the agreements manually for a total of eight market-access-related sectors of cooperation that may be included in PTAs, encompassing goods, services, investments, intellectual property rights, competition, public procurement, standards, and trade remedies. For each of these sectors, we coded a significant number of items, meaning that we have about 100 data points for each agreement. To ensure the reliability of the data, we double-coded all agreements and then resolved any differences between the two sets of data.[5]

For nearly all variables, inter-rater agreement as measured by Cohen's kappa exceeds 0.75. Moreover, cross-checks of the data set against existing ones that partially overlap with it confirm the reliability of the data.[6] The large number of agreements coded and the level of detail included in the coding mean that this data set offers a unique opportunity to test our argument. In the following, we discuss how we use this data set to measure depth, flexibility, and flexibility strings.

We use two different measures of depth below. On the one hand, we created an additive index of 48 items in our data set (*Depth* (*index*)). We only included items in the index that theoretically relate to depth and weighted each item equally (the supplementary file available on ScholarOne contains a list of all of these items). While theoretically this measure can range from 0 to 48, in practice, we observe values in the range from 0 to 40. On the other hand, we relied on item response analysis on the same set of variables to arrive at a measure of depth (*Depth* (*IR*)). Latent trait analysis is a technique that resembles factor analysis, with the advantage that it is applicable to binary data (Bartholomew, Knott, and Moustaki 2011). The specific model that we apply, which is known as the Rasch model, assumes that all items capture one underlying latent dimension. The items, however, contribute more or less to this latent dimension (that is, they have more or less discriminatory power). Using this operationalization, relatively rare provisions contribute more to depth than ubiquitous provisions. After rescaling to remove negative values, this variable ranges from 0 to 3.3.

The two measures are highly correlated ($r = 0.94$). For both depth measures, we find that over time agreements have become deeper, with the trend toward deep agreements starting in the second half of the 1980s. We also observe that North–South agreements are substan-

[5] For more details, see Dür et al. (2014).

[6] For these other data sets, see the contributions in Estevadeordal, Suominen, and Teh (2009), World Trade Organization (2011) and Kucik (2012). Estevadeordal et al. (2009) include up to 70 agreements that also feature in the present data set (varying depending on chapter). World Trade Organization (2011) analyzes the contents of 96 agreements. Kucik (2012) coded 323 of the agreements contained in our data set.

tially deeper than both North–North and South–South agreements in our data set. North–North agreements are only slightly deeper than South–South agreements. The United States (for example, Australia–US and Colombia–US), Japan (for example, Japan–Switzerland), and the European Free Trade Association (for example, European Free Trade Association–Colombia) signed the deepest agreements in the data set. On both measures of depth, 147 agreements score 0, including the African Common Market and a large number of Arab agreements. As these examples suggest, in contrast to Slapin and Gray (2014), in our data set, depth and the number of member states are not positively related.

We also rely on two measures of flexibility. On the one hand, we use a simple additive index of the presence or absence of four provisions in PTAs: a provision allowing for the suspension of tariff cuts in the case of balance of payments problems, a general safeguard provision, a provision allowing for the imposition of countervailing duties, and a provision allowing for the imposition of antidumping duties (*Escape flexibility*).[7] In the absence of these provisions, a country that suspends its tariff cuts or imposes antidumping and countervailing duties for goods covered by the agreement breaches the agreement.[8] These four provisions thus serve as escape clauses. The index can range from 0 to 4. In our data set, 73 agreements score 0 and 260 agreements score 4 on this variable. Over time, we see an increasing number of agreements including several of these escape provisions.

On the other hand, PTAs may also introduce flexibility by allowing member states to postpone tariff reductions (Chase 2003:160–165). The more time states have to achieve the agreed tariff cuts, the more flexibility exists for import-competing groups to adjust to increased competition. We thus use the maximum (across all tariff categories) number of years that countries have to achieve the liberalization of tariffs envisaged in the agreement as a second measure of flexibility (*Transitional flexibility*). Phase-out periods for tariff liberalization range between 0 years (members liberalize all tariffs at the date of entry into force of an agreement) and 25 years. We took the median for the member countries of a PTA if the transition periods vary across countries.[9] As this type of flexibility does not contain a discretionary element, it does not require flexibility strings.

Following Hypothesis 1, flexibility should be positively related to depth. To examine this proposition, we show boxplots for *Escape flexibility* based on a recoded depth variable that ranges from 0 (very shallow) to 4 (very deep). The resulting graph (Figure 1) shows a strong positive relationship between the two variables. It also suggests that the relationship may not be linear, an issue that we discuss below. In Figure 2, we show that the positive relationship between depth and flexibility also holds for our second measure of flexibility (*Transitional flexibility*). Again, we see a positive relationship between the two variables. As depth increases, transitional flexibility also

