Insuring the weak: Exit clauses in international organization*

Benjamin Daßler†, Tim Heinkelmann-Wild†, Martijn Huysmans‡

Abstract

Materially powerful states tend to dominate the creation and policymaking of international organizations (IOs). This makes it puzzling why materially weak states and states with peripheral preferences should be willing to participate in IOs, given that initial benefits might give way to policy changes and exploitation in the future. We suggest that weak states join IOs when they are granted a free exit right as an insurance against adverse policy changes. Exit clauses as insurance are especially relevant when IO decision-making is non-unanimous. A regression analysis of IOs created between 1945 and 2014 lends support to our theoretical argument: exit clauses are more common in IOs that take decisions by weighted or majority voting than in those that make decisions unanimously. Our findings matter because if the nature of cooperation changes, exit clauses may make actual exit more likely in the long run.

Keywords: Exit Clauses; International Organizations; Institutional Design; Withdrawal.

---

* We thank Inken von Borzyskowski, Christina Cottiero, and participants at the IR Colloqium (LMU Munich), the power of the weak workshop, ISA 2022, and Utrecht PPE for their comments on earlier versions of this manuscript. We are grateful to Willem Bracher, Gustav Munch, Kush Saluja and Maya von Ahnen for their research assistance. All shortcomings remain our own.

† LMU Munich.

‡ Assistant Professor, Utrecht University, School of Economics. Corresponding author. Mailing address: PO 80125, 3508 TC Utrecht, Netherlands. Email: m.huysmans@uu.nl.

1 Introduction

Recent years witnessed numerous withdrawals of states from international organizations (IOs). The United Kingdom (UK) terminated its membership of the European Union (EU). In 2018, the Philippines withdrew from the International Criminal Court. In 2004, the Seychelles withdrew from the Southern African Development Community. Exit also happens from less high-profile IOs: in 2011, Ireland withdrew from the International Organization for Wine and Vine, Panama left the International Grains Council, and Latvia withdrew from the World Tourism Organization. States' withdrawal from IOs is highly consequential for future multilateral cooperation. Withdrawals deprive IOs from economies of scale and critical material contributions such as resources and staff. Withdrawals also have the potential to de-legitimate IOs as exit is "the ultimate act of disrespect for international rules and institutions" and "may effectively discredit those entities in the eyes of other states" (Helfer 2005: 1587; see also Sommerer and Agné 2018; Tallberg and Zürn 2019; Walter 2021).

While member states' withdrawal is thus harmful, many IO constitutions explicitly provide members with the legal right to exit. In fact, so called withdrawal or exit clauses are a relatively common design feature in IOs (Helfer 2005, 2019; Koremenos 2016). Whether IO rules provide members with a legal exit right or not matters (Helfer 2005: 1589-1591; Koremenos 2016: 146-47; Koremenos and Nau 2010: 84-85; Guzman 2002; Rosendorff and Milner 2001: 834). In contrast to treaty breach, exit clauses offer a legal way to withdraw from an international institution. This helps states to avoid material and reputational costs. By reducing the cost of exit, Article 50 of the Treaty on European Union likely facilitated Brexit (Huysmans 2019).

But why then do founding governments include exit clauses when designing new IOs? In other words: why talk about divorce on your wedding day (Weiler 1985)? Scholarship from International Relations and International Law provides a functionalist answer. The choice of an IO's founding members to include exit clauses or not is assumed to be based on whether the future uncertainty of cooperation requires flexibility (Helfer 2005; Koremenos 2016; Koremenos et al. 2001; Koremenos and Nau 2010; Rosendorff and Milner 2001). This literature suggests that the design of exit clauses is affected by the likelihood of exogenous shocks (Koremenos 2005, 2016) and the nature of cooperation, e.g., enforcement and commitment problems (Koremenos 2016; Koremenos and Nau 2010).

While we agree that founding states rationally design exit clauses in IO constitutions, we contribute by addressing three gaps in this literature. First, there is surprisingly little research

on why exit clauses are included in IO constitutions. The literature often studies *how* exit clauses are designed (e.g., the length of wait or notice periods) and not *whether* they are incorporated in the first place (Koremenos & Nau 2010; Koremenos 2005, 2016: 141). Second, the bulk of the literature studies international agreements in general and *not IOs* (Koremenos & Nau 2010; Koremenos 2016; Helfer 2005). The cooperation logic of IOs typically differs from international agreements as IO policymaking usually develops and changes the initially agreed policies on a regular basis. Finally, the literature assumes that founding states include exit clauses to address future uncertainty (Koremenos & Nau 2010; Koremenos 2016; Helfer 2005) but does not explain when exit clauses are included in IOs when they are *dysfunctional*. For instance, while exit clauses in IOs providing club goods do not undermine cooperation, their inclusion in IOs providing public goods does (see Helfer 2005). Consider the International Monetary Fund (IMF). Since the IMF is a global lender of last resort, any exit diminishes the provision of global financial stability as a public good. Yet its withdrawal clause was used by Poland in 1950, Cuba in 1964, and Indonesia in 1965.

To fill these gaps, we complement the functionalist logic emphasized by the literature with a power logic to explain the inclusion of exit clauses in IO constitutions. By threatening not to join an IO, weak states can modify institutional rules in order to limit the translation of material weakness into institutional weakness. When creating an IO, materially weak states and those with peripheral preferences worry about being outvoted. This is because material weakness tends to be translated into institutional weakness, such as smaller vote shares. To the extent that IOs will be governed by majority or especially weighted voting, coalitions of states that combine sufficient votes will be in a powerful institutional position. However, at the time of creation, weak states can ask to modify the rules or refuse to join. In particular, they can ask for an exit clause. By making exit less costly and thus more credible, weak states' voice is increased. Exit clauses can thereby function as insurance devices for weak states in IOs that do not use unanimity voting. Our empirical evidence is consistent with this theory: exit clauses are more likely in IOs governed by majority or especially weighted voting.

By theorizing and empirically testing the impact of power on institutional design, we contribute to a better understanding of weak states' strategies in IOs. Scholarship recently turned to examine the power of the weak (*add references to special issue framework*). It finds that weak states can draw on strategies within IOs to compensate for their material weakness, such as forming coalitions, issue-linkage, or investing in skilled diplomats (*add references to respective contributions of the special issue*). Our paper adds to this scholarship by identifying a strategy

weak states can employ already at the time of IO creation to bolster their institutional power (see Barnett and Duvall 2005): By threatening not to join an IO, weak states can push for the inclusion of exit clauses in IO constitutions. This rule modification bolsters their credibility when pursuing cross-institutional strategies, such as forum shopping (see Morse and Keohane 2014; *add references to respective contributions of the special issue*). Lowering the hurdles of exit from IOs thereby improves weak states' voice within them.

The remainder of this paper is structured as follows: Section 2 presents our theory about how exit clauses insurance weak states in IOs. Section 3 spells out the game theoretic micro-foundations of our theory: weak states expecting to lack a veto will require them to join an IO. Section 4 presents our research design. Section 5 tests our theory for IOs created 1945-2014. Section 6 concludes by summarizing our findings and discussing their implications for when states will draw on exit clauses not to bolster their voice but indeed withdrawal from IOs.

2 Power and exit clauses

Scholarship largely agrees that cooperation in international institutions, such as IOs, has distributional consequences for its members. When it comes to losers and winners, however, scholarship diverges. *Realist scholarship* emphasizes that materially powerful states will use international institutions to dominate less capable states. Coercion through institutions thereby offers the advantage to maintain an illusion of legitimacy (Martin 1993). Small countries get rules, but great powers get their rules. By contrast, *Hegemonic Stability Theory* (HST), suggests capable member states usually carry the bulk of the burden of cooperation in international institutions. Less capable states might even free-ride on the contributions by materially more capable states (Kindleberger 1981, 1986; Snidal 1985: 581; Webb and Krasner 1989). "The small exploit the great" (Olson 1965). Similarly, *liberal-institutionalist scholarship* emphasizes that capable states push for international institutions to protect themselves from arbitrary exploitation by great powers. Capable states commit to respecting the institutionalized rights of the weak and to refrain from using their capacities outside institutions (Cowhey 1993; Ikenberry 2001; North and Weingast 1989).

We start from the assumption that states' power within international institutions is affected by the institutional context of voting rules and membership. In an institutionalized environment, whether states are weak or strong is not (only) a question of material resources but depends on

institutional rules. Institutional rules distribute voting rights and other institutional privileges among their membership (Barnett and Duvall 2005; Jupille et al. 2013; Koremenos et al. 2001; Viola et al. 2015; Zürn 2018). States with a veto or a large voting share are strong, while others are weak. Moreover, states' power in international institutions also hinges on the *other members* as they can be close to the median member state or an outlier in terms of their preferences. To the extent that IOs are governed by majority or weighted voting, coalitions of states that combine sufficient votes are in a powerful position (core), while those lacking the right to prevent undesired outcomes are in a weak position, regardless of their extra-institutional, material capacities (Blake and Payton 2015; Daßler et al. 2022; Tsebelis 2002).

