Are the Contents of International Treaties Copied-and-Pasted?  
Evidence from Preferential Trade Agreements

Todd Allee  
Department of Government and Politics  
University of Maryland  
tallee@umd.edu

Manfred Elsig  
World Trade Institute  
University of Bern  
manfred.elsig@wti.org

February 9, 2015

We thank to Andrew Lugg for excellent research assistance and acknowledge generous financial support from the Swiss Science Foundation (www.nccr-trade.org).
Introduction

A lengthy preferential trade agreement (PTA) signed in 2008 contains multiple annexes to its government procurement chapter (Chapter 14) that list entities, goods, and services in Canada and Colombia that are covered by the agreement’s procurement rules. These annexes that contain the “Schedule of Colombia” are unremarkable, until one realizes that the agreement in question is actually between Canada and Peru! Canada was indeed negotiating a PTA with Colombia around that same time, and would sign that agreement a week later. This odd insertion of language about Colombia in an agreement with Peru certainly appears to be the result of sloppy copy-and-pasting between treaties. To lawyers the above mistake may be unsurprising, since legal scholars frequently note the propensity for contracts to be drawn from standard or “boilerplate” language. Likewise, scholars from international relations and international law sometimes assert that treaties are produced from templates or models – although evidence to substantiate these claims is scant. Yet the fact that some bilateral investment treaties (BITs) have been concluded at weekend conferences among diplomats is at least consistent with this notion. But could it really be that significant portions of international treaties, even major ones, are copy-and-pasted from one agreement to the next?

The above depiction flies in the face of long-standing portrayals of international treaty-making as well as known facts about the contents of treaties and the process by which they are concluded. The international relations literature is filled with game-theoretic studies of international bargaining in which two states exchange repeated offers and tussle over differences in preferred outcomes. There also exist rich literatures on both diplomacy and international negotiation, both of which envision international agreements as resulting from various interactions over time. PTA negotiations, in fact, routinely stretch for three, four, or five years,
involving multiple meetings of the parties per year.¹ Likewise, securing domestic ratification of

treaties is often a length and contentious process, which seems odd if the language in treaties is

uncontroversial and merely “boilerplate.”² Furthermore, scholars increasingly code using

numeric indicators the varying contents of treaties like PTAs, further attesting to the ideas of

treaty heterogeneity. Finally, if one flips through the pages of PTAs, one uncovers numerous

provisions that are unique and at times peculiar, such as language in the India-Bhutan PTA

allowing reciprocal selling of one another’s lottery tickets or Singapore permitting limited

imports of U.S. chewing gum in their bilateral PTA. Looked at in this way, then, there is reason

to believe that international agreements are carefully and individually negotiated, and not merely

drawn from a template or copy-and-pasted from an existing treaty.

Revealing whether, and to what degree, the contents of international treaties are the result

of copy-pasting is important for several reasons. First, this negotiation approach – if it exists –
could be an efficient and strategic response by low-capacity states to the challenges of

negotiating international agreements. Or, stronger states might simply advocate uniform rights

and obligations that reflect their interests. A second reason to probe such dynamics is that

boilerplate language can have unintended consequences, and thus the formulaic inclusion of

certain text could produce future surprises and problems.³ Third, our findings can answer the

question of just how much fragmentation there is in the international trading system. If the rising

tide of PTAs negotiated during the recent period of WTO stagnation is more standard than

¹ For instance, the 2000 PTA between the EU and Mexico was negotiated over a period spanning five years. Similarly, Chile’s very recent PTAs with Malaysia and Vietnam each took four years to negotiate. Finally, the 2007 U.S. PTA with Panama, which followed many earlier U.S. PTAs, took more than three years to conclude.
² In 2012 U.S. Senator Rand Paul (R-KY) held up ratification of U.S. tax treaties with Hungary, Luxembourg, and Switzerland (alleging that they gave the U.S. government too much power). A lobbyist supporting the treaties responded that Paul was “…objecting to boilerplate language” and that the language was “…standard policy for the U.S.” See Bernie Becker and Alexander Bolton, “Rand Paul Blocking Tax Treaties over Fears of Government Snooping.” The Hill (online). January 13, 2012.
³ A current example of this dynamic is seen in bilateral investment treaties, whose allegedly boilerplate language on investor-state dispute settlement has had major effects on state parties.
believed, then the expanded network of PTA could serve as an imperfect substitute for multilateral rules and “multilateralizing” these agreements could potentially be less tricky than widely anticipated. Fourth, recent years have seen a shift in the study of international institutions toward coding outcomes while black-boxing processes. In this respect, our study puts the spotlight back on how diplomacy and negotiations actually transpire (e.g., Putnam 1988). Fifth, our study contributes not only to the substantive literature on international institutional design, but also to the growing collection of interdisciplinary studies that employ text-analysis methodology. Our approach supplements more conventional efforts by social scientists to numerically code international treaties. We offer an intuitive, transparent, and compelling method for text analysis that avoids some of the complex challenges associated with other forms of text analysis, which attempt to discern meaning and sentiment from text.

