How withdrawal provisions affect multilateral environmental agreement signature preferences

Claire Peacock,[†*] Jean-Frédéric Morin,[*] and Benjamin Tremblay-Auger[*]
This version: Jan 6, 2020

Abstract:

International environmental cooperation is governed by an increasing number of multilateral environmental agreements (MEAs). Despite these agreements' important role in mitigating challenges ranging from climate cooperation to increasing pollution levels, many MEAs struggle to attract and retain sufficient numbers of participants as their benefits cannot easily be contained to members. Existing research on international treaty design theorizes that the inclusion and proper design of withdrawal provisions in international treaties—i.e. their design content—may be one way in which participation incentives can be increased. We further suggest that in addition to design content, design context should also matter. Design context is provided by a state's past experience with and exposure to a particular design feature within the same regime complex. We argue that states are incentivized to prefer familiar design solutions for reasons related to validation and ingrained biases. Using MEA signature and withdrawal provision data spanning from 2004 to 2014, we show that states are more likely to sign an MEA not only based on the design content of its withdrawal provisions, but also the design context of the agreement's withdrawal provisions.

[†] Department of Political Science, Simon Fraser University
[*] ESEI, Université Laval

1 Introduction

Multilateral environmental agreements (MEAs) are required to manage many of the world's most pressing ecological problems, ranging from climate change to marine pollution to mass species extinction events. Yet, in international environmental cooperation states face elevated incentives to free-ride off of the virtuous behavior of others (Barrett 1994). The result is that MEAs, while growing in number, have struggled to attract and retain sufficient membership to meet their stated goals. Previous studies theorize however that free-riding incentives[1] may be partially mitigated by the inclusion of withdrawal provisions in international treaties (Helfer 2005:1599; Bilder 1981:52-55), and especially ones with relatively long notification periods (Koremenos & Nau 2010:97). These provisions should promote treaty participation in two ways. First, withdrawal provisions provide a state with insurance that it can voluntarily leave its international commitments[2] should other members leave or the external environment change.[3] Second, they should also reduce the prospects of sudden opportunistic behavior by others by mandating relatively long wait periods before a state can withdraw after it gives notice. Yet, we know little of the relationship between withdrawal provisions and MEA participation decisions in practice. Does the inclusion of withdrawal provisions with particular notification periods increase MEA signatories?

[1] Free-riding incentives related to public goods problems are commonly characterized as enforcement problems in the literature on international institutional design in that "[a]n enforcement problem is present when actors have incentives to defect from cooperation" (Koremenos and Nau 2010, 95).
[2] In line with this rationale, Article 50 of the Treaty of the European Union—the EU's withdrawal clause—was negotiated to convey the voluntary nature of the Union, with the intention to make it clear that the EU was not the "incipient superstate of Eurosceptic nightmares" (Kerr, 2017).
[3] Huysmans (2019) shows that prior to the EU's 2004 Eastern accession, new members, supported by Britain and Denmark, campaigned for the adoption of Article 50 by the European Union as a safeguard against undesired future policy changes.

The proliferation of MEAs and also treaties more generally since 1945 has facilitated the creation of a rich body of research on treaty design and its rational determinants (Koremenos et al. 2001; Rosendorff and Milner 2001), as well as related work on how treaty content impacts new treaty commitment decisions (Mohrenberg et al. 2019; Spilker & Koubi 2016). In exploring how withdrawal provisions affect MEA signature decisions, our paper builds on this second category of studies.[4] Yet, it is increasingly clear that new treaty signing decisions are not made in isolation. With over 600 MEAs in force by 2014, states now are equipped with a large catalogue of prior treaties situated within the environmental regime complex and knowledge of their design features and this experience is likely to impact their new treaty signature decisions. In this paper, we therefore also highlight a missing element: the importance of the treaty design context in informing its new treaty signature decisions. Treaty design context is provided by a state's past experience with and exposure to a particular design feature within the same regime complex. In this way, our work aligns with recent context-focused studies that have shown how prior cooperation amongst treaty members, and previously used legal language and substantive content are additional important determinants of treaty design outcomes in trade agreements and beyond (Copelovitch and Putnam 2014, Kim & Manger 2017; Baccini et al. 2015; Peacock et al. 2019).

We argue that states are not only more likely to sign new MEAs that have withdrawal provisions with relatively long notification periods but also those with withdrawal clauses that have similar formulations to those it has agreed to in the past. In this way, we suggest that both design content

[4] We primarily focus on treaty signature rather than ratification decisions as they are the first formal action that represents a state's intent to be legally bound by a treaty (see: Article 12 of the Vienna Convention on the Law of Treaties). Numerous political barriers may prevent a state from later ratifying a treaty despite its expressed intent to do so. We however analyze ratification decisions as a robustness check.

and design context are relevant to MEA signature decisions. Our design context explanation builds on the observation that within a regime complex, the same experts are often advisers on multiple treaty signature decisions and have rational and behavioral incentives to prefer familiar design choices—even when better solutions exist. By giving new treaties that incorporate previously deployed design solutions their approval, experts validate past design decisions. Yet, they may also prefer familiar solutions for ingrained cognitive reasons. Decades of behavioral psychology research has shown that with increased exposure to a stimulus, individuals develop a familiarity bias towards that stimulus—i.e. the "mere exposure effect: (Zajonc 1980; Zajonc 2001; Bornstein 1989).

We base our analysis on withdrawal provisions in MEAs for three reasons beyond their theorized potential to boost participation in MEAs. First, withdrawal provisions are a common procedural design feature, central to early studies in the rational design of international institutions tradition and related work (Koremenos 2001; 2005; Helfer 2005) and are included in most international treaties, facilitating cross-treaty comparison. Second, as MEAs face elevated free-riding incentives compared to those present in most other areas of international cooperation, such as international trade, or investment, if the inclusion of withdrawal provisions in international treaties increases signatories, MEAs provide a clear case to test this assertion. Finally, MEAs are sufficiently numerous to facilitate the statistical testing of our argument. The MEAs we analyze by definition involve three or more parties. By focusing on multilateral rather than bilateral treaty signature decisions, we limit our attention to circumstances in which individual states are less likely to have a veto over specific design choices during negotiations and must make their signature decisions based on the treaty design features agreed upon by negotiating parties as a whole.

Using data on withdrawal provisions in 619 MEAs and their membership between 2004 and 2014 primarily, we investigate how past experience with withdrawal provision design choices as well as the design of withdrawal provisions affects a state's new treaty signature decisions. We therefore explore both design context and design content. We also use data from the 1994 to 2004 time period to evaluate how consistent our findings are across time periods. Our MEA data covers topics as diverse as pollution, climate change, fisheries, conservation, and freshwater, offering a broader view of the global environmental governance regime complex than most previous studies. We use longitudinal statistical network analysis to test our design context argument. Network analytical techniques are well suited to studies that assess membership or affiliation decisions as they can be readily specified to account for different forms of interdependence that characterize actors' decision making over time (e.g. see Koskinen & Edling 2012). We find that states are more likely to sign MEAs with withdrawal provisions with relatively longer notification periods—though these agreements are few and far between. However, our findings show that within the environmental regime complex, a state's prior direct exposure to a given withdrawal clause type (gained through signing more previous treaties with that withdrawal clause type), also increases its likelihood of signing a new MEA with the same type of clause.

2 Design Content and Design Context

States have now signed tens of thousands of treaties and continue to add new treaties each year (Mackenzie 2010; Raustiala 2010). These treaties cluster into increasingly dense, issue-based, regime complexes, each comprised of nested and overlapping agreements (Raustiala & Victor

2003; Henning & Pratt 2019). These conditions have created an environment in which expert advisers, including a state's legal team, senior civil servants, and other policy officials, play a central role in helping states to navigate new treaty commitment decisions. These experts generally work on multiple treaties within a given regime complex over the course of their careers. This creates a continuity in expertise across treaties and an environment in which individual experts are able to accumulate detailed knowledge of and exposure to past treaty design choices—especially concerning standard procedural provisions which routinely appear across most treaties within a regime complex, such as withdrawal provisions. While it is clear that design content should influence treaty commitment decisions, as regime complexes grow, design context provided by prior design choices made by a state and others states within the regime complex, may therefore also become increasingly important.

Existing research on treaty design largely focuses on the rational determinants behind design content (e.g. Koremenos et al 2001; Rosendorff and Milner 2001) and its subsequent effectiveness (e.g. Mitchell 2006). Yet this prior scholarship also offers a number of specific theoretical predictions about how design content is relative to signature decisions. Relevant to this project, are the related predictions that withdrawal provisions should boost membership in international treaties and that in cases where states face incentives to renege on cooperation—or potentially bypass it altogether—relatively long notification periods are most relevant (Koremenos & Nau 2010).[5] If this is the case, withdrawal provisions with relatively long notification periods should increase MEA signatories. In contrast, in other issue areas where free-riding incentives are not as

[5] In their study of 86 international agreements, 24 of which were environmental agreements, Koremenos and Nau (2010) show that underlying enforcement problems are more likely to be accompanied by longer withdrawal notification periods.

prevalent, shorter notification periods, which offer more flexibility to states, may be preferable. As these theoretical predictions are already well-developed in the literature, we focus the remainder of our attention in this section on the subject of design context.