Who forms the core and who the weak – who risks exploitation by whom and to what extent – thus is shaped by the institutional context in which states interact. Materially strong states exploit materially weak states if their privileged position is institutionalized, for instance by weighted voting that grants them the formal power to dictate rules on the remaining members. Take as example the US and institutions such as the IMF or World Bank, where the US clearly calls the shots and small states must obey. In turn, marginalized states have a record of demanding more influence and complaining about Western-bias. By contrast, materially weak states can exploit materially strong states if they form a majority and decisions are taken by voting. Examples include classical public goods, such as the Paris agreement on climate change, where the US complained that other states would take advantage of them. Moreover, materially wealthy countries pay most of the budget in many IOs, while less wealthy states decide on the purposes for which this budget is used. In turn, main donor states have a record of demanding veto rights or leaving. Hence, regardless of their material capacities, states occupying a powerful position within an IO will have more opportunities to shape future policies that will tend to dissatisfy states that are in a weaker position.

The risk of exploitation thus casts a shadow over cooperation. To curb the problem of exploitation, member states can include a formal right to withdraw in an IO's constitution. Exit clauses empower weaker states' voice in an IO by making exit materially and reputationally less costly (Hirschman 1970; Huysmans 2019). In contrast to treaty breach (Schwelb 1967), exit clauses constitute an "internationally lawful act" (Helfer 2005: 1589) which "enables a state to cease cooperation with other treaty parties while avoiding or at least reducing opportunities to be penalized for noncompliance" (ibid.) and, at the same time, "the remaining states are unable to exercise their right under international law to engage in reciprocal acts of noncompliance" (ibid). Besides avoiding sanctions or retaliation, legal exit leaves the

withdrawing state's reputation for compliance unharmed (Guzman 2002; Helfer 2005: 1590-91; Rosendorff and Milner 2001: 834).

As an example, consider the International Fund for Agricultural Development (IFAD). Both developing and developed countries worried about the voting weights given to OPEC countries (Zamora 1980: 586-588). However, in the end the IO was created, with the inclusion of special majority provisions and an exit clause.

Exit clauses increase weak states' voice, as they can threaten to activate the exit clause or activate it and use the notice period to re-negotiate the terms of their membership (see Helfer 2005: 1588-1591). In fact, in the models of complete information developed in the next section, if an exit clause is present no exit threats will need to be made in equilibrium. Core states will understand they have to limit exploitation in order to forestall exit. This equilibrium reasoning implies that one should be very careful to infer from an empirical lack of use that exit clauses have no real effects.

3 A formal theory of exit clause adoption

This section develops game-theoretical models of the decision to adopt an exit clause at IO creation. The actors are the founding states, so that the models provide a micro-foundation for our theoretical argument. The models presented here show that the choices over voting rules and exit rules in an IO constitution are intertwined.

Following our contextual power considerations, a state risks being institutionally weak when it is marginalized in terms of its preferences and anticipates being vulnerable to exploitation by the remainder of the membership due to non-unanimous decision-making rules. Materially weak states are especially likely to be in this position, though materially strong states may also anticipate institutional weakness if their preferences differ strongly from the other members. We will model this type of vulnerability by plotting a priori weak states W against the remaining core of members C.

Consider the following examples of how to apply this modeling choice and identify the powerful core. When the EU was looking to expand East after having mostly let go of member states' policy vetoes, the Eastern accession states successfully demanded an exit right as protection against future policy changes (Huysmans 2019). Here, the core was formed by the existing 15 member states – they were more developed and had a larger combined voting share

than the 10 new member states. As another example, when the International Seabed Authority was being created, developing countries feared that weighted voting would favor the core of developed countries (Zamore 1980). In the end, majority voting was adopted, but an exit clause was included to insure the developing countries.

The next subsection develops and solves three models of IO decision-making under different rules. Based on the solutions of these models, a model of constitutional choice with voluntary accession is presented in the following subsection.

3.1 IO decision-making under different rules

In the models presented here, an initial set of IO policies is assumed to have been agreed upon. These initial policies generate expected utilities θ_i for member i compared to the status quo without the IO. The substance of these initial utilities varies across IOs: it could be gains from trade through liberalization, the resolution of a security dilemma, or the provision of global financial stability. Some members may expect to benefit more than others, and the initial policies include IO budget contributions that may vary across members. However, once the IO has been formed the initial set of policies may be altered. IO rules can limit the statutory competencies of the IO, as well as specify rules for decision-making and exit. The models presented here focus on different configurations of decision-making and exit rules.

When considering the interaction of voting rules with exit rules,[1] the crucial aspect of voting rules is how much protection they offer weak states against adverse policy changes. In the real world, this depends on the exact distribution of votes and preferences across members. In the models, we simplify things in order to focus on the effective amount of protection. Assume the IO makes decisions under unanimity. Then, clearly, all members can block unwanted policy

[1] For the game-theoretical models, we distinguish between voting rules that provide member states a unilateral veto or not, and the presence or absence of exit clauses. By contrast, Blake and Payton (2015) divide decision-making rules into three categories. Unanimity voting gives each member a veto. Weighted voting uses unequal voting shares, with typically larger shares for large of powerful states. Majority voting, as coded by Blake and Payton, groups all systems where the member states each have one vote – regardless of whether a simple or supermajority threshold is required. Exit rules may vary in terms of waiting and notification periods (Koremenos 2016). Theoretically, they may also specify penalties for leaving (Huysmans and Crombez 2020). For simplicity, the game-theoretical models presented here assume exit clauses allow for free exit.

changes. In all other cases, weak members may not be able to block unwanted policy changes.[2] Under weighted voting, a hegemon or a group of like-minded states are likely to have a sufficiently large voting share to impose their preferred policies. Hence, weighted voting leaves weak members especially vulnerable to unwanted policy changes. However, majority voting with equal vote shares does not offer the same protection as unanimity either. Depending on the dimensionality of politics and the distribution of preferences, a group of like-minded states may still be able to systematically push through their preferred policies under majority voting.

We model IOs as consisting of two blocs: the core C and the weak W. The core is the state or group of states that, unless unanimity voting is used, can push through its desired policy changes at the expense of the weak. The core is likely to be made up of a hegemon and/or a group of powerful states with central preferences, while the weak groups vulnerable states that are materially or contextually weak in terms of preferences.

The assumption that there are only two types of members, core and weak, is obviously simplifying. Richer models would consider each member's preferences individually and conceive of heterogeneity over all members, and not just as core versus weak. In such models, each member would have its own individual type, leading to different payoffs from the IO. Even richer models would refer to individual agents such as negotiators and heads of state, who may be thinking about re-election, audience costs, or private benefits instead of (only about) the state's general interest.

Under the simplifying assumption that the IO consists of two blocs, each member i belongs to either the core or the weak and will derive the corresponding utility θ_C or θ_W from the initial set of policies. Utility of the status quo (no IO) is normalized to 0. Assuming all members signed up to the IO voluntarily, $\theta_C, \theta_W > 0$. If the IO was designed by a hegemon to further its interests, making $\theta_W > 0$ may require that the hegemon shoulder a disproportionate share of the IO's budget.

After the IO has formed, the initial set of policies may be changed. Exogenous shocks may also occur (Huysmans and Crombez 2020), but those are left out of the model. Theoretically, shocks to the state of the world may explain why also members of the core or even the hegemon would

---

[2] In practice, supermajority or qualified majority voting can give groups of members a veto without giving each individual member a veto. For more details, see the literature on power indices and voting rules more generally (Barberà and Jackson 2006; Blake and Payton 2015; E. R. Gould 2016; M. Gould and Rablen 2017).

want an exit clause. For instance, if the world becomes more peaceful, being part of an IO aimed at mutual defense may become less valuable (Alesina and Spolaore 2003). Similarly, if world trade becomes more open, being part of a regional free trade IO may bring less benefits in terms of effective market size (Ibid.).

For each policy change, the payoffs for the core and the weak may be positive or negative and large or small. Presumably, policy changes that benefit both will be adopted without problems and will only strengthen the IO. In fact, there is no reason not to expect any such policy changes to already have been adopted before the creation of the IO, and hence to already be incorporated in the utilities θ_C, θ_W. Conversely, policy changes that hurt both groups will never be adopted.

Some policy changes may benefit the core and hurt the weak, or vice versa. It is precisely such policy changes that may affect the benefits of the IO for weaker states and ultimately the stability of the IO. Given that by definition the weak anticipate constituting a voting minority (otherwise the labels core and weak would switch), no policy changes that benefit the weak at the expense of the core will be adopted.[3] This is why the models focus on policy changes that benefit the core at the expense of the weak.

Specifically, the potential for majority-approved policy programs that benefit the core and harm the weak is modeled through the variable x, which captures the benefits to the core of such programs. For simplicity, the corresponding harm to the weak is assumed to be $-x$. The payoff of bloc i is denoted as π_i. After approval of a program x, the payoffs are $\pi_C = \theta_C + x$ for the core, and $\pi_W = \theta_W - x$ for the weak.

In each model, the core proposes a majority-approvable program $x \geq 0$; more complicated models with exogenous shocks could allow for programs with $x < 0$ aimed at compensating the weak.