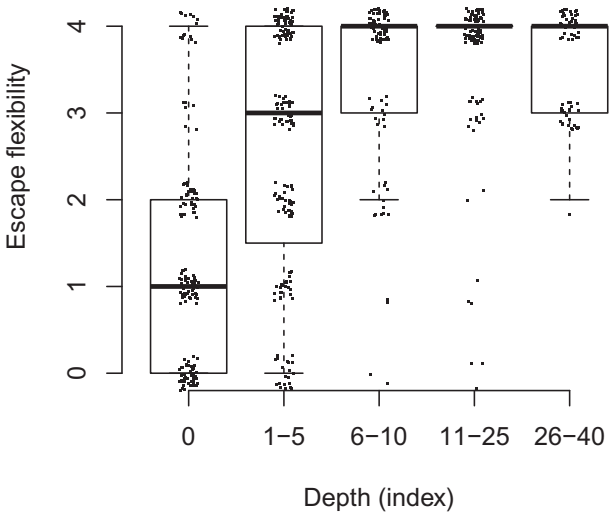

Fig 1. Depth versus Escape Flexibility

increases. The figure shows that US agreements are not only deep, but several of them also feature very long transition periods. By contrast, the African Economic Community (founded in 1991) is a stark outlier in the sense that it is very shallow and has one of the longest transition periods in the data set.

Hypothesis 2 suggests that regime type (*Regime*) conditions the relationship between depth and flexibility. We rely on the Polity IV data set to distinguish between democracies and nondemocracies.[10] In line with the existing literature, we consider countries to be democracies if they have a Polity score of 6 or higher (Marshall, Robert Gurr, and Jaggers 2010; Poast and Urpelainen 2013).[11] We use the smallest of the Polity scores for all members as the value for the PTA, but we get the same results if we use the mean. Of the 587 agreements in our data set, democracies signed 285 (48.6%) of them (we have missing values for 11 agreements). In a bivariate analysis, we find support for the conjecture put forward in Hypothesis 2. The mean value of *Escape flexibility* for shallow agreements (a value of *Depth* (*index*) of 10 or less) is 2.88 for democracies and only 1.95 for nondemocracies. For deep agreements, the value increases to 3.57 for democracies and to 3.64 for nondemocracies. Moving from a shallow to a deep agreement hence matters much more for nondemocracies than for democracies.

Finally, Hypothesis 3 refers to constraints on flexibility. We operationalize this variable (*Flexibility strings*) using six items that capture limits with respect to the imposition of antidumping duties, the provision of subsidies, and the use of the safeguard provisions.[12] These capture provisions that impose WTO rules on the use of the antidumping instrument, the safeguard provision and the provision of subsidies, stipulate that the safeguard provision is only valid during the transition period, create a common pol-

[7] We also used item response analysis as described for depth above to combine these four items into a weighted index. The resulting variable is highly positively correlated with the additive index that we use ($r = 0.90$).

[8] If an agreement is silent on the use of trade remedies, parties have a legitimate expectation that these mechanisms are not used, as they run counter to the main objective of an international agreement that foresees trade liberalization through tariff cuts.

[9] This applies to 20% of agreements. If we take the mean or the minimum value in these cases, we obtain similar results (available upon request).

[10] Results do not change if we use other measures of democracy such as that of Cheibub, Gandhi, and Vreeland (2010).

[11] Our results are not sensitive to this particular threshold. We obtain similar results, which are available upon request, if we use 7 or 8 as the threshold.

[12] This variable partly overlaps with Kucik's (2012) operationalization of "rigidity." We also consider rules concerning subsidies, however, and use references to WTO rules as a shortcut to many of the constraints included in Kucik's measure.