That a state's own experiences inform its future decision-making is unsurprising. Previous research for example has shown that the related concepts of path dependence and habit affect state decision-making (Pierson 2000; Denemark & Hoffmann 2008; Hopf 2010). However, how design context produces familiarity bias—which may in turn lead to path dependent behavior or the development of new habits—has not been directly investigated.

In the area of environmental cooperation, we expect treaty design context to influence new signature decisions in two primary ways related to individual experts employed by states,[6] which we abbreviate as the "validation rationale" and the "ingrained bias rationale". Beginning with the validation rationale, by giving new treaties with previously deployed design solutions their approval, individual experts within states validate their own and their state's past decisions.[7] Maintaining consistency across their state's agreements helps these individuals to assert their authority as experts and to conceal their uncertainty. Thus, experts may benefit professionally from supporting the adoption of MEAs with familiar design solutions that a state has agreed to in the

[6] A focus on individual decisionmakers and their incentives presents International Relations researchers with two options with regards to the field's aggregation problem. 1. These approaches can be applied to states directly, glossing over potential applicability issues or 2. instead to the individuals that comprise state decision making structures, understanding collective state decision-making as the aggregation of individual decision-making (Powell 2017; Hafner-Burton et al. 2017). We adopt the latter approach.
[7] An exception is when familiar provisions have led to undesirable outcomes. When this occurs, a state is likely to invest time in exploring alternative options. For example, states that have faced arbitration under investor-state-dispute-settlement clauses of bilateral investment treaties have been shown to be more likely to terminate or renegotiate the relevant provisions in these treaties (Haftel &Thompson 2018; Thompson et al 2019).

past. When this occurs, direct previous experiences of experts provide context for a state's future decision-making and leads to a preference for familiar solutions.

Second, individual experts may have an ingrained bias towards familiar design content. Decades of behavioral research has shown that with repeated exposure, individuals develop a preference for familiar options over alternatives (Zajonc 1980; 2001). This is called the "mere exposure effect". While over the past two decades, International Relations as a discipline has exhibited renewed interest in the application of behavioral psychology and economics to international phenomena (see: Hafner-Burton et al. 2017; Kertzer & Tingley 2018), these efforts have remained largely separate from research related to international institutional design and international regime complexes. In the international treaty context, exposure to a specific provision could feasibly be gained through prior direct experiences with that clause. When expert advisers are repeatedly exposed to a given treaty design feature, this effect is likely to lead them to develop a bias towards these familiar solutions. At the state-level, these individual preferences are aggregated, with the net effect of states being more likely to sign treaties with familiar standard features. We expect the mere exposure effect to be most relevant in situations where costs are minimal, such as is the case with many standard procedural design features, such as withdrawal provisions. Due to scarce resources, high-cost decisions in contrast are likely to lead to increased deliberation and analysis and be less susceptible to ingrained biases.

Just as the exposure effects in individuals documented by behavioral psychologists can occur as a result of both conscious and unconscious stimuli, in the context of inter-state cooperation, exposure effects can conceivably operate in two ways. First, individual experts within states may be aware of their exposure to particular treaty design choices. In this case, they may rationalize their preference for familiar solutions as being based on their knowledge of the solution and how it works or using other justifications even if their preference was formed through the mere exposure effect. Second, individuals are also known to develop positive affect towards familiar stimuli unconsciously (Kahneman 2011:67). Individual decision makers within states are subject to this unconscious form of familiarity bias. They may feel that a treaty "looks about right" or "looks like it should" without being able to pinpoint why. Whether consciously or not, familiar solutions should be more popular and conducive to treaty signature decisions, in some cases, even when not strategically optimal.

While we cannot isolate the validation and ingrained biased rationales empirically, they share similarities in that each is rooted in the idea that design context matters and that individuals involved in treaty signature decisions face incentives to pay attention to that context. In addition, both rationales lead to the similar prediction of states preferring to sign MEAs with familiar withdrawal clauses over alternatives. Yet, past experience with and exposure to particular standard design choices may of course also lead states to sign treaties with consistent design features for other reasons.

Two relevant alternative explanations are costs[8] and socialization—each best suited to different cases. Cost-based explanations are most plausible when being bound by consistent rules across treaties has strong potential to yield clear monetary or other dividends for states. Yet, the costs

[8] Costs in this context include adapting national legislation to be consistent with treaty obligations, implementing new required legislation, and updating in-house legal and policy-related expertise to reflect new treaty obligations.

associated with signing treaties with different standard design features, such as exit rules, are often minimal. Signing a treaty with a 24-month withdrawal notification period does not incur a great deal of additional costs for a state over signing a treaty with a 12-month notification period. Moreover, states do not always fulfill their treaty obligations even before they formally withdraw (Koskenniemi 1992). Socialization effects in turn are likely to be identifiable by empirical patterns of increasing conformity amongst states driven by pressures to fit in with prevailing norms. Yet, withdrawal clauses in MEAs exhibit a great deal of variation in their notification period lengths and offer no clear patterns of overall convergence. This makes it unlikely that the socialization rationale is useful at explaining this case. In addition, there is also currently only limited empirical evidence that states look to other states' treaty design decisions more generally within a regime complex and in this indirect manner, develop a preference for similar design solutions. A few notable exceptions are: the diffusion of design features from the European Court of Justice to eleven regional courts in Africa and Latin America (Alter 2012) and the adoption of EU-style institutions by ASEAN member states (Jetschke & Murray 2012). We are however able to evaluate this explanation empirically and do so in our analysis.

A number of recent empirical examples align with our expectations that states prefer to sign treaties with familiar content, whether or not it is optimal. For example, over the past few decades, states have ratified a large number of bilateral investment treaties with suboptimal but popular forms of investor-state dispute settlement clauses (Poulsen 2014). They also have routinely agreed to international financial contracts containing a standard *pari passu* clause; the meaning of which is largely unknown, and which exposes them to the risk of litigation and erroneous legal interpretations (Gulati & Scott 2012).[9] States also replicate their preferential trade agreements' now standard environmental clauses despite their uncertain consequences (Morin et al. 2019).

Using the case of withdrawal clauses in MEAs, the discussion on design context above leads to the following two testable hypotheses:

- H1: An MEA is more likely to be signed by a state when it has a withdrawal provisions with a relatively long notification period.
- H2: An MEA is more likely to be signed by a state the more prior MEAs with similar withdrawal clauses that specific state has signed in the past.

Hypothesis 1 relates to design content. It suggests that withdrawal provisions that address the underlying enforcement problem found in MEAs are likely to boost MEA signatories. The second hypothesis relates to design context, suggesting that greater amounts of prior direct experience with and exposure to a given withdrawal clause formulation increases the likelihood that a state will sign a new institution with a similar formulation.

[9] *Pari passu* means, "in equal step". *Pari passu* clauses have been included in financial contracts for over a century. See Gulati & Scott (2012).

3 Withdrawal Provisions in International Treaties

Withdrawal provisions are a standard design feature included in most international treaties, from human rights treaties to environmental treaties to tax treaties. They are a common form of flexibility mechanism, a category that also includes escape clauses and treaty reservations. Flexibility provisions, including withdrawal clauses, make international commitments less rigid (Koremenos et al. 2001). Yet, "[f]lexibility provisions are not simply chosen as a set; nor do particular pairs go together" (Koremenos & Nau 2010)—each serves a unique purpose. Even withdrawal provisions come in different forms, some allowing states to more quickly and easily terminate their treaty commitments than others. In this section, we provide a basic overview of withdrawal provisions' and how they vary across treaties.[10]

Withdrawal clauses permit member states to unilaterally and permanently leave a treaty without formally breaching their treaty-governed obligations. In that respect, withdrawal provisions differ from other flexibility measures, such as sunset clauses, which permanently govern the termination of a treaty for all members; escape clauses, which allow individual members to temporarily not be bound by specific commitments while remaining parties to the treaty; and amendment provisions, which involve states collectively and often permanently updating a treaty while remaining member. While all flexibility mechanisms may affect treaty signature decisions by reducing the constraints placed on affiliated states, withdrawal provisions offer states more policy autonomy than other flexibility mechanisms, being able to be invoked unilaterally, with permanent effect,

[10] See Helfer 2005; 2012; Curtis & Gulati 2010; and Tobin 1933 for more detailed overviews of withdrawal clauses.

and giving states an outside option. If standard flexibility features do impact treaty signature decisions, this effect is therefore likely to be clearest in withdrawal provisions, making them a logical focus for our study beyond the rationales offered in the introduction.

Withdrawal provision formulations vary across treaties within regime complexes. During treaty negotiation, the form that withdrawal clauses take is governed by state consent. Negotiating parties have the freedom to design withdrawal clauses as they see fit, given that other negotiating parties can reach an agreement—a potential stumbling block. Once a treaty text is formalized, states need not sign or accede to treaties that contain provisions that they find unacceptable. The result is that "[s]tates are the undisputed masters of treaty exit rules" (Helfer, 2012) at the initiation of a new treaty and also of treaty signature decisions after the agreement has been negotiated.

The withdrawal provisions that states negotiate and agree to be bound by often include two main components. First, they sometimes provide a timeframe after an agreement has entered into force during which states are not permitted to renounce their membership commitments. Second, withdrawal provisions usually outline a mandatory "cooling-off" period after providing notice of intent to withdraw, before treaty exit occurs. For example, Article 27 of the Kyoto Protocol reads:

> 1. At any time after three years from the date on which this Protocol has entered into force for a Party, that Party may withdraw from this Protocol by giving written notification to the Depositary.