To illuminate whether international treaties might exhibit a copy-and-paste dynamic, we systematically compare the texts of several hundred agreements that address a common issue, in our case trade liberalization. We generate overlap percentages for thousands of pairs of PTAs using unique software that allows for mass-comparison of large numbers/quantities of texts. Somewhat surprisingly, particularly in light of many multi-year negotiations, we find that most PTAs take the overwhelming majority of their content verbatim from existing agreements. Furthermore, in many cases, three-quarters or more of the text is lifted from a single, pre-existing agreement. Our overall conclusion that treaties are far from unique is robust to the method of text comparison, and our numbers are somewhat conservative. Because of this widespread tendency toward copying-and-pasting, we assert that observers should reassess and reconceptualize how international treaties are made. Although much remains to be done to illuminate the pathways by which treaty language travels, some preliminary evidence suggests
that replication of treaty text is more common in recent years, in similar geographic areas, and among the same or very similar actors.

**The Art of International Treaty-Making**

Negotiating and signing international treaties is the linchpin of international cooperation. Nearly all governments worldwide make pledges within formal agreements to address a wide variety of problems across issue areas, ranging from the environment, human rights to security. Moreover, much of this contracting between countries today occurs at the bilateral and plurilateral level. Such agreements are particularly numerous and prominent in the economic realm, since thousands of BITs and hundreds of PTAs are currently in force.

**Recognizing Treaty Heterogeneity**

Historically, scholars quite understandably took the “dummy variable” approach to understanding many types of international treaties. The presence of a PTA between two countries, for instance, would be captured by a “1” in quantitative studies. In effect, PTAs were treated as undifferentiated from one another. This was true regardless of whether the trade agreement was a dependent variable to be explained (e.g., Baccini and Dür 2012; Baier and Bergstrand 2004; Mansfield, Milner and Rosendorff 2002) or an independent variable that might affect trade or non-trade outcomes (e.g., Baier and Bergstrand 2007; Egger and Larch 2008; Mansfield and Milner 2012). More generally, the practice of treating an international institution as an undifferentiated independent or dependent variable has characterized quantitative studies of international organizations for the past few decades (Martin and Simmons 2000).
This approach has worked well as everyone has sought to learn more about treaty determinants and treaty effects, but over time there has been increased appreciation of the degree to which agreements like PTAs and BITs can differ from one another. Scholarship has moved from treating PTAs as uniform to measuring and predicting some variation across major treaty elements (see Mansfield and Milner 2012), with variation across dispute settlement being a long-appreciated area in which heterogeneity is recognized (Allee and Elsig 2015; Jo and Namgung 2012; Smith 2000). Likewise, studies of regional organizations increasingly emphasize variation across regional treaties (Acharya and Johnston 2007; Haftel 2012). Particularly noteworthy are the new generation of empirical studies that identify and explain the way in which economic agreements or regional economic organizations vary in their flexibility (Kucik 2012, Pelc 2009), depth (Dür, Baccini and Elsig 2014), and scope (Haftel 2013). Similar dynamics are present in the literature on BITs, as scholars are now chronicling and explaining variation from one BIT to the next (Allee and Peinhardt 2014; Neumayer et. al. 2014) as well as the effects of that variation (Allee and Peinhardt 2011; Berger et.al. 2010; Simmons 2014). In sum, the fact that seemingly-similar international institutions vary in important ways is becoming widely accepted and is a growing area of empirical focus.

But in many ways this is where the consensus ends. Virtually everyone acknowledges that prominent international agreements like PTAs and BITs are not fully uniform and should not be thought of as such. But the degree to which agreements like PTAs vary remains somewhat unclear. Should they be thought of as agreements that are mostly similar to one another, or that are quite distinct? Furthermore, we want to know not only how pronounced the differences are, but whether the contents of an agreement are so-called “boilerplate” taken from an existing treaty
or template, or instead are they unique to a particular agreement and carefully crafted by the parties? A wide range of literatures inform this question.

*Why Treaty Texts Should be Unique*

A first perspective views the text of international agreements as resulting from conscious choices as to what to include or exclude from a treaty. Agreements are crafted by the parties to meet their individual or collective preferences, not inspired by a one-size-fits-all model or template. This approach has its roots in literatures on negotiation, diplomacy, game theory, and international organizations more generally. The first three, in particular, emphasize the way in which agreements are generated, which are unique to a particular environment and particular actors. Attempts to produce a treaty, therefore, are far from formulaic or automatic.

The study of diplomacy long has been concerned with the making of treaties, and the defining features of diplomacy suggest that treaties should be distinct from one another. Treaties, like other diplomatic outcomes, are produced through various exchanges between often-varied governmental actors. One aspect of diplomacy which suggests treaty heterogeneity is that many actors are involved, ranging from political leadership to bureaucrats to foreign diplomats and global civil servants. Furthermore, diplomacy can be carried out in many different venues: bilateral, multilateral, conferences, summits, and ad hoc negotiations, among others. Finally, there are no set rules governing the output of structure of treaties. Even within a common treaty type, such as PTAs, states are free to structure the treaties in whatever way they like.

---

4 Cooper, Heine and Thakur 2013, part II.
5 Cooper, Heine and Thakur 2013, part III.
Scholarship concerned specifically with international negotiations also suggests that international agreements should be unique (e.g., Berton, Kimura, and Zartman 1999; Starkey, Boyer, and Wilkenfeld 2005). This is because varied inputs produce different negotiation outcomes. Culture, for instance, can shape negotiations in many ways (Faure 1999). Likewise, structural features such as power – which may differ from one negotiating setting to the next – also affect the specific negotiation results. Factors such as these will vary based on which countries are negotiating an agreement and the conditions under which they are negotiating. Individual negotiators also may differ in terms of their styles and skills. Finally, the duration of negotiations also varies, and can affect the length, content, and complexity of a resulting agreement.