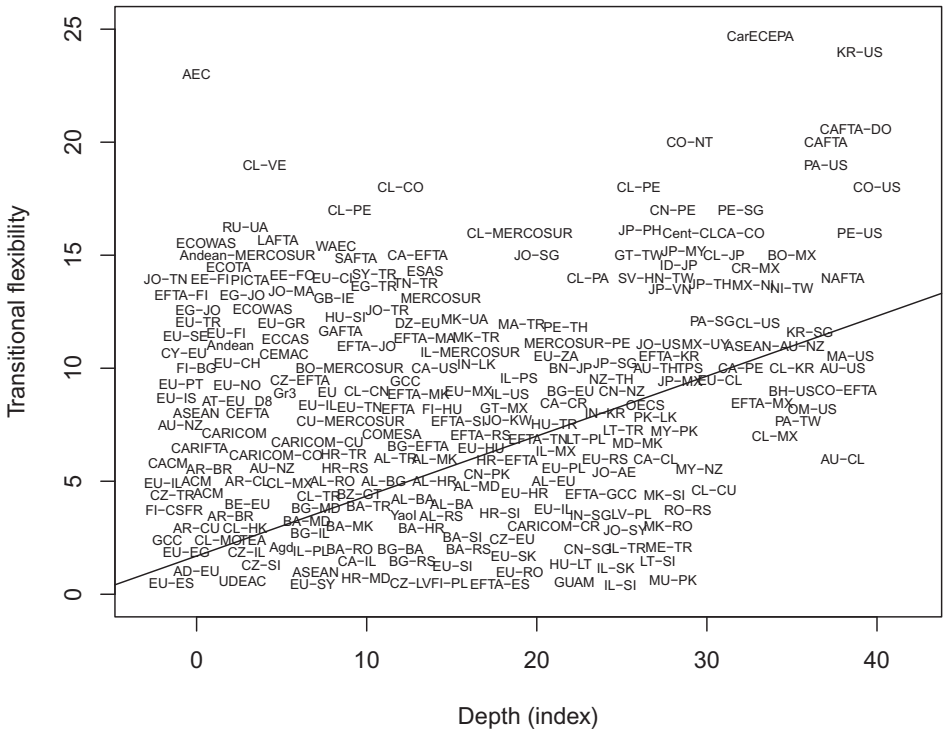

Fig 2. Depth versus Transitional Flexibility

icy on subsidies, and define a minimum dumping margin (the supplementary file contains more information on these items). The variable potentially ranges from 0 to 6, depending on the number of provisions included in a trade agreement that restrict the imposition of antidumping duties and the use of subsidies and the safeguard provisions. In practice, the variable ranges from 0 (195 agreements) to 5 (4 agreements).[13]

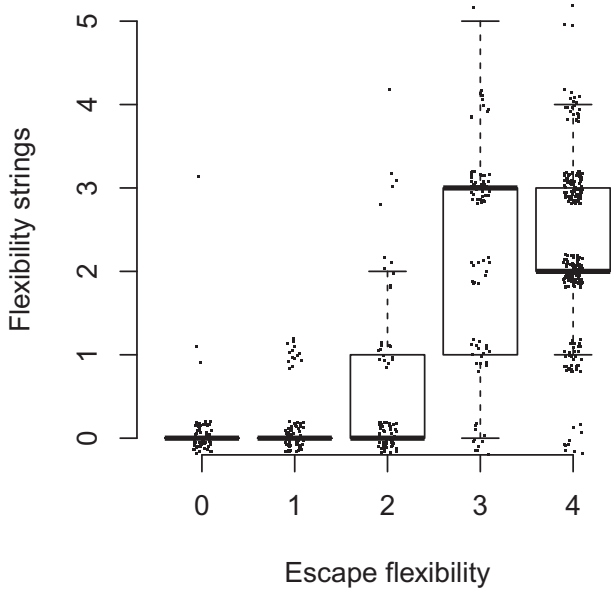

Fig 3. Escape Flexibility versus Flexibility Strings

[13] Again, this variable is highly correlated with one that relies on item response analysis ($r = 0.99$).

Figure 3 offers a first test of our argument concerning a relationship between flexibility and flexibility strings. The relationship between the two variables is positive as expected following Hypothesis 3. As escape flexibility increases, countries make its use more difficult. The relationship is slightly nonlinear, however: At the maximum of escape flexibility, flexibility strings are less pronounced than when escape flexibility takes the value of three. As can be seen in online appendix S1, this nonlinearity is driven by a decline in the number of WTO restrictions on safeguard provisions in highly flexible agreements. This drop is due to European agreements that *de facto* contain the same strings attached to the use of the safeguard clauses as the WTO, but without making an explicit reference to the WTO. We address the issue of nonlinearity in the robustness checks below.

Revisiting the Relationship between Depth and Flexibility

Going beyond the descriptive evidence, we estimate multivariate regression models with flexibility and flexibility strings as dependent variables. For two reasons, we choose the PTA as the unit of analysis and thus consciously depart from previous PTA research that mainly used dyad-year (Mansfield and Milner 2012) or PTA-country-year (Kucik 2012) as unit of analysis. First, the design of a PTA does not vary across member countries in our data set. By using dyad-year or PTA-country-year as units of analysis, therefore, we would multiply the values of our dependent variable in plurilateral agreements. Second, with the exception of a small number of regional agreements (in particular, the European Union), the design of PTAs does not vary over time in our data set. By dropping the time dimension, therefore, we do not lose any information. In fact, using the PTA as the unit

of analysis is a conservative choice to avoid inflating the number of observations and artificially reducing standard errors.

Control Variables

In the following multivariate models, we include several control variables that allow us to deal with potentially confounding factors. Since our unit of analysis is the PTA, we take the minimum value of each continuous variable across member countries to capture the weakest link.[14] In terms of economic variables, we use (the natural log of) total GDP (*GDP*) and GDP per capita (*GDPpc*) to capture the economic importance and income level of a country. Countries with relatively large markets and relatively rich countries may find it easier to design more rigid agreements. Moreover, we include the minimum value of trade flows (imports plus exports) among PTA members (*Trade*). Here, the effect could go both ways: More trade may mean that countries have an incentive to design more rigid agreements, but it could also mean that such PTAs have larger distributional effects, meaning that countries require more flexibility. As is common practice in the empirical trade literature, we take the log of this value.

We also add a variable that captures whether a country has recently undergone a transition from autocracy to democracy (*Democratization*). The expectation is that democratizing countries require more flexibility (and fewer strings) since they face high levels of uncertainty about future states of the world. This variable scores one if all member countries of a PTA have transitioned to democracy over the past ten years, with Polity IV as indicator.