Norms of universalism may impose an upper limit on x. However, in the remainder of this article we assume that even if such norms are present, the weak will take a worst-case perspective and assume that the norms may not always bind. This can be justified by a signaling story: if the core expects to be fully bound by norms of universalism, then giving a free exit right to the weak will not change its own expected benefits. Hence refusing to give such a right

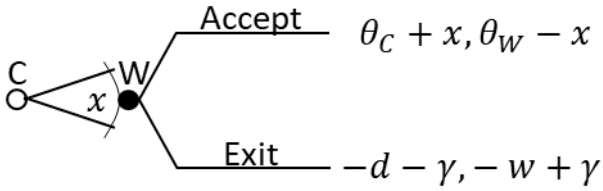
[3] Exogenous changes hurting the core may still occur; these could be a reason for hegemonic withdrawal (Heinkelmann-Wild 2021) or for compensatory policies aimed at keeping the weak from exiting.

might send a signal that the core actually expects not to be fully bound by norms of universalism, reinforcing the weak's expected need for constitutional protection.

Upon learning that the IO wants to implement the program x, the weak can respond. If they lack both a veto and an exit right, their only defense against unwanted policy programs is to break away from the IO, i.e., to exit unilaterally. Such unregulated withdrawal is assumed to entail not only the loss of IO benefits for both, but in addition costs w for the weak and d for the core. Substantively, these costs cover legal and real-world adjustment costs, as well as reputation costs for the states exiting unilaterally.

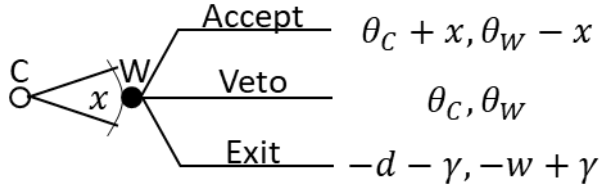
If the IO addressed a commitment problem, unilateral withdrawal is assumed to give a benefit of γ to the leaving weak, and a cost γ for the remaining core. The larger γ, the stronger the temptation to defect, i.e., the stronger the commitment problem. The resulting model is depicted in Figure 1.

Figure 1: An IO without veto or exit right.



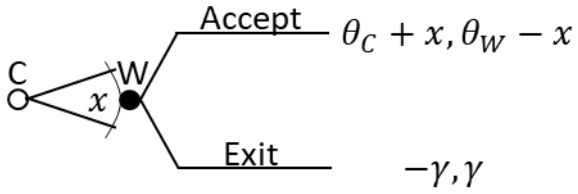
If the weak do have a veto, they can block any policy program x, as depicted in Figure 2.

Figure 2: An IO with a veto right.



An alternative safeguard for the weak is a free exit or withdrawal right. Figure 3 models this situation. If there is no commitment problem, $\gamma = 0$ and exit results in the status quo payoff of 0 for both the core and the weak. To keep things simple, this model abstracts from transaction costs of separation (see Huysmans and Crombez 2020). The next section will discuss what would happen if one relaxes the assumption that exit clauses make exit costless.

Figure 3: An IO with a free exit right.



To summarize, each of these models has two stages. In the first stage, the core C proposes a majority-approvable policy program x. In the second stage, the weak W decide on how to respond to the program. Depending on the combination of decision-making and exit rules, the weak may accept the program, exit unilaterally, veto the program, or use a free exit right.

The models are games of complete and perfect information. This means that the subgame perfect equilibria can be identified using backwards induction.

3.2 Solution of the models and discussion

This section provides the solutions to the models presented above. The model without veto or exit right was presented in Figure 1, and is solved here using backwards induction. In the second stage, W will accept a program x as long as $\theta_W - x \geq -w + \gamma$ or $x \leq \theta_W + w - \gamma$; like in the canonical ultimatum game, when indifferent W is assumed to accept. In words: the weak will accept any program x that does not make staying worse than paying the cost of unilateral exit w (or $w - \gamma$ if there is a commitment problem). Using backwards induction and recalling that $x \geq 0$, C's equilibrium strategy is to propose $x = Max\{0, \theta_W + w - \gamma\}$.

The subgame perfect equilibrium outcome and payoffs depend on θ_W. If $\theta_W \geq -w + \gamma$, the core proposes $x = \theta_W + w - \gamma$ and the weak accept, resulting in payoffs $\pi_C = \theta_C + \theta_W + w - \gamma$ and $\pi_W = -w + \gamma$. In this case the weak are maximally exploited by the core. Only if the

commitment problem is severe ($\gamma \geq w$) do the weak stay. If $\theta_W < -w + \gamma$, the core's proposal is irrelevant. Any proposal $x \geq 0$ can be made in equilibrium, but the weak will exit regardless, with payoffs $\pi_C = -d - \gamma, \pi_W = -w + \gamma$. In this case the set of initial policies was so bad for the weak that they prefer costly exit, taking away the core's benefits from the IO as well.

Figure 2 presented the model with veto right. When indifferent, we assume that W vetoes rather than exits. In the second stage, W vetoes if $\theta_W \geq -w + \gamma$ and exits if $\theta_W < -w + \gamma$. Hence C's proposal does not matter: any proposal $x > 0$ can be made in equilibrium, but it will be vetoed or lead to exit, resulting in the same outcome as if $x = 0$ had been proposed. The outcome again depends on θ_W. If $\theta_W \geq -w + \gamma$, the equilibrium payoffs are $\pi_C = \theta_C, \pi_W = \theta_W$. Thanks to the veto right, the core cannot implement new policies harming the weak. If $\theta_W < -w + \gamma$, the weak exit and the payoffs are $\pi_C = -d - \gamma, \pi_W = -w + \gamma$. Note that assuming $\theta_W > 0$, exit can only occur if there is a severe commitment problem ($\gamma > w$).

Finally, Figure 3 presented a model with a constitutional free exit right. In the second stage, assuming W accepts when indifferent, it accepts as long as $\theta_W - x \geq \gamma$ or $x \leq \theta_W - \gamma$. C's equilibrium strategy is to propose $x = Max\{0, \theta_W - \gamma\}$. If $\theta_W - \gamma \geq 0$, the core proposes $x = \theta_W - \gamma$ and the weak accept, resulting in $\pi_C = \theta_C + \theta_W - \gamma$ and $\pi_W = \gamma$. If $\theta_W < \gamma$ any proposal $x \geq 0$ can be made in equilibrium, the weak exit and the payoffs are $\pi_C = -\gamma, \pi_W = \gamma$. In keeping with the idea of a commitment problem, the weak are benefitting from defection while the core remains bound. If the model had allowed it, the core would of course then exit as well, leaving both with the status quo prior to the creation of the IO.

Note that with perfect information (as assumed here) strategic exit threats by W in order to obtain $x < \theta_W - \gamma$ would not be subgame perfect. Others in the literature have explored in more detail the implications of asymmetric information in models of exit or secession (Anesi 2012; Gradstein 2004; Olofsgård 2004).

A second important observation is that, in this model, a free exit right only increases the occurrence of exit if the IO could not have formed voluntarily, i.e., $\theta_W < \gamma$. For voluntary IOs, a free exit right merely provides a protection against undesired policy changes: the possibility of exit increases the amount of effective voice (Hirschman 1970). The equilibrium payoffs are summarized in Table 3.

Table 3: Expected ex-post payoffs from the IO.

	$\theta_W < \gamma - w$		$\gamma - w \leq \theta_W < \gamma$		$\gamma \leq \theta_W$	
	$\boldsymbol{\pi_C}$	$\boldsymbol{\pi_W}$	$\boldsymbol{\pi_C}$	$\boldsymbol{\pi_W}$	$\boldsymbol{\pi_C}$	$\boldsymbol{\pi_W}$
No veto/exit right	$-d-\gamma$	$-w+\gamma$	$\theta_C+\theta_W+w-\gamma$	$-w+\gamma$	$\theta_C+\theta_W+w-\gamma$	$-w+\gamma$
Veto right	$-d-\gamma$	$-w-\gamma$	θ_C	θ_W	θ_C	θ_W
Exit right	$-\gamma$	γ	$-\gamma$	γ	$\theta_C+\theta_W-\gamma$	γ

We now turn to the generalizability of the results and the implications of the models' simplifying assumptions. The main assumption to be discussed here is that there are only two types of members: core and weak. Relaxing this assumption would introduce many degrees of freedom and hence require a host of further assumptions and/or a large number of scenarios to be analyzed.[4] In the end, it stands to reason that what would matter for each type would still be whether it had a veto and/or exit option to protect itself against majoritarian coalitions that would be able in equilibrium to implement policy changes resulting in a negative payoff for that type.

A second assumption to be discussed is that exit based on an exit right is free. In the models of complete information presented here, the core will propose x to make the weak just indifferent between accepting and exiting. It follows that if there would be a cost of exiting even if there is an exit right, then the expected payoff of the weak would be equal to minus that cost, and the weak would not join voluntarily. However, in reality the core will likely not fully exploit the weak (because of norms or incomplete information), and even an exit right with some transaction costs could be sufficient to guarantee the weak a positive expected payoff.

Taken together, these results imply that weak states could be attracted to an IO with (1) a set of initial policies such that under these policies they would benefit, i.e., $\theta_W \geq \gamma$, and (2) a policy

[4] In particular, one would have to specify how policy changes map to payoffs for each type, how coalitions can be formed (who proposes and responds in which order), and how each member's benefits depend on the presence of each other type.