> 2. Any such withdrawal shall take effect upon expiry of one year from the date of receipt by the Depositary of the notification of withdrawal, or on such later date as may be specified in the notification of withdrawal. [...]

The first paragraph of Article 27 outlines the mandatory wait period before notification of withdrawal can be given; whereas, paragraph two provides the required notification period length after formally submitting its intent to withdraw—in this case one year—before a state's withdrawal officially occurs. In 2011, Canada used the mechanisms outlined in Article 27 to give its mandated one-year notice of its intent to terminate its treaty commitments under the Kyoto Protocol, the first country to do so under that agreement.

Our analytical focus is primarily on the length of notification provisions, the second common component of withdrawal clauses, rather than the wait periods directly following negotiations during which withdrawal cannot occur. We expect the duration of notification provisions to be more relevant to treaty signature decisions for two reasons. First, the duration of notification provisions provides a greater constraint on states' future policy autonomy than initial periods after negotiation during which states cannot terminate their commitments, as it is rare for states to decide to leave treaties directly after signature. Logically, states are more likely to desire to terminate their treaty membership as their strategic environment changes over a longer timeframe. In addition, the length of a notification period may impact the viability of re-negotiation or exit negotiation initiatives by restricting its length. For example, Article 50 of the Treaty on European Union gives EU member states up to two years after giving notice of their intent to withdraw to formalize their leaving negotiations. The outcome of Brexit negotiations hinge on Britain and the

European Union's ability to finalize their talks within the two-year timeframe—or an extension thereof. This makes withdrawal provision notification period lengths an important aspect of international treaties.

4 Research Design

Data on withdrawal clauses

We created an original dataset of MEA withdrawal clauses. We first collected the full text of 619 MEAs signed between 1994 and 2014 from the International Environmental Agreements Database Project (Mitchell 2018). Then, we relied on human coding to collect information on their withdrawal clause notification period lengths. Using this data, we first coded the specific number of months listed as official the notification period length. We then also categorized treaties that contain explicit denunciation notification periods into three categories: treaties with a notification period that is 1) less than 12 months, 2) exactly 12 months, and 3) greater than 12 months. Our focus is therefore on the broad similarity of content rather than on the exact similarity of the language used across MEAs. Our categories are relevant as a notification period of 12 months is the length of notification period outlined in Article 56(2) of the Vienna Convention on the Law of Treaty to be used when no withdrawal information is outlined in the treaty and withdrawal is not deemed implicitly prohibited. As such, it can be viewed as a baseline notification period. Notification periods of less than 12 months are less stringent than the baseline, as they allow for quicker treaty exit, and those of greater than 12 months are more stringent than the baseline.

The resulting dataset reveals that withdrawal notification period lengths vary widely, as depicted in Figure 1. Just over a quarter of MEAs in our dataset allow for a rapid exit with only a few months' official warning; whereas around 34 percent require exactly 12 months notice of a state's intent to withdraw. Despite the theoretical justification for having relatively long notification periods, only around two percent of MEAs included a withdrawal notification period of greater than 12 months. Approximately 36 percent of treaties neglect to include any notification period at all, which may preclude treaty withdrawal altogether.[11]

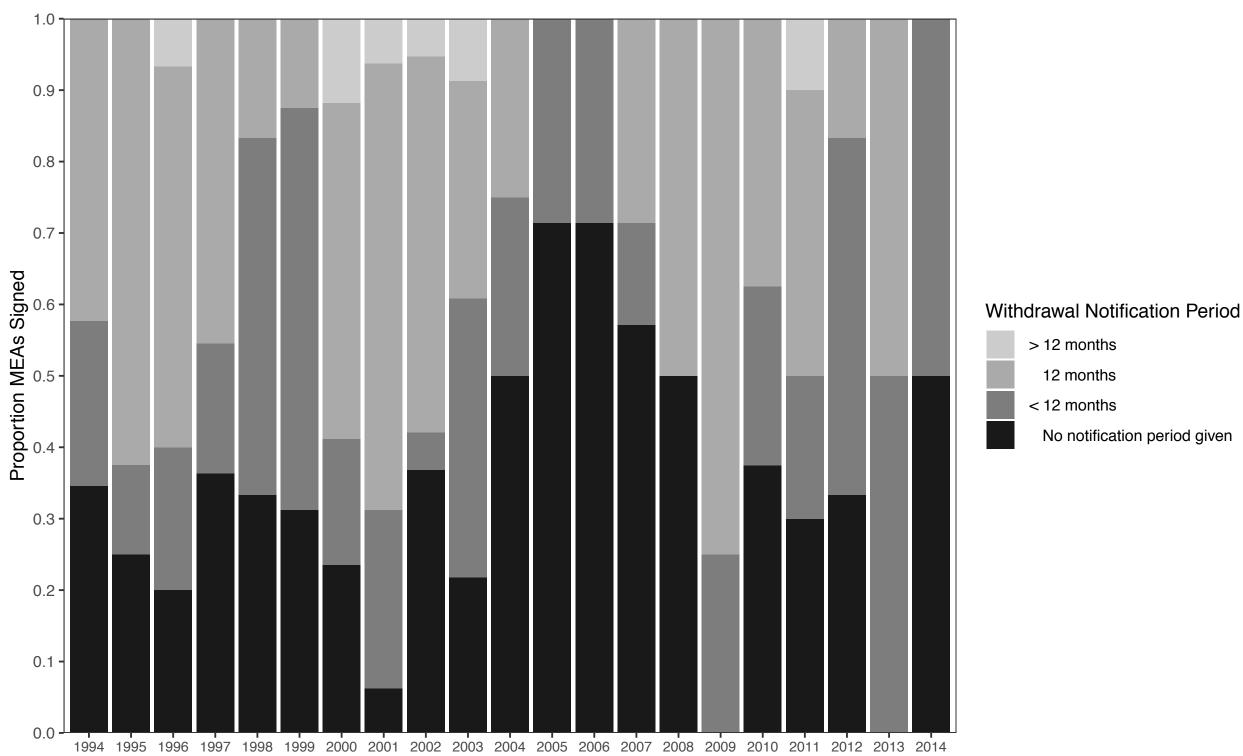


Figure 1: Proportion of MEAs with each withdrawal notification period length

Network analysis

[11] See article 56(1) of the Vienna Convention on the Law of Treaties.

We analyze our MEA withdrawal clause and signature data using longitudinal network analysis, which offers a number of advantages over traditional dyadic regression frameworks. Network analytical techniques are able to account for higher-order dependencies, e.g. if having a shared member increases the likelihood of forming a MEA signature tie or if future tie formation is dependent on the past popularity of an MEA node. Previous studies have shown that states' membership decisions in international treaties exhibit various forms of interdependence (Manger et al. 2012; Hollway & Koskinen 2016; Kinne 2013), making network analysis a particularly relevant solution for analyzing our MEA data.[12] Failure to account for such dependencies violates the independence assumption of traditional regression techniques, which when not addressed can lead to inefficient estimates. Network analysis is also able to incorporate monadic (i.e. state-level and MEA level) and dyadic (i.e. MEA signature tie-level) effects.

We use stochastic actor-oriented models (SAOMs) (see: Snijders et al. 2010), an established form of longitudinal network analysis. SAOMs are a probabilistic model for network dynamics. Actors' are assumed to form ties (i.e. here representing signature decisions) in a utility-maximizing manner—though subject to experience-related dependencies. SAOMS are based on Markov chains: future actions are assumed to be based on the current state of the network rather than all previous states of the network. However, as each new state of the network accounts for the state directly preceding it and therefore the network's evolution up until that time point, this assumption is not overly limiting in practice. Ties are modelled as forming one at a time, depending on the

[12] Spatial analysis techniques are an alternative method also suited to the analysis of interdependent data. However, they do not allow researchers the same flexibility in assessing the types of interdependence present in the data.

current network configuration. The interpretation of SAOM estimates is similar to multinomial logistic regression, with transformed estimated coefficients being interpretable as odds ratios, making them readily interpretable.

In practice, the networks used in SAOMs are modelled as waves, which are snapshots of a given network at a specific point in time. A period is the time between two waves. Our main model covers the time period from 2004 to 2014, though we use data from the 1994 to 2004 time period as a robustness check, reported in Appendix B.[13] However, all MEA signature information for agreements that remain in force during our from 1945 onwards is included in the initial snapshots of the network.

Our MEA membership network is bipartite in nature. Bipartite networks are made up of two distinct types of actors or "nodes". Actors from one node set can only form ties with actors from the other node set.[14] The two actors present in the MEA network are states and MEAs, with ties between them representing MEA signature decisions. Other examples of bipartite networks are citation networks linking researchers and papers and customer-merchant networks, linking individuals with the places where they purchase goods and services. SAOMs have previously been successfully applied to the study of bipartite networks (Koskinen & Edling 2012; Snijders et al. 2013; Fujimoto et al. 2018; Milewicz et al. 2018), making them a logical choice for our purposes.

[13] To allow heterogeneity in effects over time, we model our two periods of data separately, to avoid including time dummies for each period and to facilitate interpretation (Lospinoso et al. 2011).
[14] Bilateral environmental agreements are not bipartite in nature as they have only one node-type, i.e. state nodes, making their structure different.