Another useful conceptualization is found in the accumulated body of work on international bargaining, including game-theoretic work on international cooperation via international institutions (see Gilligan and Johns 2012). In general, actors attempting to reach an agreement hold different preferences over outcomes and make offers to one another in an attempt to reach a solution. Situations in which actors may bargain can vary based on number of players, amount of information, and structural form. One might claim that actors negotiating a PTA are facing a similar structure and thus may produce similar treaties. Yet those actors may differ by “type” or other characteristics. They also may face vary different domestic political constraints, which also should create divergence in agreement-design outcomes.

The prominent literature on international legalization (Abbott et al. 2000) also serves as an important foundation for this general perspective. In the abstract, legalization represents the “move to law” (Goldstein et al. 2000, 385) in world politics, and in reality, this is largely
accomplished through the signing of treaties, agreements, and other commitments. A key message is that although we see increased legalization overall, this occurs in various “institutional forms” and with varying degrees of obligation, precision, and delegation specified within a cooperative agreement (Abbott et al. 2000). In fact, one could claim that the entire emphasis on international legalization in recent decades is motivated by the fact that such legalization varies widely in form and degree.

Most empirical work to date on international institutions also suggests that treaties are uniquely negotiated and not copied from models or templates. Beginning fifteen years ago, several empirical studies began to emerge – mostly qualitative, and in the form of single or comparative case studies – to explain the differing “design” of international agreements. Some of these have been in the rationalist tradition (e.g., Kydd 2001, Morrow 2001, Oatley 2001). More recent empirical studies in this tradition explore agreement duration (Koremenos 2005) or institutional change (e.g., Jupille, Mattli, and Snidal 2013). Still others take a more historical-institutionalist perspective, such as Acharya and Johnston’s (2007) edited volume on regional institutional design, which contains six region-specific chapters. Regional institutions are once again the product of careful design, and as the volume’s title suggests, they are “crafted” by their member states. Finally, the empirical legal literature has showcased variation in international institutions like PTAs. Noteworthy are detailed descriptions of the design features of trade agreements, particularly with regard to prominent features such as dispute settlement procedures (Bartels and Ortino 2006; Porges 2011).

---

6 Prominent international legal bodies like the International Criminal Court (ICC) or the dispute settlement mechanism (DSM) of the WTO are born through treaties. This is also how such institutions, and more general international organizations (i.e., the European Union) evolve or are amended (e.g., Jupille, Mattli, Snidal 2013).

7 In fact, the largest section of the introduction to the most prominent study of international legalization (Goldstein et.al. 2000) centers on “the uneven expansion of legalization” including “variation in legalization.”

8 The full title is Crafting Cooperation: Regional International Institutions in Comparative Perspective.
Finally, a recent wave of large-n empirical studies is characterized by the compilation of impressive original data on institutional design. As an illustration for PTAs, the most ambitious data collection effort codes more than one hundred design-feature variables for all post-war PTAs, nearly 600 in total (see Dür, Baccini and Elsig 2014). Others have collected data on certain subsets of PTAs, ranging from trade remedies (Kucik 2012) to competition provisions (Bradford and Büthe 2015) to non-trade issues (Milewicz et al. 2014). Efforts also exist to code variation across more than 1,500 BITs on issues such as the rules for resolving future investment disputes (Allee and Peinhardt 2014). Likewise, data on environmental treaty design also have been coded for use in large n-studies on the effects of environmental treaties (Bernauer et al. 2013). Collectively, then, a large trove of data on international institutional design is being amassed, and many different explanations are being put forward to explain this newfound design variation.9

The above discussions strongly suggest that the contents of agreements like PTAs are not copied-and-pasted. Each treaty arises out of a unique process, and will contain elements that are carefully crafted, over rounds of bargaining, to meet the preferences and needs of the parties to this particular agreement. From this vantage point, one should expect to encounter relatively little overlap in treaty language, even when compared the text of the treaties within the same issue area.

Why Treaty Texts Might Look Similarly

A very different picture of international institutional design also can be seen when looking at legal scholarship on contracting and nascent scholarship in international relations on

---

9 Explanations range from other treaty design characteristics (Rosendorff and Milner 2001, Dür, Baccini and Elsig 2014), domestic politics (Allee and Peinhardt 2010; Mansfield and Milner 2012), economic asymmetry (Smith 2000), and power (Allee and Peinhardt 2014; also Broude 2004).
emulation and treaty models. This emerging portrayal acknowledges that agreements are not entirely homogenous, but envisions less variation and more consistency across agreement terms. This stems primarily from the process by which agreements are negotiated. Whereas the earlier approaches emphasized bargaining and multiple rounds of negotiations, this approach suggests a far less thorough and careful process. Instead, it assumes that much of the contents of any treaty is taken from an existing source – whether a previous treaty or some type of template or model. If true, the observable implication is that agreements of a certain type, such as PTAs, should exhibit considerable overlap with one another.