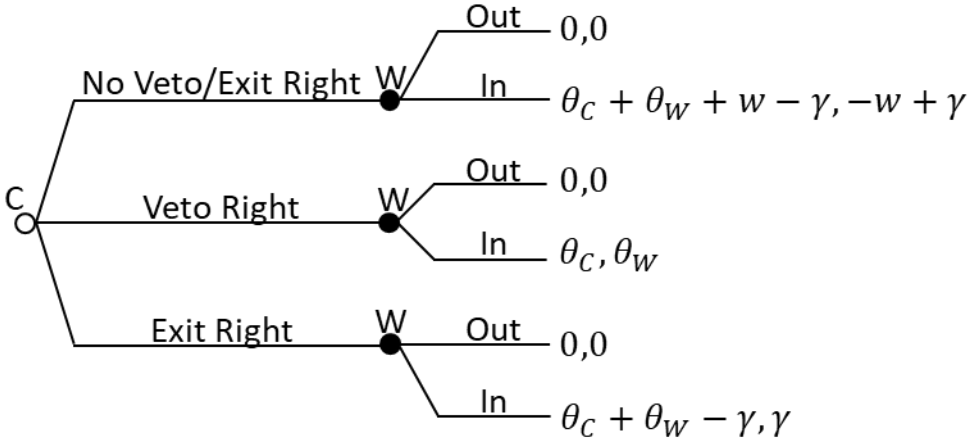
veto or a constitutional exit right. The next subsection presents a model of constitutional choice to illustrate the second implication.

3.3 A model of constitutional choice under voluntary accession

The payoffs generated by the different rules as presented in Table 3 raise the question of which rule set will be adopted. To address this question, Figure 4 presents a model of constitutional choice under voluntary IO accession. The figure represents IOs with initial policies that strongly benefit the weak, i.e., the third column of Table 3, for which $\theta_W \geq \gamma$. The other cases can be analyzed similarly, as will be discussed later.

In this model, the set of initial policies is assumed to have been agreed on, but the IO has not formed yet. At the first node of the model, the core has the choice between proposing the three different constitutional setups discussed before: "No Veto/Exit Right", "Veto Right", or "Exit Right". Next, the weak either choose "Out", in which case the IO does not form, or "In", in which case the IO is formed under the constitutional setup proposed by the core. The payoffs under each constitutional setup are those derived from the three models presented before.

Figure 4: A model of constitutional choice for an IO with $\theta_W \geq \gamma$.



Solving this model by backwards induction, the weak always accept the constitutional setups "Veto Right" and "Exit Right" because these give a positive payoff from the IO. The setup "No Veto/Exit Right" will be refused unless there is a severe commitment problem and $\gamma \geq w$. The

commitment problem is a form of protection for W, because the temptation to exit is higher and hence C needs to keep the weaks' equilibrium payoff at a higher level.

In case there is no severe commitment problem (and keeping in mind we are solving the model for $\theta_W \geq \gamma$), the core proposes an exit right, since this yields a payoff of $\theta_C + \theta_W - \gamma$ rather than only θ_C, the payoff of proposing a veto right. The outcome of the unique subgame perfect equilibrium of this game is hence that the core proposes an exit right, and that the weak accept.

If there is a severe commitment problem, the core proposes no veto/exit right, since this is now acceptable to the weak (since $-w + \gamma \geq 0$) and it yields a payoff of $\theta_C + \theta_W + w - \gamma$ rather than only $\theta_C + \theta_W - \gamma$, the payoff from proposing an exit right.

Alternative models of constitutional bargaining could have the weak propose the constitutional setup, and allow for counteroffers. We opted to model the core as making a take-it-or-leave-it offer to reflect its presumably stronger bargaining position at IO creation (Milewicz and Snidal 2016). Under this assumption, the core will successfully propose a free exit right in the absence of a severe commitment problem. Conversely, if one would model the weak as making a take-it-or-leave-it offer, then they would successfully propose a veto right (unanimous voting), since they prefer having a veto right over an exit right, and a veto right leaves a positive payoff for the core. In any case, regardless of the bargaining model one assumes, weak states will not join without either an exit right or a veto, unless there is a severe commitment problem.

The other two cases shown in Table 3 can be analyzed similarly. In particular, if the benefit is very low as in the first column ($\theta_W < \gamma - w$) the core has no interest in forming the IO because it would not be stable. If the benefit is intermediate as in the second column ($\gamma - w \leq \theta_W < \gamma$), the core does not want to give an exit right because it will be used. Unless the commitment problem is severe, the core will successfully propose a veto. If the commitment problem is severe, it will successfully propose no veto/exit right.

To summarize, it was established that with a severe commitment problem, no veto/exit right needs to be given, because the commitment problem implies a temptation to exit that the weak needs to be compensated for anyway in terms of its benefit from the IO. When the commitment problem is not severe, an exit right will be the core's preferred offer if the benefit of the IO is high ($\theta_W \geq \gamma$), and a veto right if the benefit of the IO is limited ($\gamma - w \leq \theta_W < \gamma$). The conclusions of the models are summarized in Table 4.

Table 4: Summary of theoretical expectations regarding IO design.

Commitment problem	**IO benefit**	**Veto & exit**
Severe ($\gamma \geq w$)		No veto, no exit right
Not severe ($\gamma < w$)	High ($\theta_W \geq \gamma$)	No veto, exit right
	Low ($\theta_W < \gamma$)	Veto (unanimity), no exit right

3.4 Hypothesis for empirical testing and alternative explanations

The game-theoretical models centered on individual states and groups of states. At the IO level, the models imply that weak (i.e., materially weak or peripheral states) will ask for an exit clause when decision-making is non-unanimous. The intuition is that when the voting rules will not protect weak states, they seek other forms of formal protection before agreeing to join. We therefore derive the following hypothesis:

Hypothesis: *IOs with non-unanimous voting are more likely to have exit clauses.*

Note that this hypothesis is not claimed to be causal. In fact, as in the theoretical models, voting rules and exit modalities are both determined at creation of the IO – by underlying characteristics of cooperation that are hard to observe and operationalize empirically, such as the benefits from the IO. Process-tracing may map out actual negotiation trajectories for individual IOs (e.g. Huysmans 2019; Zamora 1980). For our empirical tests, we will treat the decisions as simultaneous and only look for the theoretically expected combination of exit clauses with non-unanimous voting.

What about exogenous shocks and the *rational design argument* of uncertainty requiring flexibility? Would they lead to alternative hypotheses or be observationally equivalent to our theory focusing on power and voting rules? When member states worry about asymmetric exogenous shocks, they fear being the only ones to be negatively affected by a shock, without being able to obtain mitigating policy changes. This means that the fear of asymmetric shocks will lead all members to desire an exit clause, irrespective of the voting system. Hence exit clauses as a mere functional response to asymmetric shocks would be consistent with the null hypothesis. When member states worry about symmetric exogenous shocks that affect everyone similarly, they know that others will want to implement mitigating policy changes as well. Their main fear would be not being able to do so, because of a blocking minority. The fear of symmetric shocks will thus lead states to prefer majority decision-making, to be able to be

responsive to shocks. Hence symmetric shocks are unlikely to drive a demand for exit clauses unless the IO takes decisions by unanimity. This is the opposite of our hypothesis. In conclusion, our hypothesis would be able to distinguish between our power and voting-rule based theory of exit clauses on the one hand, and a more basic functionalist theory of exit clauses as response to uncertainty on the other hand.

Leaving the rational design framework altogether, one may wonder about *historical or symbolic drivers* of voting and exit clauses. Newly independent states may join IOs to signal their newfound sovereignty, without a rational cost-benefit calculation underlying the decision. However, in this framework it seems more likely that they would seek both a veto and an exit clause, rather than being content with either. This would go against our hypothesis. Similarly, authoritarian states may only join IOs if they feel their sovereignty is protected – ideally by both unanimous voting and an exit clause.

Another alternative explanation for design patterns is *institutional isomorphism* (Meyer and Rowan 1977; Powell and DiMaggio 1983): organizations copy the design of other prestigious organizations, without there necessarily being a good fit to the circumstances or rational design behind it. However, since there is a large variety in the specifics of exit provisions (Koremenos and Nau 2010), simple isomorphism seems unlikely to be the main driver of our hypothesized design pattern.

4 Data and methods

To test our hypotheses, we draw on the dataset "Voting rules, founding membership and issue area in intergovernmental organizations, 1944-2005" compiled by Blake and Payton (2015). We extended this sample with IOs founded between 2005 and 2014 that are also included in the dataset of von Borzyskowski and Vabulas (2019), who provide info on exit clauses. Hence, our sample covers 269 IOs created between 1944 and 2014.[5]

[5] The starting point of Blake and Payton (2015) is the Correlates of War IGO database version 2.3. They coded 266 IOs, of which 4 were no longer covered in COW 3.0. Since many of our covariates are related to COW 3.0, we dropped those 4 IOs. In addition, 5 IOs in their dataset did not specify a voting rule at creation, so we also drop those. We then extend the sample with the 7 IOs founded between 2005-2014 coded by von Borzyskowski and Vabulas (2019) for which we could identify the voting rule.