The bipartite nature of our MEA signature network is illustrated in Figure 2 below, which depicts four sub-networks taken from our sample: A) the North American Agreement on Environmental Cooperation (NAAEC), B) the International Convention for the Regulation of Whaling (ICRW) and the International Convention for the High Seas Fisheries of the North Pacific (ICHSF) C) the Kyoto Protocol, and D) the United States' ego network of MEA signature ties. Figure 2 also highlights three network structural properties that can contribute to interdependence in our bipartite network: treaty popularity (i.e. membership size) in subplots A) and C), the presence of shared memberships in subplot B)—i.e. four-cycles in network terms, and an individual state's tie-formation activity in subplot C).

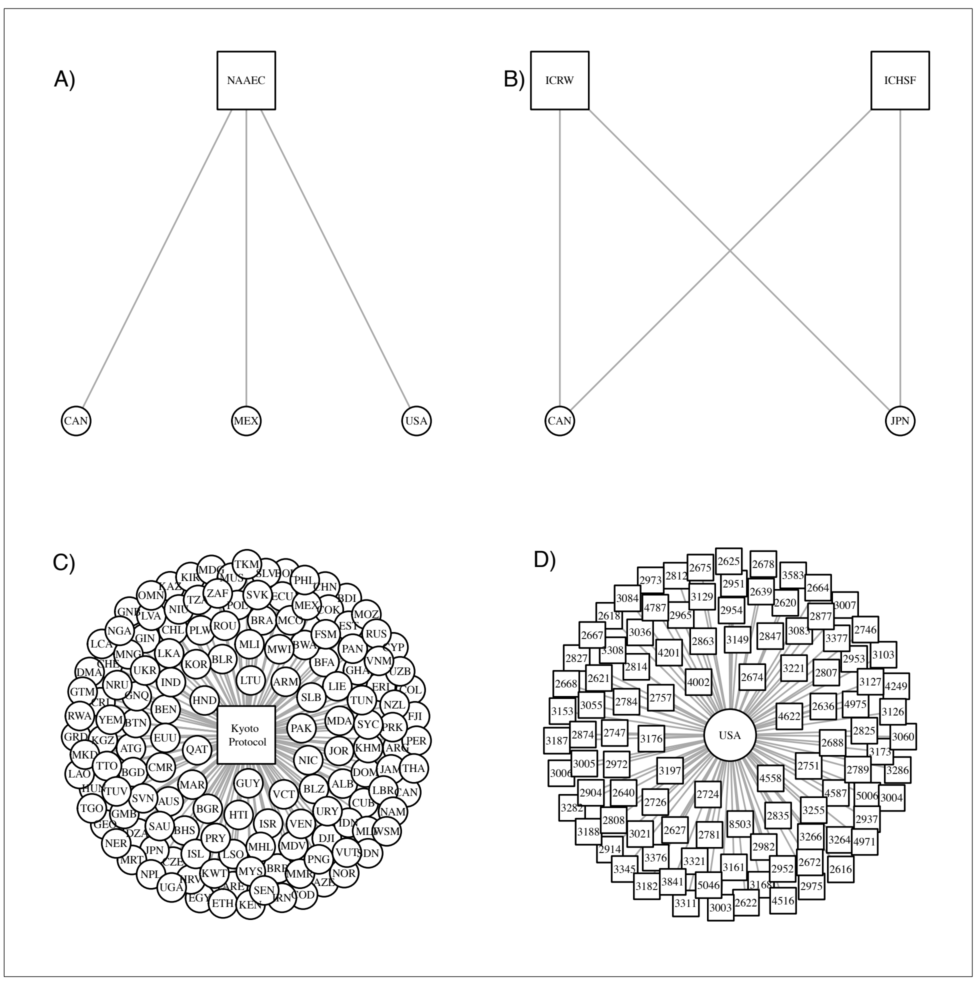


Figure 2: Four Sample MEAs

In Figure 2, circular nodes represent states, square nodes represent MEAs, and ties represent signature decisions.[15] Varying numbers of ties around MEA nodes therefore indicate varying levels of MEA popularity, with the NAAEC (subplot A),[16] having been ratified by three states, compared to the Kyoto Protocol's 146 state members by 2014 (subplot C). Subplot B) of Figure 2 provides an illustration of a four-cycle (i.e. shared membership), amongst Canada, Japan, the ICRW and the ICHSF.[17] State nodes within the MEA network also are characterized by unique signature activity levels. To illustrate tie-formation activity, we include the US' MEA signature ties in subplot D) of Figure 2.

	2004	**2014**
Density	0.039	0.049
Avg. Signature Ties/State	24.445	30.118
Total MEA Signature Ties	5158	6355

Table 1: MEA Network Descriptive Statistics by Wave

While individual MEAs, such as the Kyoto Protocol, have been ratified by a large number of states, overall, the MEA network is sparse, as shown in Table 1. The density—i.e. the number of ties in a network divided by the possible number of ties—of the MEA network ranges from 0.039 in to 0.049 in 2014. The low density of this network indicates that states only sign a small number of the MEAs that they could sign. Within the network, states on average had ratified around 24

[15] The numbers in the MEA nodes in subplot D) are the MEA identification codes found in the underlying data.
[16] The NAAEC was only open to the three member states of NAFTA.
[17] Other members of these two agreements have been excluded for visualization purposes.

MEAs in 2004 and just over 30 in 2014. By 2014, the network had 6335 ties overall, spread across 619 MEAs.

Dependent Variable

Our dependent variable is a matrix of signature ties linking states with MEAs.[18] The data for this variable is from the International Environmental Agreements Database Project (Mitchell 2018). When a given state signs an MEA, MEA SIGNATURE is coded 1 and 0 otherwise. Signature decisions commit states to refrain "from acts which would defeat the object and purpose of a treaty"—unless a state makes its "intention clear not to become a party to the treaty" (Article 18, VCLT). They are a signal of a state's intent to be legally bound by a treaty and as such are relevant indicators of a state's desire to participate in a treaty.

Independent Variables

Our first set of independent variables are measured at the MEA-level and capture withdrawal clause notification period length and relate to Hypothesis 1. According to rational design research, longer notification periods in particular are preferred by states in areas suffering from enforcement problems, where individual states face increased incentives to defect from continued cooperation (Koremenos & Nau 2010), such as in the case of MEAs, which are subject to free-riding incentives. These variables are grouped into four mutually exclusive types: MEA NOTIFICATION PERIOD >12,

[18] An alternate modelling strategy would be to divide our sample into four dependent variable networks, with one network corresponding to each withdrawal notification period category. We tested this strategy but insufficient data remained within each network for the proper functioning of the algorithm.

MEA NOTIFICATION PERIOD <12, MEA NOTIFICATION PERIOD =12, and NO MEA NOTIFICATION PERIOD. Each variable is coded 1 when a notification period falls under that category and 0 otherwise. We exclude MEA NOTIFICATION PERIOD =12 as a reference category as it is the baseline notification period length outlined in the Vienna Convention on the Law of Treaties from which states may choose to depart and also is the most common notification period length in our MEA data.[19] Positive and statistically significant coefficients for these variables indicate that a state is more likely to sign an MEA that has a withdrawal notification period of that particular type as compared with MEAs with withdrawal notification periods that are 12 months long.

Our two main design context-related independent variables are measured at the dyad-level, i.e. tie-level, and capture a state's exposure to a given withdrawal clause formulation. The first, STATE MEAS WITH SAME NOTIFICATION PERIOD (SQRT), is a count variable. It is measured as the square root of the number of MEAs a state has previously signed that have the same withdrawal notification period length measured in months as the MEA in question. A positive and statistically significant coefficient for STATE MEAS WITH SAME NOTIFICATION PERIOD (SQRT) would indicate that a state is more likely to sign an MEA with a particular notification period length the more times it has signed other MEAs with that notification period length, all else equal. We also include an alternative formulation of this variable, PROPORTION STATE MEAS WITH SAME NOTIFICATION PERIOD. These two variables correspond to Hypothesis 2.

Controls

[19] When a treaty does not include a notification period, while withdrawal is occasionally precluded altogether (Helfer 2005:1582), in other cases, there is a 12-month withdrawal notification period as specified in Article 56(2) of the Vienna Convention on the Law of Treaties.

We include a number of relevant control variables. First, we incorporate three variables related to the structure of the network, which further capture potential interdependent decision making within the network. MEA POPULARITY (SQRT) is the sum of the total number of members of each MEA a state has signed. It captures whether "popular" MEAs that have more members are more likely to attract additional members. STATE MEAS (SQRT) is measured as the square root of the number of MEAs a state has previously signed. It captures the tendency of states that have signed many MEAs to be more likely to sign additional MEAs, all else equal. SHARED MEMBERS (SAME REGION) counts the number of four-cycles a state is involved in. It evaluates how co-membership within a region affects signing decisions.[20] A positive and statistically significant SHARED MEMBERS (SAME REGION) coefficient indicates that having a shared member makes signature more likely.

We also include the variable GLOBAL MEAS WITH SAME NOTIFICATION PERIOD (SQRT) as a means of capturing the alternative socialization explanation. It is measured as the square root of the number of MEAs existing globally with the same withdrawal notification period length measured in months as the new MEA in question—minus a state's own MEAs with that feature. A positive and statistically significant coefficient for GLOBAL MEAS WITH SAME NOTIFICATION PERIOD (SQRT) would suggest that a state is more likely to sign an MEA with a particular notification period length the more times that notification period length has been included in any other MEA within the environmental regime complex, all else equal. In this way, it captures a socialization dynamic. We also create an alternative formulation of this variable: PROPORTION GLOBAL MEAS WITH SAME NOTIFICATION PERIOD.