The body of scholarship that most directly informs this perspective is the legal literature on contracting. Standard-form contracts, which can be thought of as a template, are widely used in domestic law. Across a wide variety of settings, the actors that write up any type of legal document will rely heavily on standard or boilerplate language from an existing document (e.g., Ben-Sharar and White 2006; Gulati and Scott 2012; Hill 2001; Kahan and Klausner 1997; Radin 2013). Within the contracting literature, then, boilerplating is a fact of life – a common and seemingly ubiquitous practice for the drawing up of binding legal documents. One potential advantage is efficiency, since language can be taken “off the shelf” instead of being created from scratch. Boilerplate language also can become a focal point that is inserted when parties disagree on elements of a contract. The primary risk with relying on boilerplate language – which is a substantial one – is that users may not be fully aware of what they are specifying or its myriad consequences (Radin 2013)

Gulati and Scott’s (2012) study of the prevalence of boilerplating in sovereign debt contracts is a particularly relevant application. The compelling title of their book, The Three and a Half Minute Transaction: Boilerplate and the Limits of Contract Design, suggests the speed
with which a large law firm can turn out a boilerplate sovereign debt contract. They interviewed hundreds of lawyers at firms that write sovereign debt contracts and find widespread evidence of boilerplate contracting, even in the presence of reasons to modify or tailor contracts as well as unexpected shocks that cast doubt on particular boilerplate provisions. Gulati and Scott (2012, ch. 3) outline several possible reasons for the prevalence of boilerplating, including but not limited to: satisficing, other cognitive biases and limitations, risk aversion, organizational routine, free riding, and complexity. One could see how many of these ideas would apply to international negotiations, with trade negotiators being asked to negotiate multiple treaties at once, a trade ministry following standard operating procedures, or underfunded bureaucratic searching for existing treaty models.

Within the realm of international treaty-making, the presence of models and the possible use of templates is particularly relevant to BITs, as legal and policy-oriented scholars have long highlighted. At the broadest level of thinking about BITs, the presence of two “types” of BITs – a North American one and a European one – is well established in the legal community. Interestingly, some governments devise their own “model BIT,” with the idea that it will be used as the basis for future negotiations with partners. The practice dates back nearly three decades to the first U.S. model BIT (Vandevelde 1988), although dozens of countries now produce their own models. The model treaties are frequently analyzed and dissected, both individually (e.g., Bernasconi-Osterwalder and Johnson 2010, 2011) and comparatively (e.g., Brown 2013; Dolzer and Schreuer 2012).

It is widely believed that these model treaties serve as the basis for BIT negotiations, although whether and how they shape treaty outcomes is somewhat less clear. Scholars often assert that significant clauses of negotiated BITs are boilerplate (e.g., Moon 2012) and/or look a
lot like a powerful country’s model BIT. Anecdotal evidence certainly supports this claim.\textsuperscript{10} Even developing countries in Africa seem to follow Western BIT models, despite more applicable models and incentives to do otherwise (Ofodile 2013). But these common perceptions of model-using and BIT homogeneity are not systematically tested across treaties.\textsuperscript{11} Nevertheless, it is clear that models exist for BIT negotiations and it is assumed that negotiations are built, at least in part, upon these models.

Most relevant for us is the fact that some political scientists have begun to put forward case studies that emphasize models and templates, often in the PTA context. A group of studies focus on how mostly E.U. institutions serve as a model for other regional agreements (\textit{e.g.}, Alter 2012; Börzel and Risse 2012; Jetschke and Lenz 2013; Jetschke and Murray 2012; Lenz 2012). Likewise, Arnold and Rittberger (2013) argue that the design of the Mercosur’s dispute settlement system was inspired by both the World Trade Organization’s legal system and the European Court of Justice. Some recent work has focused on selected aspects of PTAs and how different templates could be defined (Estevadeordal et al. 2009), quite some of it in Asia (Fink and Molinuevo 2008). Focusing on trade remedies, Elsig and Serrano (2014) look at the inclusion of antidumping (AD) provisions in PTAs by seven major countries in an attempt to identify AD “templates.”

Several recent large-n studies also claim that PTA contents are the product of models or boilerplating. In a quantitative study on dispute settlement provisions in PTAs, Jo and Namgung (2012) argue that states that include legal dispute settlement provisions in past PTAs are more


\textsuperscript{11} One exception is a working paper by Waibel (2012) in which he numerically codes three treaty provisions for 137 BITs to see how much negotiated treaties deviate from the model BITs of seven countries.
likely to include them in future PTAs, evoking the idea of legal “boilerplates.” Rühl (2012) pushes this idea further, claiming that dispute settlement mechanisms (DSMs) of the WTO and NAFTA have served as models for PTA DSMs and pursuing this idea with several numeric indicators. Perhaps most directly applicable is a book chapter by Baccini, Dür and Haftel (2015), in which they identify the consistent presence of three PTAs types or “models” (an E.U. model, a U.S./NAFTA model, and a “Southern” one) across the universe of post-war PTAs. They arrive inductively at these three models, utilizing cluster analysis and drawing upon dozens of quantitative indicators of PTA provisions. All of the aforementioned studies are unified by the idea that PTAs are not independent and that the design of PTAs draws heavily upon a selected number of models or templates.

The preceding literatures suggest, then, that when one probes more carefully into the actual texts of PTAs, one should observe significant overlap in the language across agreements. It is not just that a given PTA should be inspired by another PTA, but the contents should be taken directly from another agreement. From a text analysis standpoint, then, a significant percentage of the text in a given PTA should be taken from one or more existing PTAs.