Our *dependent variable* measures the presence or absence of withdrawal clauses in IOs' constitutional treaty. Our observations are thus IOs at foundation. The dependent variable is '1' if an IO's constitution contains a withdrawal clause and '0' otherwise. We source data on IO exit clauses from von Borzyskowski and Vabulas (2019) who coded the presence of exit clauses in IO constitutions for the period 1945 through 2014. Figure 5 plots the cumulated number of IOs with and without a withdrawal clause throughout our period of observation.

As Figure 5 illustrates, the divergence between the number of IOs with and without withdrawal clauses increases from the mid-1960s onwards. This coincidence with the 1969 *Vienna Convention on the Law of Treaties* (VCLT) has led us to include a dummy control for the post-VCLT era to our model instead of linear time trend. We argue that the VCLT might have made exit clauses more likely if states became aware they had to include an explicit clause if they wanted to deviate from default international treaty law on the matter.[6]

To measure *voting rules*, we draw on the coding by Blake and Payton (2015: 386): "IGOs were coded as having either unanimity (0), majoritarian (1), or weighted (2) voting rules".[7] For our main analysis, we created a dummy variable *Majoritarian* indicating the presence of majority or weighted voting rules. Thus, "1" indicates majoritarian voting rules, and "0" unanimous decision-making. We also take advantage of the fine-grained coding of Blake and Payton (2015) and calculate additional models in the robustness check section using a categorical independent variable that can take the values: unanimity, majority, or weighted voting. Figure 6 plots the accumulated share of different voting rules across our sample.

---

[6] See Helfer (2005: 1594-1595) and others (Huysmans 2019; von Borzyskowski and Vabulas 2019; Zbíral 2007).

[7] There is also a fourth category when no voting rule is specified initially. However, as only 5 IOs within our sample lack an explicitly formulated voting rule, we exclude this category from our analysis.

Figure 5. Number of IOs including and excluding withdrawal clauses in their founding documents by year (1944-2005).

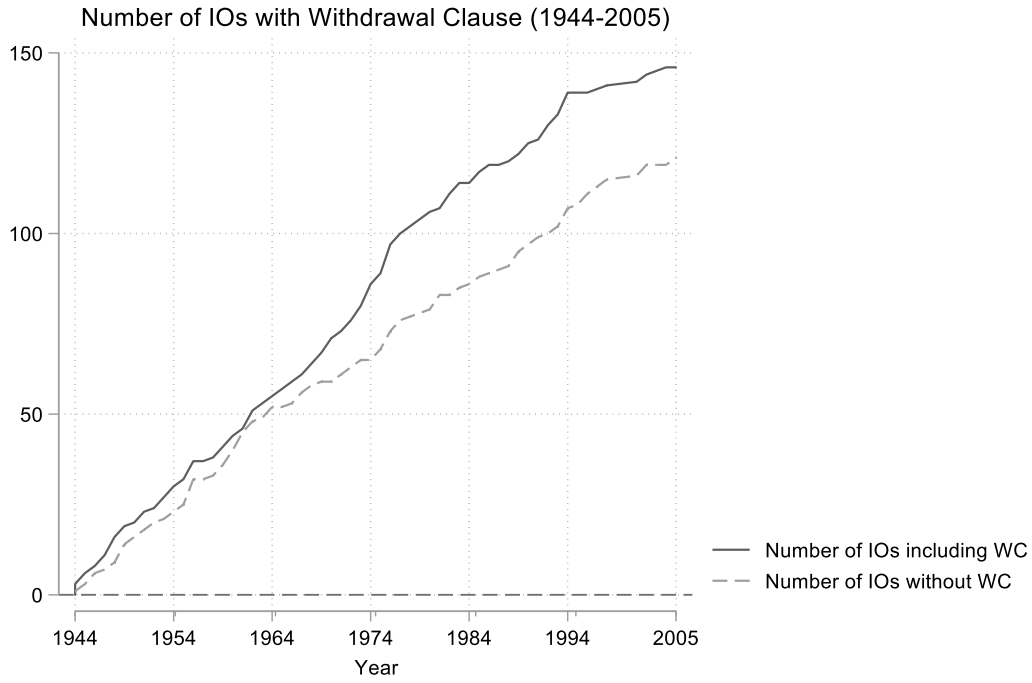
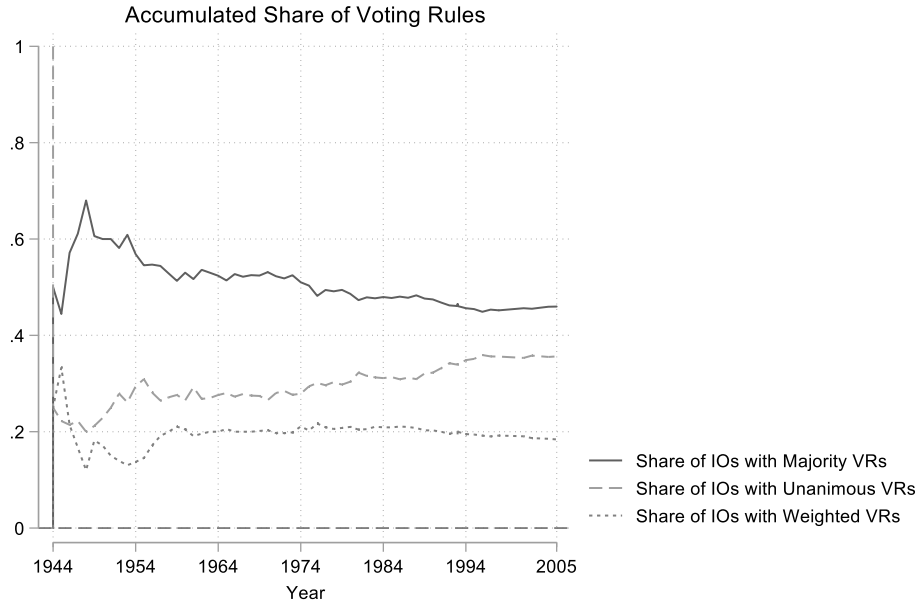


Figure 6. Accumulated Share of Voting Rules by Year (1944-2005).



Our hypothesis predicts a positive association between majoritarian voting (and weighted voting in particular) and exit clauses. Even without taking into account any control variables such as issue area, descriptive Table 5 bears out this positive association: IOs with unanimous voting are less likely to have a withdrawal clause (both in absolute and relative terms). IOs with weighted voting are most likely to have a withdrawal clause. In other words: the grey cells of the table show design equilibria, taking into account the design interaction between voting rules and exit provisions (Koremenos et al. 2001: 795). For illustrative purposes, each cell includes one example per 10 IOs in that cell.

Table 5. Number of IOs classified by withdrawal clause and voting rules, with examples.

	Voting Rule			
Withdrawal Clause	Unanimity	Majority	Weighted	Total
0	55	52	11	118
	Arctic Council	Afric. Intell. Prop. Org.	EURATOM	
	ASEAN	LatAm. Center Physics		
	Benelux Community	Org. of African Unity		
	Gulf Coop. Council	World Health Org.		
	OSCE	World Intell. Prop. Org.		
	Wassenaar Arrang.			
1	39	70	37	146
	ECOWAS	Andean Parliament	Int'l Fund Agri. Devpt.	
	Int'l Bauxite Assoc.	FAO	Int'l Grains Council	
	NAFTA	Int'l Criminal Court	Int'l Monetary Fund	
	OPEC	Int'l Whaling Comm.	World Bank	
		UNIDO		
		World Tourism Org.		
		World Trade Org.		
Total	94	122	48	264

Abbreviations: ASEAN (Association of Southeast Asian Nations), OSCE (Organization for Security and Co-operation in Europe), EURATOM (European Atomic Energy Community), ECOWAS (Economic Community of West African States), OPEC (Organization of the Petroleum Exporting Countries), NAFTA (North American Free Trade Agreement), FAO (Food and Agriculture Organization), UNIDO (United Nations Industrial Development Organization).

While our theory assumes that states will always be heterogenous to some extent – otherwise there would be no need for institutionalized cooperation – the level of heterogeneity can vary. The larger the heterogeneity among members, the more likely core states will take advantage of weak or peripheral states, prompting them to seek an exit clause as a condition of their accession. We thus included three different measures of IO membership heterogeneity.

First, we include a measure for *material power heterogeneity*. After all, cooperation preferences might vary according to a states' material power (Milewicz and Snidal 2016). We draw on the Composite Indicator of National Capability (CINC) which contains annual values for total population, urban population, iron and steel production, energy consumption, military personnel, and military expenditure of all state members, from 1816-2012 as developed by Singer (1987) and provided by the COW project (Pevehouse et al. 2004). To capture the heterogeneity of CINC scores among member states at creation, we calculated the coefficient of variation (CV) of CINC scores for the year of creation. We use the CV as it is a simple measure of the dispersion or heterogeneity.

Second, we include the well-known *UNGA voting* indicator, based on data compiled by Bailey et al. (2017) on states' voting behavior in the UN General Assembly (UNGA). To capture the heterogeneity of preferences for each IO at the time of its creation, we calculated the coefficient of variation of UN voting divergences from the IO average for each of the founding member of the IOs in our sample for the year of creation.