We also add a dummy that scores one if all PTA members are also members of the WTO (*WTO*). WTO members tend to implement trade policies that differ from those of countries that are not members (Mansfield and Reinhardt 2003:830). They also have discretionary flexibility provisions upon which they can rely. We thus expect WTO members to design more rigid agreements. Finally, we include the number of member countries of a PTA (*No. Members*), which previous research hypothesized to be positively related to flexibility (Koremenos et al. 2001). Table 2 summarizes the descriptive statistics of the dependent and independent variables.

Table 2. Descriptive Statistics

Variables	*Mean*	*SD*	*Min*	*Max*
Escape flexibility	2.68	1.46	0	4
Transitional flexibility	3.44	5.04	0	25
Flexibility strings	1.53	1.35	0	5
Depth	8.13	10.03	0	40
GDP	21.30	1.81	14.13	27.39
GDPpc	8.41	0.99	5.07	10.64
Trade	2.96	2.70	0	11.98
Regime	1.55	7.14	−10	10
Regime dummy	0.49	0.50	0	1
Democratization	0.25	0.43	0	1
WTO	0.50	0.50	0	1
No. members	5.65	8.65	2	91

[14] Our results are similar if we use the median or average across member countries. Results are available upon request.

Statistical Models

Since *Escape flexibility* and *Flexibility strings* are ordinal variables, we use ordered probit to estimate the models that include them as dependent variables.[15] We rely on zero-inflated negative binomial regression in the equations that have *Transitional flexibility* as dependent variable, since this is a count variable with a large number of zeros (about 50% of our observations score zero).[16] This estimation technique predicts first the existence of excess zeros using a logistic regression and then predicts the number of years until the end of the transition period using a negative binomial estimation. We include the same variables in the first and second stages.

Findings

Table 3 shows the results of the baseline models. In models 1 and 2, we test Hypothesis 1 for both of our measures of flexibility. In Model 3, we add an interaction term between depth and regime type to test Hypothesis 2. Models 4 and 5, finally, have *Flexibility strings* as dependent variable, allowing us to test Hypothesis 3. In the latter two models, we keep *Depth* as a control variable, to make sure that the distinction between detailed and less detailed agreements does not drive our results. In Model 5, we drop PTAs that have a value of zero on *Escape flexibility*, since we cannot expect states to include flexibility strings in agreements with no discretionary flexibility.

Across all models, the findings conform to our expectations.[17] Indeed, the coefficients of both *Depth* and *Escape Flexibility* are positive and statistically significant at the conventional level.[18] Thus, deeper PTAs are more flexible. Similarly, more escape flexibility goes hand in hand with a greater number of flexibility strings, even after controlling for the depth of a PTA (Model 4). This latter finding proves robust to dropping PTAs with no escape flexibility (Model 5).

To assess whether regime type conditions the relationship between depth and flexibility, Figures 4 and 5 show the marginal effect of nondemocracy and democracy on *Escape flexibility* for different values of *Depth*. We focus on the probability of including no flexibility provisions (Figure 4) and including the maximum number of escape clauses (Figure 5). The figures show, first, that for a given level of *Depth*, democracies tend to sign more flexible PTAs than nondemocracies. Second, and in line with our second hypothesis, the effect of *Depth* on flexibility is larger for nondemocracies than for democracies. Figure 4 shows that as *Depth* increases, the probability of including no flexibility into PTAs decreases substantially more for nondemocracies than for democracies. Similarly, Figure 5 shows that as *Depth* increases, the probability of making a

[15] Results are similar if we use multinomial logistic regression (available upon request).

[16] The test of α shows that a negative binomial model is more appropriate than a zero-inflated poisson model (except for Model A16). The Vuong test shows that a zero-inflated negative binomial model outperforms an ordinary negative binomial regression.

[17] In all models, we lose some observations because some of our control variables have incomplete coverage. Our results are not sensitive to dropping control variables with missing observations.

[18] For model 2, we only show the coefficients of the second equation. In the first equation, as expected, *Depth* and *Trade* are negative predictors of the probability of having zero tariff transition. The other variables are not statistically significant.