**Empirical Tests**

Although both vantage points are compelling, they are seemingly diametrically opposed and real-world evidence in support of either is lacking or at best incomplete. This is why carrying out careful empirical tests are crucial. One on hand, there certainly is a belief among many that treaties vary considerably, thus the growth in studies of international institutional design and the launching of numerous efforts to code variation across treaties. But these coding

---

12 Jo and Namgung (2012) evoke the idea of treaty emulation and boilerplates, even specifying the later in their title. Their tests, however, are based on a numeric coding of a treaty provision instead of an analysis of treaty text.
efforts are necessarily selective and incomplete: many code only one or two design variables, which often are binary. Our approach, to engage the entire contents of treaties, provides a very different way to potentially highlight the differences across treaties. On the other hand, there is well-grounded and growing support for the idea that treaty-making is based on models and templates – but this is an inherently difficult idea to test with conventional data and regression analyses. To really examine whether treaties (or another other legislation or document) are copied-and-pasted one has to analyze actual text.

Our approach to text analysis is very different from the wave of other text-based studies that have become popular across the social sciences. Political scientists, for instance, have utilized numerous methods to extract selected bits of information from text documents for purposes of classification or scaling (e.g., Grimmer and Stewart 2013). In our analyses, by contrast, we are not forced to prioritize certain words or to attribute meaning to elements that we specify. Instead, we analyze and compare entire documents, which is consistent with our research question. To carry out our large-scale comparisons, we need a methodology that allows us to do three important things: i) to compare the entire text of one treaty with the entire text of another treaty, ii) to provide a quantifiable indicator of the percentage of overlapping text between treaties, and iii) to do this in an efficient manner, for hundreds of treaties and thousands of comparisons.

The best fit for our analyses is DupFreePro, a text-comparison software that performs all three tasks well with minimal limitations. It allows for the visual, side-by-side comparison of two texts, highlights areas of overlap, and calculates the percentage of content that is unique to,

---

13 The primary limitation is with the size of documents that DupFreePro can compare in bulk. The multi-document comparisons do not work well when documents size begin to exceed 10,000 words, so we limit comparisons to treaties under 75kb file sizes. We still retain more than 85% of treaties that meet our case selection criteria, however.
as well as shared among, both treaties. We use the Bulk Compare option, which for each of our 211 PTAs generates pairwise comparisons of the percentage of text that is drawn from each of the PTAs in existence at that time. Among the two primary specification options, we specify five words as the minimum amount of text needed to constitute a “match” and include common words in these counts – although we modify these two assumptions later with no effect on our central conclusions.

To carry out the comparisons, we standardize the texts of all agreements in our analyses. Since nearly all PTA texts are in .pdf format, we first convert those texts to Microsoft Word before ultimately converting them to plain text format for analysis in DupFreePro. For standardization purposes, and to avoid biasing results in the direction of low overlap, we systematically eliminate the initial and concluding sentences of each document, which provide unique location, date, and other information. Since key commitments are in the main text of the treaty, we eliminate any annexes to the agreements unless a particular annex clearly contains content that in other PTAs is included in the main text (articles and chapters). Annexes, when they exist, typically contain agreement-specific and somewhat esoteric language and data on rules of origin, tariff schedules, etc.  

We run the comparisons using all English-language PTAs between two actors (state-state or regional actor-state) signed during the postwar period. We eliminate approximately 25 agreements because they involve a micro-state or non-sovereign territory, have limited scope

---

14 Product descriptions in tariff schedules are harmonized, so including them would likely bias upward our conclusions about text overlap.

15 At this stage of the project, we focus on two-party (“bilateral”) PTAs because they are more tractable and it is easier to identify patterns among signatories, as compared to the cumbersome nature of some multilateral PTAs, which can have more than fifty members. Given the nature of the endeavor we focus only on treaties in a common language (English). In the future we also plan to compare among Spanish-language treaties. We also exclude the very longest treaties because of their propensity to drive findings and due to software limitations (see fn. 10).
(i.e., services only), or are merely supplemental protocols.\textsuperscript{16} We only compare an agreement to PTAs that were signed in previous year, or the same year as the agreement in question. All told, we examine and compare the contents of 211 bilateral PTAs signed between 1954-2009 (Dür et al. 2014). This results in a sizeable 22,351 pairwise comparisons of the percentage of text in a given PTA that matches the text in each of the other PTAs in existence at the time.

\textit{Findings}

Our analyses show, quite powerfully, that most of the language in PTAs is indeed taken directly from other PTAs. For each of the 211 PTAs we analyze, we first look at the percentage of text that is taken from its closest match among existing PTAs. The results are striking. Figure 1 shows the distribution of maximum-match percentages for each of the 211 treaties. Overall the distribution is tilted heavily to the right. On average, a PTA will share 58\% of its text with its most-similar treaty partner. Furthermore, the median treaty among our 211 PTAs takes a very sizeable 69\% of its text from an existing treaty source. The most likely outcome, in fact, is for a PTA to take 85-89\% of its text from a single source treaty (see Figure 1). Moreover, the three most common options depicted in Figure 1 all entail a PTA drawing more than 75\% of its text from an existing source. We also note that the minority of PTAs that overlap relatively little with another treaty fall in the lowest overlap percentage (0-20\%) instead of more moderate ranges (20-40\%).