Finally, we also included *IO size* as the number of member states might be related to the heterogeneity among them (Koremenos 2016). The variable measures the (logged) number of other IO member states in the IO in the previous year.

Moreover, in line with the functionalist approach and the game-theoretical model, issue area characteristics might condition the relationship between voting rules and exit clauses. To measure whether an IO addresses a *commitment problem*, we use its institutional design as a proxy. We assume that commitment problems especially require the enforcement of IO policies, and proxy commitment problem by whether an IO entails enforcement functions (Reinsberg and Westerwinter 2021; Westerwinter 2021).

Note that our quantitative approach has intrinsic limits, and as such is complementary to in-depth qualitative case studies. For instance, the International Criminal Court allows for withdrawal under Article 127 of its governing document, the Rome Statute. However, there is arguably a commitment problem: states may wish to withdraw when investigations concern their citizens. A more in-depth look at the exit clause in this case shows a careful balance in

institutional design: under Article 127, states can withdraw subject to a one-year notice (there is an exit option as insurance), but withdrawal does not allow states to get out of ongoing cases against individuals (the commitment problem is covered). In addition, Article 121 gives additional protection to states in case of adverse policy changes: after being outvoted on a potential amendment, states can withdraw with immediate effect. Such subtleties, in this case fully consistent with our theory, but in others perhaps counter to our theory, will not be detected by our quantitative approach.

We also included information on *issue areas* as the presence and severity of commitment problems varies across issue areas (Koremenos 2016). Specifically, research on institutional cooperation indicates that the realm of security is particularly prone to commitment problems as compared to issue areas such as trade. We therefore include dummy variables that indicate whether an IO addresses an issue from the realm of security, trade, finance, and environment. Moreover, we include a variable for IOs addressing multiple issues, which therefore cannot be assigned to a specific category. We took this information from Reinsberg and Westerwinter (2021) and coded the 66 IOs that were not part of their dataset.

5 Results

To evaluate our hypothesis, we first estimate the association of the main explanatory factor, the presence of majoritarian voting rules, on the probability of exit clauses in IO constitutions. We then include controls to test our expectations about the scope of our argument and to control for alternative explanations in the literature.

5.1 Assessing the association of voting rules and exit clauses

In our main analysis, we run three logistic regression models of IO withdrawal clauses (see Table 6). Model 1 only includes our dummy voting rule variable as the independent variable. In Model 2, we add our preference heterogeneity indicators: material power heterogeneity, UN voting heterogeneity and IO size. In Model 3, we include all other control variables, i.e., a dummy variable indicating the presence (absence) of enforcement mechanisms, dummies indicating the issue area, as well as our post-VCLT era dummy.

Table 6. Three Logistic Regression Models of IO Withdrawal Clause Existence.

	Model 1	Model 2	Model 3
Majoritarian (vs. Unanimity)	0.873***	0.980***	1.398***
	(0.263)	(0.351)	(0.443)
Heterogeneity Indicators			
Material Power Heterogeneity		0.637**	0.749**
		(0.282)	(0.314)
UN Voting Heterogeneity		-1.099*	-1.210*
		(0.632)	(0.672)
IO Size		0.229	0.180
		(0.186)	(0.180)
Additional Control Variables			
Enforcement Mechanism			-0.0545
			(0.453)
Issue Trade			1.072**
			(0.465)
Issue Security			1.282**
			(0.581)
Issue Finance			1.282**
			(0.581)
Issue Environment			0.655
			(0.498)
Multiple Issues			1.228*
			(0.632)
Post VCLT			0.455
			(0.375)
Constant	-0.344	-0.986*	-2.083***
	(0.210)	(0.548)	(0.746)
AIC	365.412	228.689	219.667
BIC	372.601	244.597	253.419
Observations	264	178	171

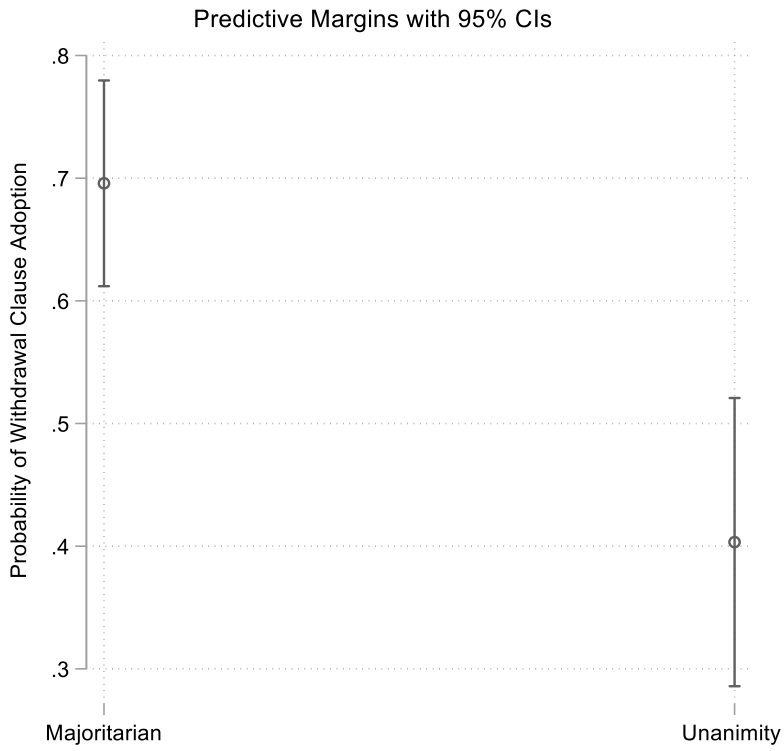
Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

In line with our expectations, the usage of *majoritarian*, i.e., non-unanimous voting rules is associated with significantly higher probabilities of an IO constitution having a withdrawal clause. As Model 2 and Model 3 demonstrate, this correlation remains strongly significant even when including all our control variables.

Turning to our other independent variables, we, firstly, find that the adoption of withdrawal clauses is more likely in IOs exhibiting higher levels of material power heterogeneity. This correlation remains significant on the 95% level in both constellations of independent variables of Model 2 and 3. This indicates that withdrawal clauses are significantly more likely in institutional settings where material power is asymmetrically distributed among the respective founding members. To the contrary, we find no significant relationship between preference constellations and the existence of withdrawal clauses. This might be because with low enough or high enough preference heterogeneity IOs are either not needed or not possible. Moreover, while the UNGA voting variable represents a general and broad measurement of state interest alignment, it can be considered a relatively rough measurement of policy-specific interest divergence among states negotiating the founding treaties of new IOs. Turning to the issue areas, Model 3 indicates that the probability of IO withdrawal clause adoption is significantly higher in the realm of trade and finance (95% level).

The independent variable of theoretical interest – the voting rules – is not only statistically significant but also has a strong substantive association with the probability of the adoption of a withdrawal clause. Figure 5 presents *the predictive margins of non-unanimous voting rules* based on Model 3 including all control variables. As illustrated by the margins plot, while the presence of voting rules is on average associated with an almost 70% probability of withdrawal clause adoption, the presence of unanimous decision-making is associated with a probability of only 40% that a withdrawal clause is adopted.

Figure 5. Predictive Margins of (Non-) Unanimous Decision-Making Rules.



Note: 95% confidence intervals.

We further calcuated conditional predictions of withdrawal clause probabilites for different constellations of voting rules and the significant material power heterogeneity indicator. In line with our theory, Figure 6 shows that the adoption of withdrawal clauses is most likely in cases where material power heterogeneity reaches its maximum and there is no unanimity. Here, the projected probability of withdrawal clause adoption amounts to more than 90 percent. To the contrary, the smallest predicted probability of withdrawal clause adoption occurs for homogenous material power distributions among members and decision rules based on unanimity. Here, the projected probability of withdrawal clause adoption is less than 20 percent.

Figure 6. Predictive Margins of (Non-) Unanimous Decision-Making Rules and different levels of material Power Heterogeneity.