TABLE 3. Baseline Models

Covariates	*(1) Escape*	*(2) Transitional*	*(3) Escape*	*(4) Strings*	*(5) Strings*
Depth	0.05***	0.02***	0.07***	0.06***	0.06***
	(0.01)	(0.003)	(0.01)	(0.01)	(0.01)
Escape Flexibility				0.71***	0.73***
				(0.05)	(0.06)
Regime	0.04***	0.01*		−0.01	−0.01
	(0.01)	(0.007)		(0.01)	(0.01)
Regime Dummy			0.77***		
			(0.14)		
Depth × Regime					
GDP	−0.02	−0.05	−0.02	0.17***	0.16***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
GDPpc	0.06	−0.17***	0.05	−0.04	0.002
	(0.07)	(0.05)	(0.07)	(0.07)	(0.07)
Trade	−0.07**	0.06***	−0.07***	−0.05**	−0.06**
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Democratization	−0.07	−0.21**	−0.06	−0.28**	−0.30***
	(0.12)	(0.11)	(0.12)	(0.13)	(0.13)
WTO	0.04	−0.10	0.04	0.40***	0.44***
	(0.11)	(0.10)	(0.11)	(0.12)	(0.12)
No. Members	−0.0005	−0.01	−0.004	−0.002	−0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Cut1	−0.96		−0.83	4.86***	5.14***
	(0.89)		(0.88)	(0.99)	(1.03)
Cut2	−0.42		−0.29	5.71***	6.00***
	(0.89)		(0.88)	(1.00)	(1.04)
Cut3	0.09		0.22	6.70***	7.02***
	(0.88)		(0.88)	(1.01)	(1.05)
Cut4	0.61		0.74	8.32***	8.63***
	(0.89)		(0.88)	(1.02)	(1.07)
Cut5				9.60***	9.91***
				(1.04)	(1.09)
Constant		4.17***			
		(0.71)			
Test of α		61.38***			
Vuong test		9.04***			
Nonzero Obs.		208			
Zero Obs.		263			
Pseudo R^2	0.10		0.10	0.33	0.29
Observations	559	471	559	543	482

(*Notes*. Models 1, 2, 4, and 5 are ordered probit. Model 3 is a zero-inflated negative binomial model.
Standard errors in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$).

PTA highly flexible increases substantially more for nondemocracies than for democracies (when moving from 0 to 20, we see an increase of 40% for nondemocracies and 20% for democracies). The positive correlation between depth and flexibility thus hinges on domestic institutions, as presented in Table 1.

In Table 4, we summarize the magnitude of the substantive effects of Model 1. Moving *Depth* from its minimum to its maximum value increases by 63 [52, 71]% the probability of having the highest value of *Escape Flexibility*, that is, four. The results of Model 2 indicate that moving *Depth* from its minimum to its maximum value increases the tariff transition period by 8 [6, 10] years. The transition period increases by 3 [2, 4] years if *Depth* moves from a standard deviation below the mean to a standard deviation above the mean.

In Table 5, we summarize the magnitude of the substantive effects of Model 4. Moving *Escape Flexibility* from its minimum to its maximum value decreases by 83 [−89, −75]% the probability of having the lowest value of *Flexibility Strings*, that is, zero. Similarly, moving *Escape Flexibility* from its minimum to its maximum value increases by 40 [34, 45]% the probability of observing a value of 3 on *Flexibility Strings*.[19]

The results for the control variables also are largely in line with our expectations. Large countries sign agreements that contain more flexibility strings. Rich countries have shorter transition periods than poor ones. Equally, democratization correlates negatively and WTO membership correlates positively with the number of flexibility strings. Of the coefficients that are statistically significant, only the negative sign for the coefficient for *Democratization* in Model 2 is surprising. Contrary to our expectation, democratizing countries sign agreements with a shorter transition period. The reason here may be that these countries use PTAs as a commitment device and thus want these agreements to take effect as soon as possible (Liu and Ornelas 2014). As expected, the effect of *Trade* on flexibility is ambiguous. Paired with the descriptive evidence provided above, these models thus offer encouraging support for our hypotheses.

[19] In our data set, very few PTAs have values that are equal to 4 or 5.

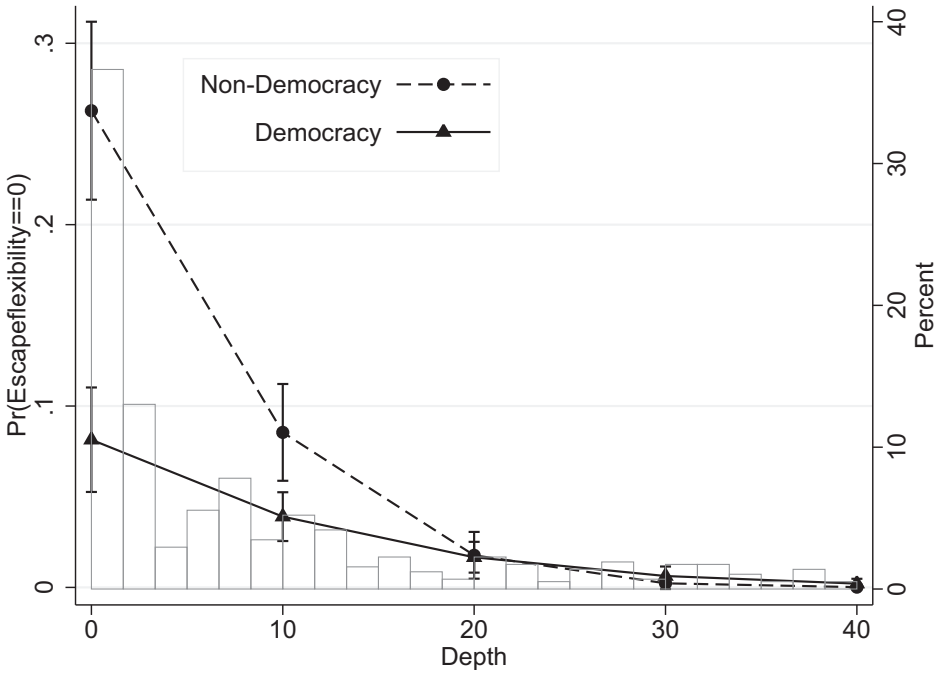

Fig 4. The Effect of *Depth* on the Probability of *Escape Flexibility* Equaling Zero for Democracy and Autocracy