---

[20] This is a closure affect akin to triadic closure in one-mode networks.

In addition, we include a number of MEA-level controls that are common in the treaty commitment literature.[21] First we include the variables NORTH-NORTH, SOUTH-SOUTH, and NORTH-SOUTH. These three variables capture whether members of an MEA are all Global North states, Global South states or a combination thereof. NORTH-SOUTH is the excluded reference category. Next, treaties that are of a hard law nature have been shown to reduce the likelihood a state will commit to an MEA (Spilker & Koubi 2016). We therefore include the following three variables: HARD ENFORCEMENT is coded 1 when an MEA includes provisions that allow for non-compliance to be met with sanctions or the revocation of the treaty and 0 otherwise. DELEGATION measures the degree to which states have created and transferred authority to an MEA's institutional body. It ranges from 0 to 3, with higher values representing a more delegation. ENVIRONMENTAL OBLIGATION captures the degree to which an MEA includes specific environmental obligations such as limiting emissions, limiting environmentally damaging activities, and enforcing existing domestic and international environmental laws. It ranges from 0 to 3, with 3 corresponding to the highest possible level environmental obligation.[22]

We also expect that the overall degree of flexibility within an MEA may affect signature patterns. We therefore include FLEXIBILITY, an MEA-level index ranging from 0 to 3, with higher scores indicating more flexibility. Our FLEXIBILITY measure specifically focuses on flexibility provisions

[21] We do not include a control for general NGO membership in our model specification as Spilker and Koubi (2016) find that it is not a statistically significant predictor of MEA participation. However, in our analysis we control for a state's level of participation within the MEA network specifically.
[22] For additional information on the coding of these indices, refer to Appendix A.

that reduce the rigidity of commitments for states, while they remain a member of the treaty.[23] These include: the possibility to make reservations, the ability to modify an MEA in the future, and the presence of "exceptions" clauses that allow a state to have provision-specific opt-outs of the treaty. We also control for UNIVERSAL MEMBERSHIP at the MEA-level. It captures whether an treaty is open for membership to any state. It is coded 1 when this is the case and 0 otherwise.

Finally, at the MEA-level, we include a series of binary SUBJECT variables as it is likely there is variation in signature preferences within the broad environmental regime complex according to unique environmental regimes within the complex, such as fisheries and pollution. The inclusion of these subject variables also controls for uncertainty, which may vary from one subject to another. SUBJECT – POLLUTION is the excluded reference category.

Our last set of control variables are measured at the state-level. REGIME is taken from Polity IV (Marshall et al. 2018) and ranges from -8 to 8, with higher values indicating that a state is more democratic. POWER IS based on the widely used Composite Index of National Capability (Singer 1987; Greig & Enterline 2017), which creates a score incorporating total population, urban population, iron and steel production, energy consumption, military personnel, and military expenditure. Scores range from 0 to 1, with higher values corresponding to an increased degree of power. ENVIRONMENTAL PROTECTION values originate from the World Database on Protected Areas (UNEP-WCMC & IUCN 2019) and are a measure of terrestrial protected areas as a percent of total land area.

[23] For additional information on the coding of this index, refer to Appendix A.

A standard density parameter is also included in our models, akin to the intercept in logistic regression models. In SAOMs, density is correlated with other included variables, and is thus not interpreted on its own. A standard rate parameter is also included for each period, modelling the number of tie change opportunities given to each actor during that period.

5 Analysis

The main results of our statistical analysis are presented in Table 2, which models the 2004-2014 time period. In Model (1), the main independent variable corresponding to Hypotheses 2 is STATE MEAS WITH SAME NOTIFICATION PERIOD (SQRT). In Model (2), we substitute the main independent variable for PROPORTION STATE MEAS WITH SAME NOTIFICATION PERIOD.

	(1)	(2)
MEA NOTIFICATION PERIOD > 12 MONTHS	2.370 (0.228)***	2.033 (0.225)***
MEA NOTIFICATION PERIOD < 12 MONTHS	-1.804 (0.285)***	-1.972 (0.278)***
NO MEA NOTIFICATION PERIOD	-0.557 (0.226)*	-0.793 (0.212)***
MEA NOTIFICATION PERIOD = 12 MONTHS (reference category)	-	-
STATE MEAS WITH SAME NOTIFICATION PERIOD (SQRT)	0.737 (0.123)***	-
PROPORTION STATE MEAS WITH SAME NOTIFICATION PERIOD	-	1.317 (0.495)**
Structural controls		
STATE MEAS (SQRT)	-0.021 (0.059)	0.189 (0.055)***
MEA POPULARITY (SQRT)	0.335 (0.032)***	0.325 (0.033)***
SHARED MEMBER (SAME REGION)	0.016 (0.002)***	0.015 (0.002)***
Other controls		
GLOBAL MEAS WITH SAME NOTIFICATION PERIOD (SQRT)	-0.011 (0.036)	-
PROPORTION GLOBAL MEAS WITH SAME NOTIFICATION PERIOD	-	-3.342 (0.503)***
DELEGATION	-0.545 (0.092)***	-0.439 (0.093)***
ENVIRONMENTAL OBLIGATION	1.228 (0.139)***	1.288 (0.139)***
FLEXIBILITY	0.728 (0.108)***	0.673 (0.109)***
REGIME	-0.012 (0.019)	-0.015 (0.022)
POWER	3.317 (3.703)	3.278 (3.630)
NORTH-NORTH	-3.391 (0.814)***	-3.702 (0.792)***
SOUTH-SOUTH	-0.227 (0.200)	-0.355 (0.196)$^{\cdot}$
HARD ENFORCEMENT	-1.916 (0.331)***	-2.066 (0.324)***
ENVIRONMENTAL PROTECTION	0.003 (0.008)	0.004 (0.008)
SUBJECT - AGRICULTURE	-1.306 (0.347)***	-1.106 (0.318)***
SUBJECT - CONSERVATION	-2.727 (0.365)***	-2.538 (0.352)***
SUBJECT - ENERGY	-0.171 (0.293)	0.111 (0.283)
SUBJECT - FISHERIES	-0.302 (0.224)	-0.183 (0.229)
SUBJECT - FRESHWATER	-0.089 (0.314)	0.041 (0.307)
SUBJECT - HABITAT	-1.663 (0.271)***	-1.454 (0.279)***
SUBJECT – POLLUTION (reference category)	-	-
SUBJECT - WEAPONS	-0.859 (0.633)	-0.502 (0.586)
SUBJECT - OTHER	-1.338 (0.484)**	-1.181 (0.476)*
UNIVERSAL MEMBERSHIP	-0.097 (0.167)	-0.264 (0.168)
RATE	1.536 (0.087)***	1.538 (0.088)***
DENSITY (INTERCEPT)	-2.906 (0.646)***	-3.195 (0.650)***
Iterations	3021	3021

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 2: SAOM Results

Beginning with our independent variables related to withdrawal clause content, our results suggest that as rational design research predicts, the type of withdrawal provision included in an MEA affects the likelihood that a state will sign the agreement. In support of Hypothesis 1, we find that MEA's with notification periods of greater than 12 months, are over 10 times more likely to attract an additional member than those with notification periods of the standard 12 months in Model 1, with the corresponding value for the later time period being 7.6. Yet, it is important to note that only around two percent of MEAs fall within this category. MEAs with no withdrawal notification period are less likely to attract additional signature ties than those with notification periods of 12 months across both Models, all else equal.

We also find that in addition to design content, design context matters. MEAs that include withdrawal clause formulations that states have had greater prior individual experience with, are more likely to attract members across Models 1 and 2. In Model 1 for example, for each additional unit increase in STATE MEAS WITH SAME NOTIFICATION PERIOD (SQRT), there is a 2.09 fold increase in the odds that a state will form sign the new MEA, all else equal.[24] A one-unit increase here corresponds to a state going from having one MEA with the same withdrawal clause notification period length to having four MEAs with a similar clause. In Model 2, each unit increase in the proportion of MEAs a state has signed with a similar withdrawal clause notification period length, increases its odds of signing a new MEA with a similar withdrawal clause by 3.73 times. These results align with the expectations of Hypothesis 2.

[24] While a procedure for evaluating how strongly individual variables contribute to tie-change probabilities in SAOMs has been developed (see Indlekofer and Brandes 2013), at this time, it is not applicable to bipartite network configurations. We therefore rely on odds ratio-based interpretations.

Our network structural variables perform largely as expected in Models 1 and 2, with some variation in the STATE MEAS (SQRT) effect. In Model 1 for example, having an additional shared member from the same region for example makes a state 1.6 percent more likely to sign a new MEA. The general popularity of an MEA is also a clear predictor of it gaining additional members across both models, as illustrated by both the positive and statistically significant coefficients for MEA POPULARITY (SQRT). The odds of an MEA that goes from having four to nine members of attracting additional members—corresponding to a one-unit increase of MEA POPULARITY (SQRT)—increase by 40 percent in Model 1 and 38 percent in Model 2. A state's own previous signature behavior in general, i.e. STATE OUTDEGREE ACTIVITY (SQRT), is only a statistically significant predictor of new signature ties in Model 2.