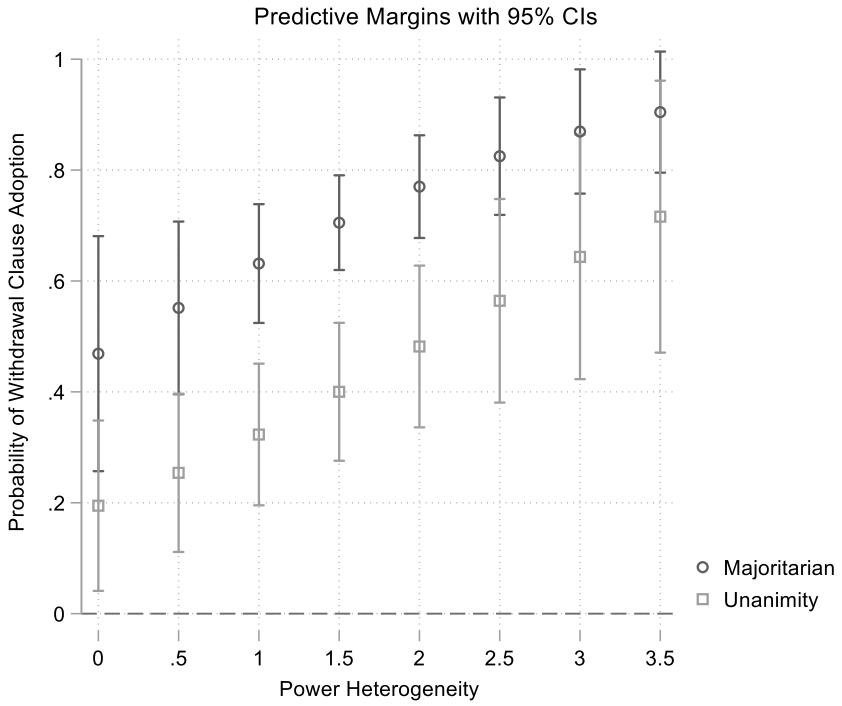
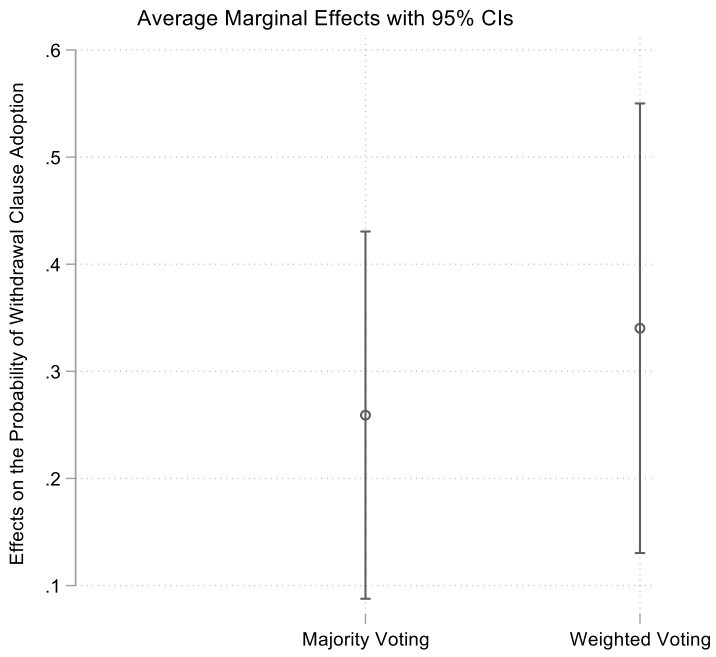


Figure 7 shows the results using the more fine-grained categorical indicator of voting rules, differentiating between unanimous, majority, and weighted voting procedures (see Online Appendix section 2). Weighted voting, implying high weights for the materially powerful, gives more risks of being outvoted for weak states than does majority voting. Hence among non-unanimous voting rules, we expect weighted voting to be more likely to be combined with an exit right than majority voting. We find that majority voting is associated with an on average 26 percentage point increase in the probability of withdrawal clause versus unanimity. In line with our theoretical expectations, weighted voting is associated with a higher increase of on average 34 percentage points.

Figure 7. Average Marginal Effects of Different Voting Rules (Baseline Unanimity).



Note: 95% confidence intervals. dy/dx for factor levels is the discrete change from the base level (Unanimity).

Moreover, we ran several additional models with alternative and additional control variables. First, we calculated a full model drawing on an alternative binary voting rule indicator as provided by Reinsberg and Westerwinter (2021) indicating the presence or absence of formalized voting rules (see Online Appendix section 4). Despite the smaller number of observations (n=160), we again find a significant and strong association of voting rule existence to exit clause probability. Furthermore, we calculated an additional model using a different construction of our indicator of preference heterogeneity. Instead of using the CV of preferences divergences from the IO average, we calculated the simple mean of preference divergences. The results remain unchanged, and the preference heterogeneity variable remains insignificant (see Online Appendix section 3). Finally, we calculated a full model which includes alternative measures of member state heterogeneity (see Online Appendix section 4). We used the GDP growth data by country and year as provided by Vabulas and Borzyzkowski (2019) and calculated its CV for each IO at the year of its creation. Moreover, for each IO, we included the CV of member states' democracy level, drawing on states' polity2 scores in the previous year sourced from the polity4 dataset (Marshall *et al.* 2010) ranging from a minimum score of -10 to a maximum score of 10, representing the highest level of state democracy. While the CV of

the GDP growth thus allows us to control for economic heterogeneity among members, the CV of member states' level of democracy allows us to check whether regime type heterogeneity matters. Despite the loss of observations due to the inclusion of both variables (n=116), the correlation between voting rule existence and withdrawal clause adoption remains strong and significant, while we do not find significant effects of the alternative heterogeneity indicators. This further increases the confidence in our results.

6 Conclusion

Powerful states tend to dominate IOs. Yet, to achieve economies of scale and legitimacy, they need to attract weak states as well. Materially weak states or those with peripheral preferences will worry about adverse policy changes if their material weakness is translated into institutional weakness in the form of low voting shares. Exit clauses provide them with insurance against future exploitation by institutionally powerful states within the IO. Prospective member states do not have to insist on an exit clause when decision-making is anticipated to be unanimous. Unanimity allows them to block unwanted policy changes in the future. Yet when voting is planned to be non-unanimous, they will want an exit clause. This is especially true when voting is weighted in favor of the powerful, rather than simply majoritarian. The empirical analysis of IOs created between 1945 and 2014 corroborated our theoretical argument. Exit clauses are more common in IOs that take decisions by weighted or majority voting than in those that make decisions unanimously.

This paper adds to scholarship interested in the power of the weak (*add references to special issue framework*). This literature emphasizes that materially weak states have several strategies at their disposal to bolster their position within IOs, such as coalition building or skilled diplomacy (*add references to respective contributions of the special issue*). Yet, these strategies play out in an already institutionalized setting that in turn shapes states' opportunities and constraints (see Barnett and Duvall 2005; Jupille *et al.* 2013). Our theory suggests that weak states will seek to improve already at the time of IO creation their institutional position in ensuing IO policymaking. By withholding their participation in an IO, weak states can push for the institutionalization of exit clauses. By making exit easier and thus more credible, the inclusion of exit clauses bolsters weak states' voice within them.

We also contribute to scholarship on rational design more generally. This literature emphasizes the role of uncertainty in the functional demand for flexibility (Koremenos et al. 2001; Rosendorff and Milner 2001; Helfer 2005; Koremenos and Nau 2010; Koremenos 2016). Yet, its focus on the functional benefits of exit clauses and their respective design tends to overlook the power politics of institutional design. This paper adds to the rational design literature by theorizing how power relations affect institutional rules and particularly how exit clauses can compensate for the lack of veto power. As an insurance for weak states against future exploitation, exit clauses might even be included where they are dysfunctional to achieve cooperation benefits.

Our theory also yields implications for states' withdrawal from IOs. Our theory predicts that, in equilibrium and under the conditions of the model, the probability of withdrawal should actually not increase with the inclusion of exit clauses. Politicians should thus not be afraid "to talk about divorce at the wedding day" but use exit clauses as a tool to increase the voice of weak states. This likely explains why, in equilibrium, exit clauses generally do not appear to make exit more likely (von Borzyskowski and Vabulas 2019). When exogenous conditions change, however, the institutional design must be adapted to prevent dysfunctional outcomes. Yet, the adaption of institutions has often proved to be difficult (Fioretos 2011; Hanrieder 2015; Jupille et al. 2013; Pierson 2000). Where exit clauses are granted but if the nature of cooperation shifts from club to public goods, or material power diffuses among the membership, the design equilibrium is upset. Unintended by its designers, states – both the weak and the powerful – are then more likely to draw on exit clauses not to increase their voice but to withdraw from IOs.

References

Alesina, A., & Spolaore, E. (2003). *The Size of Nations*. Cambridge, Mass: MIT Press.

Anesi, V. (2012). Secessionism and minority protection in an uncertain world. *Journal of Public Economics*, *96*(1–2), 53–61. http://dx.doi.org/10.1016/j.jpubeco.2011.08.004

Bailey, M. A., Strezhnev, A., & Voeten, E. (2017). Estimating Dynamic State Preferences from United Nations Voting Data. *Journal of Conflict Resolution*, *61*(2), 430–456. doi:10.1177/0022002715595700

Barberà, S., & Jackson, M. O. (2006). On the Weights of Nations: Assigning Voting Weights in a Heterogeneous Union. *Journal of Political Economy*, *114*(2), 317–339. doi:10.1086/501172

Barnett, M., & Duvall, R. (2005). Power in international politics. *International Organization*, *59*(1), 39–75. doi:10.1017/S0020818305050010

Blake, D. J., & Payton, A. L. (2015). Balancing design objectives: Analyzing new data on voting rules in intergovernmental organizations. *Review of International Organizations*, *10*(3), 377–402. doi:10.1007/s11558-014-9201-9

Cowhey, P. F. (1993). Domestic institutions and the credibility of international commitments: Japan and the United States. *International Organization*, *47*(2), 299–326. doi:10.1017/S0020818300027958

Daßler, B., Heinkelmann-wild, T., & Kruck, A. (2022). Wann eskalieren westliche Mächte institutionelle Kontestation? Interne Kontrolle, externe Effekte und Modi der Kontestation internationaler Institutionen. *Zeitschrift für Internationale Beziehungen*, *29*(1).