[Figure 1 here]

Figure 2 focuses on the upper half of PTAs that copy more than 70\% of their text from another treaty, depicting the number of treaties that overlap at various high percentages. At the

\textsuperscript{16} We also encountered a handful of situations in which a treaty text was damaged and we could not identify an undamaged alternative.
highest end, six treaties take more than 95% text from another, single treaty – with two taking a staggering 99%. An additional 17 PTAs share between 90 and 95% of their language with another treaty. As suggested earlier, the largest collection of treaties fall in the 80-90% match-range; nearly one-quarter of treaties fall in this area, with 86% overlap being the most common outcome.

These overlap numbers are not only striking, but also robust. Our default is to require five or more consecutive words of any size or commonality to constitute a text “match,” since our trial runs with the software suggested this requirement struck an appropriate balance. Thus all figures reported in this paper are based on these parameters. Yet to check for robustness we also explore more liberal and more restrictive requirements for text-matching. Table 1 contains the average increase (decrease) in text overlap, using different parameters, based on a random sample of treaty comparisons. The results are consistent with expectations, and the changes are fairly modest. They certainly do not affect our substantive conclusions. In three cases, lower-threshold parameters increase the resulting overlap, but in most cases only by one or two percentage points. In three other cases, higher-threshold parameters reduce the reported match, but by 10% or less, on average (see Table 1).

Furthermore, the reported overlap numbers are somewhat conservative. For our overlap percentages, we report the amount of text drawn from a single, existing treaty – the treaty with the greatest single overlap with the PTA being analyzed. We do this for simplicity, consistency, and tractability. But there often are additional treaties – the second-, third-, or even tenth-best matches – that also overlap significantly with the treaty in question. If we cycled through each
subsequent closest match, we likely would generate a (considerably) higher overall overlap percentage as we locate additional text in other existing treaties that is utilized in the new treaty.

Figure 3 provides an illustration of how much overlapping text is also found in the second- and third-closest PTA matches. The upper bar indicates how much language is drawn from the closest-matching PTA – as reported in Figures 1 and 2. Below that are bars indicating how much overlapping text is found in the next-two-closest matches. In a few cases it is a clear that a single treaty is the clear source of the copied language – with the next-closest match falling to the level of 50-60% overlap, or occasionally less. Yet in most cases there is relatively little “drop off,” with second-best matches overlapping to the tune of 80% of more. Thus, even if the closest-match PTA had never been signed, the subsequent PTA might still look a lot like various other PTAs. It is difficult to disentangle thos web of reinforcing text overlap, which is something we plan to address in future work. Nevertheless, once one moves beyond the closest-match, or primary source, for a given PTA’s text, several other treaties “fill in” at least some of the remaining uniqueness.

[Figure 3 here]

It also is helpful to identify the PTAs that borrow the most text from elsewhere, in order to get an initial sense of when copying-and-pasting is most likely to occur. Table 2 lists the 30 PTAs that share the greatest amount of text with another, earlier PTA. One thing that jumps out is the sizeable number of countries from central and eastern Europe. In addition, several countries in southeastern Europe and west-central Asia also make this list. One argument mentioned earlier is that low-capacity countries might find it desirable to copy from an existing treaty as an efficiency shortcut. Probing further, the time periods in which these countries are doing the copying are periods of political and economic transition. This adds other possible
motivations, such as embracing a certain template or concluding an agreement quickly as a signal towards embracing market economy status. It also raises the possibility that common PTA terms are somehow encouraged by outside actors, such as the European Union, IMF, or World Bank. More recent years are heavily represented in the list, too. It is difficult to know whether it simply reflects that many countries have undergone transitions in the past two decades or whether it reflects some increased tendency over time toward “borrowing” language or treaty convergence, perhaps drawing on selected templates (e.g. NAFTA, new generation of EC agreements, WTO treaties). Nevertheless, Table 2 also shows that high levels of copying-and-pasting occurred as far back as the 1970s (the two Finland PTAs). Finally, larger entities like the EU and EFTA countries are prominent on this list. One explanation is that these bigger economic actors often come to the table with a PTA template, which is accepted largely intact by a weaker partner, and may be helped by the fact that their PTA partners often have similar interests and similar economic structures.

[Table 2 here]

The picture becomes clearer if we pair the “borrower” treaties in Table 2 with the treaty that is the “source” of their text. Table 3 depicts the top 20 (including ties) high-overlapping treaty pairings. Interestingly, copy-and-pasting is almost always internal; that is, a country replicates its own earlier treaty text. Nearly all instances of the most rampant copy-and-pasting (85% or greater from a single treaty) fit this pattern in which a common country is present on both sides (see Table 3). This could be a power-play by the country common to both sides of the pairing. Or it may be an unintentional first mover advantage bestowed simply by having an existing PTA, which then sets a precedent for treaty negotiations. It also could be just a matter

---

17 In a few cases, both treaties in the pair are signed in the same year and thus the entry could be double-entered. In these instances, we randomly delete one of the entries.
of practical necessity or transactional efficiency. In general countries on both sides tend to be developing or lower-capacity, which is unsurprising given the countries listed in Table 2 coupled with the copy-your-own dynamic. That said, the EU and EFTA enter into Table 3 in several places, starting with the EU in the 1970s and its Norway agreement and carrying into more recent agreements with the Baltic states. Finally, we also note that many of the pairings that fall just outside of Table 3’s domain (overlap between 70-80%) do not include a common country, but instead include collections of similar states within the same region.