The presence of MEAs with a given withdrawal notification clause globally—i.e. corresponding to the alternative socialization explanation—is less straightforward. In Model 1, GLOBAL MEAS WITH SAME NOTIFICATION PERIOD (SQRT) is not statistically significant. In Model 2, PROPORTION GLOBAL MEAS WITH SAME NOTIFICATION PERIOD is negative and statistically significant. This result suggests that states are less likely to sign new MEAs with more globally popular withdrawal clause designs. We therefore find no clear evidence in support of the alternative socialization explanation, which suggests that states should be more likely to prefer globally popular solutions.

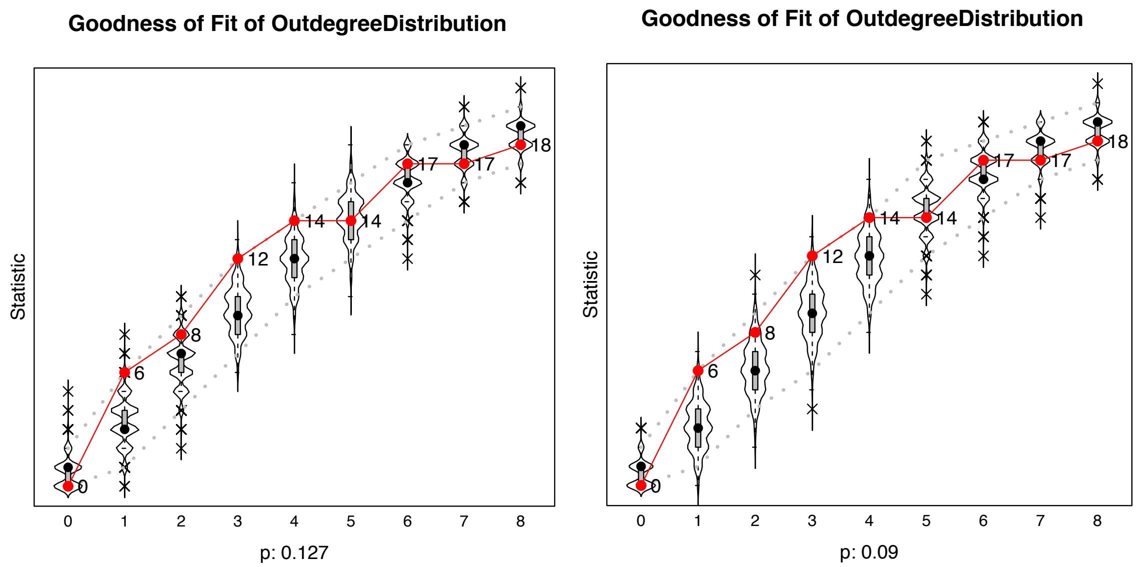
Turning to MEA-level covariates, DELEGATION and HARD ENFORCEMENT have a negative and statistically significant impact on treaty signature across both models. In contrast, the odds that a state will sign an MEA for each one-unit increase in FLEXIBILITY—here in terms of flexibility within the treaty rather than flexibility that facilitates treaty exit—increases by a factor of 2.07 in

Model 1. We also find that states are less likely to sign North-North than North-South MEAs, potentially reflecting that MEAs are often required to solve problems that affect states from all regions of the world. In addition, our SUBJECT controls indicate that there is variation in states' propensity for MEA signing across subject areas within the global environmental governance regime complex, in line with expectations. We find no clear evidence that being open to universal membership impacts the likelihood that a state will sign an MEA.

Finally, moving on to our state-level controls, we find no evidence that REGIME, ENVIRONMENTAL PROTECTION or POWER are strong predictors of MEA signature behavior in either model.

Goodness of Fit

Goodness of fit in SAOM's, as with most network analysis techniques, is assessed by comparing observed values for auxiliary network statistics at the end of each period with simulated values using the Mahalabonis distance. We are primarily interested in our model's ability to predict whether a state will sign new MEAs as measured by the number of signature ties it possesses. We therefore measure goodness of fit by comparing differences in "outdegree distributions", which correspond to the number of outgoing nodal ties (here MEA signature ties). Observed values are superimposed on the plots and the *p*-values included at the base of each plot indicate whether there is a statistically significant difference between observed and simulated values.



Model 1 Model 2

Figure 3: Outdegree Distribution Goodness of Fit Plots

Our models perform sufficiently well, as indicated by the solid line passing through the bodies of the violins in the violin plots and the *p*-values underneath the plots, which both exceed the conventional threshold of 0.05. This indicates that there is no statistically significant difference between observed and simulated values. However, it is worth noting that both of our models slightly underpredict the number of states with four or less ties, visible where the solid line appears at the top of the violins.

Robustness Checks

In addition to our main analyses, we perform a series of robustness checks, available in Appendix B. First, we modify our dependent variable by fitting a similar model to Model 1 using ratification data rather than signature data as our main measure of an MEA tie. While a stronger form of commitment than signature, in that they legally bind a state to fulfil its treaty obligation, due to domestic political hurdles, it is not unusual for states to fail to ratify treaties that they intended to eventually ratify after signing. For this reason, ratification ties are our secondary rather than primary measure of a state's intention to participate in an MEA. Second, we use the same specification as Model 1 on a model using data from the 1994 to 2004 time period. Third, we alter the reference category of our notification period length variables to NO MEA NOTIFICATION PERIOD.

Next, we add in a control for IDEAL POINT DIFFERENCE, a variable based on United Nations State voting pattern data (Bailey et al. 2017). It captures the difference between a state's own ideal point score, and the average ideal point score of members of the treaty it is considering signing. Our measure IDEAL POINT DIFFERENCE ranges from 0 to 5.3, with higher scores corresponding to a greater heterogeneity of ideal point scores between a state and other MEA members on average. We include IDEAL POINT DIFFERENCE in our models as a heterogeneity of preference between prospective members of an MEA and existing member could the attractiveness of forming a signature tie with that MEA.

Finally, we reduce our sample size to only include the MEA signature ties of states that joined an MEA after its entry into force—i.e. late signers. This specification helps us to ensure that what we are witnessing is an effect related to past experience with design solutions impacting new decisions to sign MEAs (i.e. rather than capturing the effect of a state designing an MEA is subsequently want to join). While our focus on multilateral treaties rather than bilateral ensures that individual states do not generally have a full say over treaty design choices, this additional specification provides another way of capturing this dynamic. However, we caution that this specification is not overly relevant within a network analysis framework, as network structural variables necessarily correspond to the full network (e.g. a count of shared members or membership size would be inaccurate in this sub-sample). For this reason, in these models, we do not include network structural effects.

Overall, our additional models perform largely as expected. We find similar results in our ratification tie models as in Model 1. In our model using the 1994 to 2004 data, we find that relatively long notification periods of greater than 12 months are a negative and statistically significant predictor of signature ties in this time period. This suggests that the effect of having long notification periods on signature decisions has changed over time. IDEAL POINT DIFFERENCE is not statistically significant. Lastly, our "late signer" models perform as expected.

6 Conclusions

The UK's withdrawal negotiations with the European Union, the US' departure from the Paris Agreement in 2019, and Canada's withdrawal from the Kyoto Protocol in 2012 have brought withdrawal provisions role in facilitating treaty exit into the public eye. However, withdrawal provisions also play an important role in making treaty participation more appealing both through their design content and design context.

Our results show that the design content and design context of withdrawal provisions both impact treaty signature decisions. More specifically, not only does the specific formulation of a withdrawal provision but also a given state's prior direct experience with it impact its signing propensity. MEAs within the environmental regime complex that have relatively long notification periods as well as those that exploit familiar standard design features are more likely to attract members than other MEAs. While our focus has been on design context within a particular international regime complex, we expect that design context may also be formed by experiences across regime complexes as many treaty design features such as enforcement provisions and membership rules appear throughout all or most treaties.

Our findings also have important implications for a number of related research agendas. First, they affirm that familiarity with design is an important element of research on international signature and commitment decisions more broadly—in addition to the role of design content. We also show that treaty proliferation does not always lead to increased incentives for harmful fragmentation. Rather than leading to fragmentation, growth within regime complexes may in some cases foster coherence through the presence of familiarity bias. Treaty proliferation exposes states with specific standard treaty design features, providing the conditions under which the validation and ingrained biases rationales are likely to arise. In this way, prior experience may encourage coherence within a regime complex. Finally, our findings affirm that cognitive and behavioral research can yield important insights related to international cooperation.

A number of questions also emerge from this study. For example, while we explored exposure that stems from direct and indirect prior experience, we did not differentiate between systematic and automatic decision-making in our analysis. While the rational design research agenda is based on the assumption of systematic decision-making, ingrained biases stemming from exposure effects could realistically be the result of either or both types of processes. Differentiating when each process is likely to dominate, will be an important area of future research. We expect that high-stakes areas are more likely to be dominated by systematic decision-making; whereas, routine and low-impact decisions should be characterized by—though not limited to—more automatic decision-making processes.

Taking into account both design content and design context has great potential to augment our understanding of international treaty design and signature practices and to inform future design choices. By including familiar but also well-suited forms of standard treaty design features, MEAs are more likely to increase their memberships. At a time when international environmental cooperation is increasingly important, smart design choices, while only a partial solution towards encouraging all relevant states to sign new agreements, are likely to be a step in the right direction.

References

Alter, Karen J., 2012. The Global Spread of European Style International Courts. *West European Politics*, 35(1): 135-154.