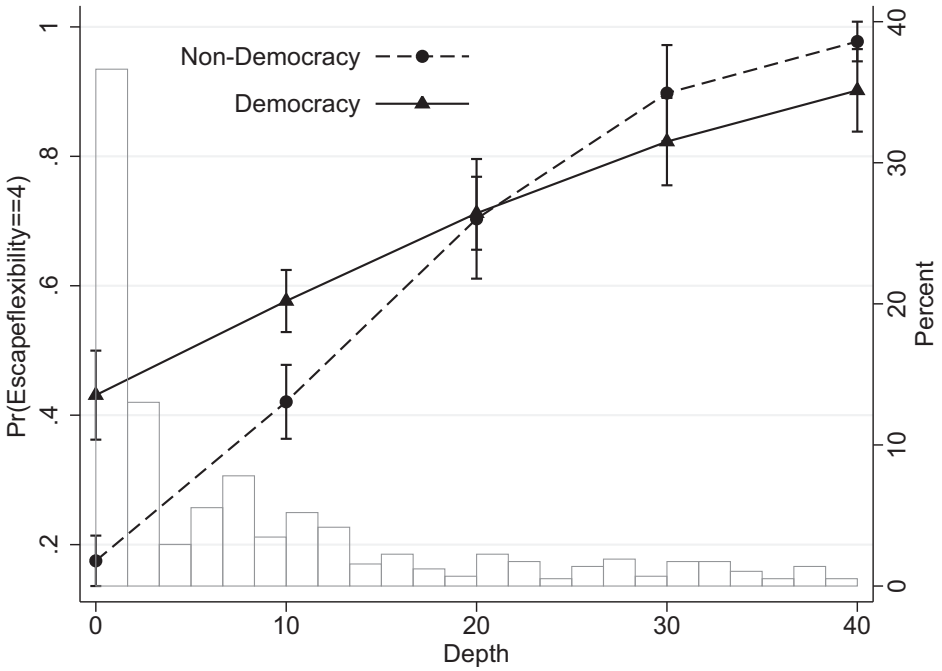

Fig 5. The Effect of *Depth* on the Probability of *Escape Flexibility* Equaling Four for Democracy and Autocracy

Table 4. Predictions for the Effect of *Depth* on *Escape Flexibility* (Model 1)

Value	*Min → Max*	*CI*
0	−0.16	[−0.20, −0.12]
1	−0.16	[−0.19, −0.12]
2	−0.17	[−0.22, −0.13]
3	−0.14	[−0.18, −0.10]
4	0.63	[0.52, 0.71]

Robustness Checks

We perform several tests to check the robustness of our findings. First, we replace the indicators of depth, escape flexibility, and flexibility strings obtained by summing provisions with the indicators of depth, escape flexibility, and flexibility strings that rely on item response analysis (see the discussion above). Doing so does not change our results (see Table A1). Second, Figures 1 and 3 indicate that the relationships between the relevant variables may be nonlinear. To address this issue, we recode *Depth* so

TABLE 5. Predictions for the Effect of *Escape Flexibility* on *Flexibility Strings* (Model 4)

Value	*Min → Max*	*CI*
0	−0.83	[−0.89, −0.75]
1	0.06	[0.01, 0.10]
2	0.34	[0.28, 0.39]
3	0.40	[0.34, 0.45]
4	0.04	[0.02, 0.06]
5	0.002	[0.0002, 0.01]

that it scores one for values between one and five, two for values between six and ten, and so on. Then, we include each value as a dummy on the right-hand side of a model (leaving *Depth*=0 as baseline) in which *Escape Flexibility* is the dependent variable. The model does indeed offer some evidence of nonlinearity (Table A2). The effect of very high values of *Depth*, namely between 36 and 40, is lower than that of values of *Depth* between 31 and 35 (the difference is statistically significant with $p < .1$). Nevertheless, and importantly, the effect of very deep agreements remains positive, as expected in Hypothesis 1. Similarly, we include dummies of each value of *Escape Flexibility* in a model in which *Flexibility Strings* is the dependent variable. The results show that the positive effect of *Escape Flexibility* on *Flexibility Strings* is larger when *Escape Flexibility* is equal to three than when it is equal to two (the difference is statistically significant at $p < .01$). Again, however, the model confirms our key expectation of a positive relationship between *Escape Flexibility* and *Flexibility Strings*. We also exclude these PTAs that have a value of zero in our *Depth* variable (147 observations in our data set). Even in this case, the main results remain unchanged (see models A8, A9, and A10 in the supplementary file).