[Table 3 here]

Conclusions and Future Work

This paper provides some initial yet powerful evidence to resolve a fundamental tension that looms over scholarship on international institutions. Much of the literature envisions treaties as resulting from protracted negotiations and bargaining, resulting in unique treaties that are deliberately crafted. Another perspective holds that treaty-making is much less deliberate and unique, and that major portions of agreements are likely to be copy-and-pasted from existing treaties. We find strong and robust empirical support for this latter position. One next step is to dive deeper into PTA texts to determine what language is, and is not, copied. Another immediate goal is to follow up on the patterns discussed in the previous three pages by conducting bivariate and multivariate tests to determine which PTAs are most likely to be copied-and-pasted. A logical extension is to test dyadic arguments about who copies from whom, and under what circumstances.

We see several future avenues and applications beyond these first steps. We are breaking PTAs into their various articles or “subsets” (e.g., antidumping, subsidies, procurement) to see if
some parts of PTAs are more likely to be copied-and-pasted than others. It may be that the least important, or most obscure, portions of PTAs are most likely to be copied. Much future effort will be devoted to testing the pathways by which popular treaty language travels. Various WTO agreements could serve as a template, as could prominent PTAs like NAFTA. The transmission of text also might travel bilaterally, through a game of global “tag.” There also could be competing treaty models. Finally, we see great potential to apply this analytical approach to other treaty regimes, such as bilateral investment treaties (BITs) and extradition treaties. BITs are particularly ripe for study in this way, since their provisions are increasingly controversial and several countries have “model BITs” that could be tested as templates.

In sum, we view this as a truly novel methodology that allows us to examine important dynamics that are not well understood. This paper establishes that much of the language that goes into seemingly different PTAs is in fact quite similar. This revelation not only serves as a foundation for additional research, but shakes the core of how we theorize about international cooperation. Treaty-making may be less about intense bargaining and more about who will pick up a computer mouse and where they will point it.
Figure 1: Overlap between each PTA and Closest Treaty Match
(n=211 PTAs)
Figure 2: Fine-Tuned Distribution of Treaties with the Most Text Taken from Another PTA
Table 1: Sensitivity Check for Alternate Text-Match Parameters

<table>
<thead>
<tr>
<th>Match Parameters</th>
<th>Effect on Average Overlap Percentage (baseline is 5 words, include common)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 words, include common words</td>
<td>+7%</td>
</tr>
<tr>
<td>3 words, exclude common words</td>
<td>+2%</td>
</tr>
<tr>
<td>7 words, include common words</td>
<td>+1%</td>
</tr>
<tr>
<td>5 words, include common (baseline)</td>
<td>---</td>
</tr>
<tr>
<td>5 words, exclude common words</td>
<td>-4%</td>
</tr>
<tr>
<td>7 words, exclude common words</td>
<td>-6%</td>
</tr>
<tr>
<td>1 sentence, include common words</td>
<td>-10%</td>
</tr>
</tbody>
</table>
Figure 3: Amount of Text Shared with Top 3 Treaty Matches (among PTAs that copy most heavily)
<table>
<thead>
<tr>
<th>PTA</th>
<th>Year</th>
<th>% Overlap with Most Similar Treaty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania-Bosnia and Herzegovina</td>
<td>2003</td>
<td>99</td>
</tr>
<tr>
<td>Albania-Moldova</td>
<td>2003</td>
<td>99</td>
</tr>
<tr>
<td>EFTA-Estonia</td>
<td>1995</td>
<td>97</td>
</tr>
<tr>
<td>EFTA-Latvia</td>
<td>1995</td>
<td>97</td>
</tr>
<tr>
<td>Czech Republic-Turkey</td>
<td>1997</td>
<td>96</td>
</tr>
<tr>
<td>Slovakia-Turkey</td>
<td>1997</td>
<td>96</td>
</tr>
<tr>
<td>Lithuania-Slovakia</td>
<td>1996</td>
<td>94</td>
</tr>
<tr>
<td>Azerbaijan-Ukraine</td>
<td>1995</td>
<td>93</td>
</tr>
<tr>
<td>EC-Norway</td>
<td>1973</td>
<td>93</td>
</tr>
<tr>
<td>Latvia-Slovakia</td>
<td>1996</td>
<td>93</td>
</tr>
<tr>
<td>Albania-Romania</td>
<td>2003</td>
<td>92</td>
</tr>
<tr>
<td>EC-Latvia</td>
<td>1994</td>
<td>92</td>
</tr>
<tr>
<td>EC-Lithuania</td>
<td>1994</td>
<td>92</td>
</tr>
<tr>
<td>Latvia-Poland</td>
<td>1997</td>
<td>92</td>
</tr>
<tr>
<td>Macedonia-Romania</td>
<td>2003</td>
<td>92</td>
</tr>
<tr>
<td>Armenia-Turkmenistan</td>
<td>1995</td>
<td>91</td>
</tr>
<tr>
<td>Czech Republic-Estonia</td>
<td>1996</td>
<td>91</td>
</tr>
<tr>
<td>Czech Republic-Slovenia</td>
<td>1993</td>
<td>91</td>
</tr>
<tr>
<td>Estonia-Slovakia</td>
<td>1996</td>
<td>91</td>
</tr>
<tr>
<td>Slovakia-Slovenia</td>
<td>1993</td>
<td>91</td>
</tr>
<tr>
<td>Australia-Papua New Guinea</td>
<td>1991</td>
<td>90</td>
</tr>
<tr>
<td>Estonia-Slovenia</td>
<td>1996</td>
<td>90</td>
</tr>
<tr>
<td>Montenegro-Turkey</td>
<td>2008</td>
<td>90</td>
</tr>
<tr>
<td>Albania-EFTA</td>
<td>2009</td>
<td>89</td>
</tr>
<tr>
<td>Bulgaria-EFTA</td>
<td>1993</td>
<td>89</td>
</tr>
<tr>
<td>EFTA-Serbia</td>
<td>2009</td>
<td>89</td>
</tr>
<tr>
<td>Bulgaria-Finland</td>
<td>1974</td>
<td>88</td>
</tr>
<tr>
<td>Czechoslovakia-Finland</td>
<td>1974</td>
<td>88</td>
</tr>
<tr>
<td>Israel-Slovenia</td>
<td>1998</td>
<td>88</td>
</tr>
<tr>
<td>Lithuania-Poland</td>
<td>1996</td>
<td>88</td>
</tr>
<tr>
<td>Lithuania-Slovenia</td>
<td>1996</td>
<td>88</td>
</tr>
</tbody>
</table>
Table 3: PTA Pairs with the Greatest Text Overlap