Baccini, Leonardo, Andreas Dür, and Yoram Z. Haftel, 2015. Imitation and Innovation in International Governance: The diffusion of trade agreement design. In *Trade Cooperation: The purpose, design and effects of preferential trade agreements*, 167-194.

Bailey, Michael A., Anton Strezhnev, and Erik Voeten, 2017. Estimating Dynamic State Preferences From United Nations Voting Data. *Journal of Conflict Resolution*, 61(2): 430-456.

Bilder, Richard B., 1981. *Managing the Risks of International Agreement*. University of Wisconsin Press.

Bornstein, Robert F., 1989. Exposure and Affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin,* 106(2): 265-289.

Bradley, Curtis A., and Mitu Gulati, 2010. Withdrawing From International Custom. *Yale Law Journal,* 120: 202-275.

Copelovitch, Mark S. and Tonya L. Putnam, 2014. Design in Context: Existing international agreements and new cooperation. *International Organization*, 68(2): 471-493.

Denemark, Robert A., and Matthew J. Hoffmann, 2008. Just Scraps of Paper? The dynamics of multilateral treaty-making. *Cooperation and Conflict*, 43(2): 185-219.

Fujimoto, Kayo, Tom AB Snijders, and Thomas W. Valente, 2018. Multivariate Dynamics of One-Mode and Two-Mode Networks: Explaining similarity in sports participation among friends. *Network Science*, 6(3): 370-395.

Greig, J. Michael, and Andrew J. Enterline. 2017. Correlates of War Project National Material Capabilities (NMC) Data Documentation. Version 5.0. Available at: http://www.correlatesofwar.org/data-sets/national-material-capabilities. Accessed 6 August 2019.

Gulati, Mitu, and Robert E. Scott, 2012. *The Three and a Half Minute Transaction: Boilerplate and the limits of contract design*. University of Chicago Press.

Hafner-Burton, Emilie M., Stephan Haggard, David A. Lake, and David G. Victor, 2017. The Behavioral Revolution and International Relations. *International Organization*, 71(S1): S1-S31.

Haftel, Y.Z. and Thompson, A., 2018. When Do States Renegotiate Investment Agreements? The impact of arbitration. *The Review of International Organizations*, *13*(1): 25-48.

Helfer, Laurence R., 2005. Exiting Treaties. *Virginia Law Review*, 91: 1579-1648.

Helfer, Laurence R., 2012: Terminating Treaties. In Hollis, Duncan B. ed., *The Oxford Guide to Treaties*. Oxford University Press. 634-650.

Henning, C. Randall, and Tyler Pratt, 2019. Dimensions of Order in International Regime Complexes: Authority, Differentiation, and Collaboration. Presented at APSA Annual Meeting, Washington, D.C., August 31, 2019.

Hollway, James. and Johan Koskinen, 2016. Multilevel Embeddedness: The case of the global fisheries governance complex. *Social Networks*, 44: 281-294.

Hopf, Ted, 2010. The Logic of Habit in international Relations. *European Journal of International Relations*, 16(4): 539-561.

Huysmans, Martijn, 2019. Enlargement and Exit: The origins of Article 50. *European Union Politics,* 0(0): 1–21.

Indlekofer, Natalie. and Ulrik Brandes, 2013. Relative Importance of Effects in Stochastic Actor-Oriented Models. *Network Science*, 1(3): 278-304.

Jetschke, Anja. and Philomena Murray, 2012. Diffusing Regional Integration: The EU and Southeast Asia. *West European Politics*, 35(1): 174-191.

Kahneman, Daniel, 2011. *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Kerr, John, 2017. Full Text of Lord Kerr's Speech 'Article 50: The Facts'. Open Britain. Accessed 1 January 2019. https://www.open-britain.co.uk/full_text_of_lord_kerr_s_speech_article_50_the_facts

Kertzer, Joshua D. and Dustin Tingley, 2018. Political Psychology in International Relations: Beyond the paradigms. *Annual Review of Political Science*, 21: 319-339.

Kim, Soo Yeon. and Mark S. Manger, 2017. Hubs of Governance: Path dependence and higher-order effects of preferential trade agreement formation. *Political Science Research and Methods*, 5(3): 467-488.

Kinne, Brandon J., 2013. Network Dynamics and the Evolution of International Cooperation. *American Political Science Review*, 107(4): 766-785.

Koremenos, Barbara, 2001. Loosening the Ties that Bind: A learning model of agreement flexibility. *International Organization,* 55(2): 289-325.

Koremenos, Barbara., Charles Lipson, Duncan Snidal, 2001. The Rational Design of International Institutions. International organization, 55(4): 761-799.

Koremenos, Barbara., 2005. Contracting Around International Uncertainty. *American Political Science Review*, 99(4): 549-565

Koremenos, Barbara, and Allison Nau, 2010. Exit, No exit. *Duke Journal of Comparative and International Law*, 21: 81-119.

Koskenniemi, M., 1992. Breach of Treaty or Non-compliance? Reflections on the Enforcement of the Montreal Protocol. *Yearbook of International Environmental Law*, *3*(1): 123-162.

Koskinen, Johan, and Christofer Edling, 2012. Modelling the Evolution of a Bipartite Network—Peer referral in interlocking directorates. *Social Networks,* 34(3): 309-322.

Lospinoso, Joshua A., Michael Schweinberger, Tom A.B. Snijders, and Ruth M. Ripley, 2011. Assessing and Accounting for Time Heterogeneity in Stochastic Actor Oriented Models. Advances in data analysis and classification, 5(2): 147-176.

MacKenzie, David Clark. 2010. *A World Beyond Borders: An Introduction to the History of International Organizations*. Toronto: University of Toronto Press.

Manger, Mark S., Mark A. Pickup, and Tom A.B. Snijders, 2012. A Hierarchy of Preferences: A longitudinal network analysis approach to PTA formation. *Journal of Conflict Resolution*, 56(5): 853-878.

Marshall, Monty G., Keith Jaggers, and Ted R. Gurr, 2018. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2017. Centre for Systemic Peace.

Milewicz, Karolina, James Hollway, Claire Peacock, and Duncan Snidal, 2018. Beyond Trade: The expanding scope of the nontrade agenda in trade agreements. *Journal of Conflict Resolution*, 62(4): 743-773.

Mitchell, R.B., 2006. Problem structure, institutional design, and the relative effectiveness of international environmental agreements. *Global Environmental Politics*, *6*(3): 72-89.

Mitchell, Ronald B. 2018. International Environmental Agreements Database Project (Version 2018.1). Available at: http://iea.uoregon.edu/. Accessed 01 July 2018.

Mohrenberg, Steffen., Vally Koubi, and Thomas Bernauer, 2019. Effects of Funding Mechanisms on Participation in Multilateral Environmental Agreements. *International Environmental Agreements: Politics, law and economics*, 19(1): 1-18.

Morin, Jean-Frédéric, Dominique Blümer, Clara Brandi, and Axel Berger, 2019. Kick-Starting Diffusion: Explaining the varying frequency of PTAs' environmental clauses by their initial conditions, *World Economy* 00:1–2

Peacock, Claire, Karolina Milewicz, and Duncan Snidal (2019). Boilerplate in International Trade Agreements. *International Studies Quarterly*. TBD.

Pierson, Paul, 2000. Increasing Returns, Path Dependence, and the Study of Politics. *American Political Science Review*, 94(2): 251-267.

Powell, Robert. 2017. Research Bets and Behavioral IR. *International Organization* 71 (S1): S265–S277.

Raustiala, Kal, 2012. Institutional Proliferation and the International Legal Order in Dunoff and Pollack, eds., *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art*, Cambridge.

Ripley, Ruth M., Tom A.B. Snijders, Zsófia Boda, Andràs Vörös, and Paulina Preciado, 2019. Manual for RSiena, University of Oxford Department of Statistics and Nuffield College.

Rosendorff, B. Peter, and Milner, Helen V., 2001. The Optimal Design of International Trade Institutions: Uncertainty and escape. I*nternational Organization*, 55(4): 829-857.

Skovgaard Poulsen, Lauge N., 2014.. Bounded Rationality and the Diffusion of Modern Investment Treaties. *International Studies Quarterly* 58(1): 1-14.

Singer, J. David. 1987. Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985. *International Interactions*, 14: 115-32.

Snijders, Tom A.B., Gerhard G. Van de Bunt, and Christian EG Steglich, 2010. Introduction to Stochastic Actor-Based Models for Network Dynamics. *Social Networks* 32(1): .44-60.

Snijders, Tom AB. 2011. Statistical Models for Social Networks. *Annual Review of Sociology* 37: 131-153.

Snijders, Tom A.B., Alessandro Lomi, and Vanina Jasmine Torló, 2013. A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks* 35 (2): 265-276.

Spilker, Gabriele. and Vally Koubi, 2016. The Effects of Treaty Legality and Domestic Institutional Hurdles on Environmental Treaty Signature. I*nternational Environmental Agreements: Politics, Law and Economics*, 16(2): 223-238.

Thompson, A., Broude, T. and Haftel, Y.Z., 2019. Once Bitten, Twice Shy? Investment Disputes, State Sovereignty, and Change in Treaty Design. *International Organization*, *73*(4): 859-880.

Tobin, Harold James, 1933. *The Termination of Multipartite Treaties*, Columbia University Press.

UNEP-WCMC and IUCN (2019). Protected Planet: The World Database on Protected Areas (WDPA), [2016 Version], Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net. Accessed 28 February 2019.