Third, as suggested above, power asymmetry could affect the relationship between depth, flexibility, and flexibility strings. If a large country such as the United States does not fear competition from a small country, import-competitors in the former only have weak incentives to lobby for the inclusion of flexibility provisions in a PTA. At the same time, the large country could impose flexibility strings on the small country. To test this counterargument, we split our sample into North–South PTAs, and North–North and South–South PTAs. Should the aforementioned objection hold, *Depth* should not be statistically significant in the subsample of North–South PTAs. Table A3 in the supplementary file, however, shows that this is not the case. *Depth* remains positive and statistically significant in these models. In fact, we find no statistically significant difference between the coefficients in the two subsamples.

Fourth, we use Kucik's (2012) data to operationalize *Escape flexibility* and *Flexibility strings* and rerun our main models. We rely on the same operationalization of *Escape flexibility* as above; to measure *Flexibility strings*, we employ Kucik's (2012) measure of rigidity.[20] Table A4 and Figures A10 and A11 show that our main results hold even using this different data set, which includes only half the number of observations that we have in our data set.

Finally, we include a variable counting the years since the signing of the first PTA in our data set, that is, since 1948. This variable accounts for the possibility that a time trend could drive our results, as both flexibility and depth increase over time. It allays concerns that a third variable, not included in our model, which also increases over time, causes the growth in both depth and flexibility that we capture. Table A5, however, shows that our main results are unchanged even after including this trend variable.[21]

Conclusion

Based on an original data set on the design of 587 PTAs, we find that two qualifications apply to the conventional understanding of the depth–flexibility nexus in the design of international institutions. On the one hand, while deep trade agreements tend to be more flexible than shallow ones, this relationship proves less pronounced for democracies than for nondemocracies. This stems from how democracies rely more heavily on nontariff barriers and other forms of contingent protection, such as trade remedies, than do autocracies (Kono 2006). On the other hand, while deep PTAs offer more flexibility, states restrict the additional, discretionary flexibility with relevant strings.

This article speaks to the literature on PTAs (Manger 2009; Baccini and Dür 2012; Mansfield and Milner 2012). The quantitative PTA literature tended to treat all PTAs as equivalent, but recent studies have started to pay attention to the diversity of the design of such agreements. Explaining this variation should help to elucidate the reasons that regimes sign PTAs, as states' motivations to form PTAs may also affect the design of PTAs. Our findings regarding the relationship between depth and flexibility also indicate the need for caution in research on the consequences of PTAs. Such research needs to take into consideration the effects of *both* of these dimensions of PTA design.

Our findings also contribute in several important ways to the literature on the design of international institutions (for example, Koremenos et al. 2001). First, our analysis shows that earlier work overly simplifies the determinants of institutional design. Domestic institutions play an important, but so far overlooked, role in conditioning the relationship between different dimensions of the design of international treaties. Second, we introduce the concept of flexibility strings to illustrate how governments can deal with time inconsistency problems. Governments have strong incentives to ensure that flexibility avoids threatening the stability of deep agreements. Flexibility strings allow countries to maintain predictability even when engaging in deep cooperation. Finally, we add an analysis of PTAs to the literature on the design of international institutions that so far has mainly focused on a few prominent international organizations (but see Baccini 2010; Kucik 2012). Our empirical focus provides an important analytical advantage: In the negotiations on PTAs, all aspects of design are up for discussion. This makes PTAs a particularly interesting case for understanding the interrelation of different dimensions of design.

[20] Since this measure of *Flexibility strings* ranges from 0 to 15, we estimate the model relying on ordinary least squares regression. The results are similar if we use an ordered probit model.

[21] The results are similar if we include the square and cube of the time-trend variable (Carter and Signorino 2010). Since the three variables are highly correlated ($\rho > .95$), we prefer to include only one time-trend variable.

In view of the negotiations for mega PTAs such as the Trans-Pacific Partnership (TPP) and the Trans-Atlantic Trade and Investment Partnership (TTIP), this article also has important real-world implications. It shows that governments can find a balance between depth and flexibility that addresses the concerns of diverse sets of domestic actors. Increasingly, deep agreements, however, require ever greater attention to the fine-tuning of other aspects of institutional design, explaining why the negotiations for new PTAs drag on for so long. Our study also suggests that copying provisions from existing agreements and then presenting them to stakeholders is a tricky exercise for negotiators due to the interdependence between the various dimensions of a treaty, which we have highlighted in this article. Designing an international institution is thus an even more complex task than previously acknowledged.

References

Baccini, Leonardo. (2010) Explaining Formation and Design of EU Trade Agreements: The Role of Transparency and Flexibility. *European Union Politics* 11 (2): 195–217.

Baccini, Leonardo, and Andreas Dür. (2012) The New Regionalism and Policy Interdependence. *British Journal of Political Science* 42 (1): 57–79.

Baier, Scott L., and Jeffrey H. Bergstrand. (2007) Do Free Trade Agreements Actually Increase Members' International Trade? *Journal of International Economics* 71 (1): 72–95.