<table>
<thead>
<tr>
<th>First PTA in Pair</th>
<th>Year</th>
<th>Second PTA in Pair</th>
<th>Year</th>
<th>% Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania-Bosnia and Herzegovina</td>
<td>2003</td>
<td>Albania-Moldova</td>
<td>2003</td>
<td>99</td>
</tr>
<tr>
<td>Slovakia-Turkey</td>
<td>1997</td>
<td>Czech Republic-Turkey</td>
<td>1997</td>
<td>96</td>
</tr>
<tr>
<td>Lithuania-Slovakia</td>
<td>1996</td>
<td>Czech Republic-Lithuania</td>
<td>1995</td>
<td>94</td>
</tr>
<tr>
<td>Azerbaijan-Ukraine</td>
<td>1995</td>
<td>Turkmenistan-Ukraine</td>
<td>1994</td>
<td>93</td>
</tr>
<tr>
<td>EC-Norway</td>
<td>1973</td>
<td>EC-Switzerland/Liechtenstein</td>
<td>1972</td>
<td>93</td>
</tr>
<tr>
<td>Latvia-Slovakia</td>
<td>1996</td>
<td>Lithuania-Slovakia</td>
<td>1996</td>
<td>93</td>
</tr>
<tr>
<td>Albania-Romania</td>
<td>2003</td>
<td>Macedonia-Romania</td>
<td>2003</td>
<td>92</td>
</tr>
<tr>
<td>EC-Latvia</td>
<td>1994</td>
<td>EC-Lithuania</td>
<td>1994</td>
<td>92</td>
</tr>
<tr>
<td>Latvia-Poland</td>
<td>1997</td>
<td>Lithuania-Poland</td>
<td>1996</td>
<td>92</td>
</tr>
<tr>
<td>Macedonia-Romania</td>
<td>2003</td>
<td>Albania-Romania</td>
<td>2003</td>
<td>92</td>
</tr>
<tr>
<td>Armenia-Turkmenistan</td>
<td>1995</td>
<td>Armenia-Moldova</td>
<td>1993</td>
<td>91</td>
</tr>
<tr>
<td>Estonia-Slovakia</td>
<td>1996</td>
<td>Czech Republic-Estonia</td>
<td>1996</td>
<td>91</td>
</tr>
<tr>
<td>Slovakia-Slovenia</td>
<td>1993</td>
<td>Czech Republic-Slovenia</td>
<td>1993</td>
<td>91</td>
</tr>
<tr>
<td>Australia-Papua New Guinea</td>
<td>1991</td>
<td>Australia-Papua New Guinea</td>
<td>1976</td>
<td>90</td>
</tr>
<tr>
<td>Estonia-Slovenia</td>
<td>1996</td>
<td>Estonia-Slovenia</td>
<td>1996</td>
<td>90</td>
</tr>
<tr>
<td>Montenegro-Turkey</td>
<td>2008</td>
<td>Albania-Turkey</td>
<td>2006</td>
<td>90</td>
</tr>
<tr>
<td>Albania-EFTA</td>
<td>2009</td>
<td>EFTA-Serbia</td>
<td>2009</td>
<td>89</td>
</tr>
<tr>
<td>Bulgaria-EFTA</td>
<td>1993</td>
<td>EFTA-Romania</td>
<td>1992</td>
<td>89</td>
</tr>
<tr>
<td>Czechoslovakia-Finland</td>
<td>1974</td>
<td>Bulgaria-Finland</td>
<td>1974</td>
<td>88</td>
</tr>
<tr>
<td>Israel-Slovenia</td>
<td>1998</td>
<td>Israel-Slovakia</td>
<td>1996</td>
<td>88</td>
</tr>
<tr>
<td>Lithuania-Slovenia</td>
<td>1996</td>
<td>Lithuania-Poland</td>
<td>1996</td>
<td>88</td>
</tr>
</tbody>
</table>
Bibliography