World Trade Organization, 2018. World Trade Statistical Review 2018. Available at: https://www.wto.org/english/res_e/statis_e/wts2018_e/wts2018_e.pdf Accessed 28 February 2019.

Zajonc, R.B., 1980. Feeling and Thinking: Preferences need no inferences. *American Psychologist*, 35(2): 151-175.

Zajonc, R.B., 2001. Mere Exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, 10(6): 224-228.

How withdrawal provisions affect multilateral environmental agreement signature preferences

Appendix A: Additional Coding Information

In this Appendix, we provide additional information related to the coding of our MEA design-related variables. Further details related to the coding of the underlying variables can be found in the DITES Codebook (Morin 2019).[1]

FLEXIBILITY is an additive index comprised of underlying variables measuring design content related to exceptions, reservations, and amendments—i.e. flexibility unrelated to facilitating withdrawal. The exceptions component of the index is coded as 1 when at least one of underlying exceptions variables outlined in section 8.01 of the DITES Codebook is coded 1. The reservations component of the index ranges from 1 to 3 and uses the data explained in section 8.02 of the DITES Codebook. Higher scores indicate a greater potential to make reservations. These values are recoded to fall between 0 and 1. Agreements that score 0 offer no possibility for making reservations. Finally, the amendments component of the index is coded 1 when there are explicit rules that allow an agreement or its annexes to be amended in the future and 0 otherwise (see section 9 of the DITES Codebook). The resulting three sets of values are added together to create the flexibility index we use in the paper, which ranges from 0 to 3.

DELEGATION also ranges from 0 to 3. It is based on the variables outlined in sections 12.01; 12.02; 12.03; and 12.05 of the DITES Codebook, with different points given for each sub-component. The details of these choices are available in the R script that accompanies this paper.

ENVIRONMENTAL OBLIGATION ranges from 0 to 3. Its components are the variables outlined in 5.09; 10.03; and 6.01 to 6.07 of the DITES Codebook. The specific coding of this index is available in the R script that accompanies this paper.

Finally, HARD ENFORCEMENT is a dummy variable. It is codded as 1 when the variable outlined in X13.05 or 13.06 is coded as 1 (i.e. "Sanction or suspension of benefits in case of non-conformity" and "Suspension of the agreement or revocation of membership").

[1] Soon available at <XXXX>.

How withdrawal provisions affect multilateral environmental agreement signature preferences

Appendix B: Robustness Checks

The results of our robustness checks are presented in Table B1 below. In Model 1, we replace our treaty signature data with treaty ratification data. In Model 2, we substitute our 2004 to 2014 data presented in the paper for data from the 1994 to 2004 time period. Next, in Model 3, we we alter the reference category of our notification period length variables to NO MEA NOTIFICATION PERIOD from MEA NOTIFICATION PERIOD = 12 MONTHS. In Model 4, we include an additional variable IDEAL POINT DIFFERENCE. Finally, in Model 5, we limit ourselves to evaluating "late signing" decisions that occur after an MEA is officially signed.

	(1) Ratification	(2) 1994-2004	(3) Changed Ref. Cat.	(4) Ideal Point diff.	(5) Late Signing
MEA NOTIFICATION PERIOD > 12 MONTHS	1.031 (0.173)***	-4.085 (0.648)***	2.393 (0.225)***	2.400 (0.227)***	1.361 (0.221)***
MEA NOTIFICATION PERIOD < 12 MONTHS	-0.467 (0.207)*	0.393 (0.240)	-1.244 (0.297)***	-1.805 (0.286)***	-5.165 (1.289)***
NO MEA NOTIFICATION PERIOD	-0.428 (0.171)*	0.887 (0.200)***	-	-0.548 (0.228)*	0.148 (0.459)
MEA NOTIFICATION PERIOD = 12 MONTHS	-	-	0.697 (0.222)**	-	-
STATE MEAS WITH SAME NOTIFICATION PERIOD (SQRT)	1.213 (0.199)***	0.380 (0.115)***	0.488 (0.144)***	0.539 (0.137)***	1.408 (0.618)*
Structural controls					
STATE MEAS (SQRT)	2.678 (2.553)	-0.065 (0.079)	-0.005 (0.059)	-0.027 (0.065)	-
MEA POPULARITY (SQRT)	0.337 (0.022)***	0.472 (0.033)***	0.329 (0.033)***	0.335 (0.032)***	-
SHARED MEMBER (SAME REGION)	0.030 (0.002)***	0.015 (0.002)***	0.016 (0.001)***	0.016 (0.001)***	-
Other controls					
GLOBAL MEAS WITH SAME NOTIFICATION PERIOD (SQRT)	-0.142 (0.044)**	-0.160 (0.039)***	-0.079 (0.035)*	-0.076 (0.037)*	-0.386 (0.105)***
DELEGATION	0.063 (0.061)	0.160 (0.090)$^{.}$	-0.552 (0.095)***	-0.539 (0.091)***	-0.382 (0.091)***
ENVIRONMENTAL OBLIGATION	0.549 (0.098)***	0.613 (0.170)***	1.214 (0.142)***	1.237 (0.144)***	0.978 (0.154)***
FLEXIBILITY	0.180 (0.073)*	-0.662 (0.108)***	0.726 (0.103)***	0.732 (0.109)***	0.670 (0.132)***
REGIME	0.454 (0.471)	0.000 (0.016)	-0.013 (0.020)	-0.010 (0.019)	0.098 (0.080)
POWER [2]	0.000	25.852 (13.968)$^{.}$	3.272 (3.856)	3.160 (3.899)	-4.171 (18.637)
NORTH-NORTH	-2.444 (0.508)***	-6.116 (1.479)***	-3.378 (0.772)***	-3.342 (0.861)***	-5.250 (2.643)*
SOUTH-SOUTH	-0.367 (0.142)**	-0.524 (0.191)**	-0.239 (0.196)	-0.224 (0.201)	-1.198 (0.235)***
HARD ENFORCEMENT	-1.201 (0.249)***	0.026 (0.225)	-1.903 (0.333)***	-1.927 (0.353)***	-2.131 (0.990)*

[2] While we include POWER in each model, the algorithm was unable to reach a precise estimate of the value of its coefficient in Model 1, though it exhibited a large and positive coefficient, with a large standard error, suggestive of the Donner Hauck phenomenon (Hauck & Donner 1977). We therefore test POWER separately using a score-type test as recommended for such variables in SAOM analyses (Ripley et al. 2019:85; 89-90). Using this procedure, we find evidence that POWER is an important predictor of ratification decisions.

ENVIRONMENTAL PROTECTION [3]	-	0.010 (0.010)	0.003 (0.008)	0.003 (0.008)	0.011 (0.057)
SUBJECT - AGRICULTURE	-0.733 (0.204)***	0.726 (0.270)**	-1.278 (0.333)***	-1.292 (0.327)***	-1.522 (0.428)***
SUBJECT - CONSERVATION	-1.306 (0.204)***	1.384 (0.222)***	-2.691 (0.383)***	-2.716 (0.377)***	-8.161 (2.518)**
SUBJECT - ENERGY	-0.345 (0.185)·	1.282 (0.313)***	-0.164 (0.275)	-0.174 (0.277)	-0.265 (0.290)
SUBJECT - FISHERIES	-0.463 (0.203)*	1.770 (0.208)***	-0.266 (0.222)	-0.283 (0.228)	-1.221 (0.275)***
SUBJECT - FRESHWATER	0.220 (0.181)	0.791 (0.337)*	-0.100 (0.315)	-0.070 (0.315)	-0.364 (0.296)
SUBJECT - HABITAT	-0.641 (0.172)***	1.076 (0.203)***	-1.656 (0.271)***	1.639 (0.276)***	-2.861 (0.494)***
SUBJECT – POLLUTION (reference category)	-	-	-	-	-
SUBJECT - WEAPONS	0.058 (0.261)	-0.951 (0.405)*	-0.810 (0.590)	-0.854 (0.630)	0.052 (0.866)
SUBJECT - OTHER	1.510 (0.160)***	0.966 (0.557)·	-1.210 (0.510)*	-1.320 (0.490)**	-1.523 (0.502)**
UNIVERSAL MEMBERSHIP	-0.293 (0.113)**	-0.352 (0.152)*	-0.134 (0.164)	-0.092 (0.160)	0.391 (0.197)*
IDEAL POINT DIFFERENCE.	-	-	-	0.088 (0.096)	-
RATE	4.598 (0.154)***	1.764 (0.093)***	1.539 (0.091)***	1.538 (0.089)***	1.287 (0.079)***
DENSITY (INTERCEPT)	-13.777 (8.901)	-2.513 (0.682)***	-3.431 (0.671)***	-2.963 (0.714)***	2.005 (0.660)**
Iterations	2232	3021	5769	3051	2909

***p < 0.001, **p < 0.01, *p < 0.05, ·p < 0.1

Table B1: Robustness Check SAOM Models

[3] The algorithm was unable to converge in this specification when the ENVIRONMENTAL PROTECTION variable was included. We therefore chose to remove it. As ENVIRONMENTAL PROTECTION is not a central variable to our theoretical predictions and is not statistically significant in our other model specifications, we do not expect this change to have a meaningful impact on our results.

References

Ripley, Ruth M., Tom A.B. Snijders, Zsófia Boda, Andràs Vörös, and Paulina Preciado, 2019. Manual for RSiena, University of Oxford Department of Statistics and Nuffield College.