Bartholomew, David J., Martin Knott, and Irini Moustaki. (2011) *Latent Variable Models and Factor Analysis: A Unified Approach*. Chichester, UK: John Wiley and Sons.

Büthe, Tim, and Helen V. Milner. (2014) Institutional Diversity in Trade Agreements and Foreign Direct Investment: Credibility, Commitment, and Economic Flows in the Developing World, 1970–2007. *World Politics* 66 (1): 88–122.

Carter, David B., and Curtis S. Signorino. (2010) Back to the Future: Modeling Time Dependence in Binary Data. *Political Analysis* 18 (3): 271–292.

Chase, Kerry A. (2003) Economic Interests and Regional Trading Arrangements: The Case of NAFTA. *International Organization* 57 (1): 137–174.

Cheibub, Jose A., Jennifer Gandhi, and James R. Vreeland. (2010) Democracy and Dictatorship Revisited. *Public Choice* 143 (1–2): 67–101.

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 Database. *Review of International Organizations* 9 (3): 353–375.

Egger, Peter, and Sergey Nigai. (2015) Effects of Deep Versus Shallow Trade Agreements in General Equilibrium. In *Trade Cooperation: The Purpose, Design and Effects of Preferential Trade Agreements*, edited by Andreas Dür, and Manfred Elsig. Cambridge: Cambridge University Press.

Elsig, Manfred, and Cédric Dupont. (2012) European Union Meets South Korea: Bureaucratic Interests, Exporter Discrimination and the Negotiations of New Trade Agreements. *Journal of Common Market Studies* 50 (3): 492–507.

Estevadeordal, Antoni, Kati Suominen, and Robert Teh., eds. (2009) *Regional Rules in the Global Trading System*. Cambridge: Cambridge University Press.

Grossman, Gene M., and Elhanan Helpman. (1994) Protection for Sale. *The American Economic Review* 84 (4): 833–850.

Helfer, Laurence R. (2005) Exiting Treaties. *Virginia Law Journal* 91 (7): 1579–1648.

Helfer, Laurence R. (2013) Flexibility in International Agreements. In *Interdisciplinary Perspectives on International Law and International Relations*, edited by Jeffrey Dunoff, and Mark Pollack. Cambridge: Cambridge University Press.

Johns, Leslie. (2014) Depth Versus Rigidity in the Design of International Trade Agreements. *Journal of Theoretical Politics* 26 (3): 468–495.

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.

Kucik, Jeffrey. (2012) The Domestic Politics of Institutional Design: Producer Preferences Over Trade Agreement Rules. *Economics & Politics* 24 (2): 95–118.

Kucik, Jeffrey, and Eric Reinhardt. (2008) Does Flexibility Promote Cooperation? An Application to the Global Trade Regime *International Organization* 62 (3): 477–505.

Liu, Xuepeng, and Emanuel Ornelas. (2014) Free Trade Agreements and the Consolidation of Democracy. *Forthcoming in American Economic Journal: Macroeconomics* 6 (2): 29–70.

Manger, Mark. (2009) *Investing in Protection: The Politics of Preferential Trade Agreements Between North and South*. Cambridge: Cambridge University Press.

Mansfield, Edward D., and Helen V. Milner. (2012) *Votes, Vetoes, and the Political Economy of International Trade Agreements*. Princeton: Princeton University Press.

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.

Marshall, Monty, Ted Robert Gurr, and Keith Jaggers. (2010) Polity IV Project: Political Regime Characteristics and Transitions, 1800–2009. Center for Systemic Peace.

Maskus, Keith E., and Mohan Penubarti. (1995) How Trade-Related Are Intellectual Property Rights? *Journal of International Economics* 39 (3): 227–248.

Neumayer, Eric. (2007) Qualified Ratification: Explaining Reservations to International Human Rights Treaties. *The Journal of Legal Studies* 36 (2): 397–429.

Pelc, Krzysztof J. (2009) Seeking Escape: The Use of Escape Clauses in International Trade Agreements. *International Studies Quarterly* 53 (2): 349–368.

Piermartini, Roberta, and Michele Budetta. (2009) A Mapping of Regional Rules on Technical Barriers to Trade. In *Bank Regional Rules in the Global Trading System*, edited by Antoni Estevadeordal, Kati Suominen, and Robert Teh. Cambridge: Cambridge University Press.

Poast, Paul, and Johannes Urpelainen. (2013) Fit and Feasible: Why Democratizing States Form, Not Join, International Organizations. *International Studies Quarterly* 57 (4): 831–841.

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* 54 (4): 829–857.

Slapin, Jonathan B., and Julia Gray. (2014) Depth, Ambition and Width in Regional Economic Organizations. *Journal of European Public Policy* 21 (5): 730–745.

World Trade Organization. (2011) *World Trade Report – The WTO and Preferential Trade Agreements: From Co-Existence to Coherence*. Geneva: World Trade Organization.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. The politics of trade agreement design operationalizing depth.