Fioretos, O. (2011). Historical institutionalism in international relations. *International Organization*, *65*(2), 367–399. doi:10.1017/S0020818311000002

Gould, E. R. (2016). What consensus? Explaining the rise of consensus decision-making in international organizations. *Working paper presented at PEIO X*, 1–42.

Gould, M., & Rablen, M. D. (2017). Reform of the United Nations Security Council : equity and efficiency. *Public Choice*, *173*(1), 145–168. doi:10.1007/s11127-017-0468-2

Gradstein, M. (2004). Political bargaining in a federation: Buchanan meets Coase. *European Economic Review*, *48*, 983–999.

Guzman, A. T. (2002). A Compliance-Based Theory of International Law. *California Law Review*, *90*(6), 1823–1888.

Hanrieder, T. (2015). The path-dependent design of international organizations: Federalism in the World Health Organization. *European Journal of International Relations*, *21*(1), 215–239. doi:10.1177/1354066114530011

Helfer, L. R. (2005). Exiting treaties. *Virginia Law Review*, *91*(7), 1579–1648.

Helfer, L. R. (2019). Taking Stock of Three Generations of Research on Treaty Exit. *Israel Law Review*, *52*(01), 103–117.

Hirschman, A. O. (1970). *Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations and States*. Cambridge, Mass.: Harvard University Press.

Huysmans, M. (2019). Enlargement and exit: The origins of Article 50. *European Union Politics*, *20*(2), 155–175. doi:10.1177/1465116519830202

Huysmans, M., & Crombez, C. (2020). Making exit costly but efficient: the political economy of exit clauses and secession. *Constitutional Political Economy*, *31*(1), 89–110. https://doi.org/10.1007/s10602-019-09295-1

Ikenberry, G. J. (2001). *After Victory: Institutions, Strategic Restraint and the Rebuilding of Order after Major Wars*. Princeton, NJ: Princeton University Press.

Jupille, J., Mattli, W., & Snidal, D. (2013). *Institutional choice and global commerce*. *Institutional Choice and Global Commerce*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139855990

Kindleberger, C. P. (1981). Dominance and Leadership in the International Economy: Exploitation, Public Goods, and Free Rides. *International Studies Quarterly*, *25*(2), 242. doi:10.2307/2600355

Kindleberger, C. P. (1986). Hierarchy versus inertial cooperation. *International Organization*, *40*(4), 841–847. doi:10.1017/S0020818300027399

Koremenos, B. (2005). Contracting Around International Uncertainty. *American Political Science Review*, *99*(4), 549–565.

Koremenos, B. (2016). *The Continent of International Law: Explaining Agreement Design*. Cambridge University Press.

Koremenos, B., Lipson, C., & Snidal, D. (2001). The Rational Design of International Institutions. *International Organization*, *55*(4), 761–799. doi:10.1162/002081801317193592

Koremenos, B., & Nau, A. (2010). Exit, No Exit. *Duke Journal of Comparative & International Law*, *21*, 81–119.

Martin, L. R. (1993). The Rational State Choice of Multilateralism. In J. G. Ruggie (Ed.), *Multilateralism Matters: The Theory and Praxis of an Institutional Form* (pp. 91–122). New York: Columbia University Press.

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, *83*(2), 340.

Milewicz, K. M., & Snidal, D. (2016). Cooperation by Treaty: The Role of Multilateral Powers. *International Organization*, *70*(4), 823–844. doi:10.1017/S002081831600031X

North, D. C., & Weingast, B. R. (1989). Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England. *The Journal of Economic History*, *49*(4), 803–832. doi:10.1017/S0022050700009451

Olofsgård, A. (2004). Secessions and Political Extremism: Why Regional Referenda Do Not Solve the Problem. *Journal of the European Economic Association*, *2*(5), 805–832.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Pevehouse, J., Nordstrom, T., & Warnke, K. (2004). The correlates of war 2 international governmental organizations data version 2.0. *Conflict Management and Peace Science*, *21*(2), 101–119. doi:10.1080/07388940490463933

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

Powell, W. W., & DiMaggio, P. J. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2), 147–160.

Reinsberg, B., & Westerwinter, O. (2021). The global governance of international development: Documenting the rise of multi-stakeholder partnerships and identifying

underlying theoretical explanations. *Review of International Organizations*, *16*(1), 59–94. doi:10.1007/s11558-019-09362-0

Rosendorff, B. P., & Milner, H. V. (2001). The Optimal Design of International Trade Institutions: Uncertainty and Escape. *International Organization*, *55*(4), 829–857.

Schwelb, E. (1967). Withdrawal From the United Nations: The Indonesian Intermezzo. *American Journal of International Law*, *61*(3), 661–672. doi:10.1177/106591294800100103

Snidal, D. (1985). The limits of hegemonic stability theory. *International Organization*, *39*(4), 579–614. doi:10.1017/S002081830002703X

Sommerer, T., & Agné, H. (2018). Consequences of Legitimacy in Global Governance. In J. Tallberg, K. Bäckstrand, & J. A. Scholte (Eds.), *Legitimacy in Global Governance: Sources, Processes, and Consequences* (pp. 153–168). Oxford: Oxford University Press.

Tallberg, J., & Zürn, M. (2019). The legitimacy and legitimation of international organizations: introduction and framework. *Review of International Organizations*. doi:10.1007/s11558-018-9330-7

Tsebelis, G. (2002). *Veto players: how political institutions work*. Princeton, N.J: Princeton University Press.

Viola, L. A., Snidal, D., & Zürn, M. (2015). Sovereign (In)equality in the Evolution of the International System. In S. Leibfried, E. Huber, M. Lange, J. D. Levy, F. Nullmeier, & J. D. Stephens (Eds.), *The Oxford handbook of transformations of the state* (pp. 221–236). Oxford: Oxford University Press.

von Borzyskowski, I., & Vabulas, F. (2019). Hello, goodbye: When do states withdraw from international organizations? *Review of International Organizations*, *14*(2), 335–366. doi:10.1007/s11558-019-09352-2

Walter, S. (2021). Brexit Domino? The Political Contagion Effects of Voter-endorsed Withdrawals from International Institutions. *Comparative Political Studies*. doi:10.1177/0010414021997169

Webb, M. C., & Krasner, S. D. (1989). Hegemonic stability theory: An empirical assessment. *Review of International Studies*, *15*(2), 183–198. doi:10.1017/S0260210500112999

Weiler, J. H. (1985). Alternatives to withdrawal from an international organization: the case of the European Economic Community. *Israel Law Review*, 282–298.

Westerwinter, O. (2021). Transnational public-private governance initiatives in world politics: Introducing a new dataset. *Review of International Organizations*, *16*(1), 137–174. doi:10.1007/s11558-019-09366-w

Zamora, S. (1980). Voting in International Economic Organizations. *American Journal of International Law*, *74*(3), 566–608. doi:10.2307/2201650

Zbíral, R. (2007). Searching for an optimal withdrawal clause for the European Union. In M. Niedobitek & J. Zemánek (Eds.), *The constitutional treaty: A critical appraisal* (pp. 308–323). Berlin: Duncker & Humboldt.

Zürn, M. (2018). *A theory of global governance: Authority, legitimacy, and contestation*. Oxford: Oxford University Press. doi:10.1093/oso/9780198819974.001.0001

Appendix

In this Appendix, we provide descriptive statistics of all variables used in our models as well as the result of additional statistical models. These include the following sections:

A.1. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Voting Rule categorical	264	1.826	.714	1	3
Unanimity Dummy	269	.349	.478	0	1
Power Heterogeneity	234	1.456	.684	.049	3.476
Preference Hetereogeneity (CV)	185	.677	.258	0	1.583
Preference Hetereogeneity (mean)	262	.607	.069	.095	.795
IO Size	205	2.171	1.144	0	4.905
Enforcement Mechanism	195	.21	.409	0	1
Issue Trade	269	.257	.438	0	1
Issue Finance	269	.123	.329	0	1
Issue Security	269	.078	.269	0	1
Issue Environment	269	.16	.367	0	1
Multiple Issues	269	.097	.296	0	1
Post VCLT	269	.528	.5	0	1

A.2. Models 1-3 with the categorical voting rule variable (baseline unanimity)

	(1) Model 1	(2) Model 2	(2) Model 3
Baseline Unanimity			
Majority Voting	0.641**	0.751**	1.315***
	(0.278)	(0.371)	(0.474)
Weighted Voting	1.557***	1.550***	1.798***
	(0.402)	(0.500)	(0.595)
Power Asymmetry		0.649**	0.728**
		(0.296)	(0.320)
UNGA voting heterogeneity		-1.032	-1.144
		(0.660)	(0.709)
IO Size		0.186	0.160
		(0.189)	(0.196)
Multiple Issues			1.154*
			(0.650)
Environmental Issues			0.672
			(0.520)
Economic Issues			1.097**
			(0.483)
Issue Security			0.327
			(0.652)
Financial Issues			1.147*
			(0.682)
Post VCLT Era			0.458
			(0.386)
Enforcement Mechanism			-0.186
			(0.499)
AIC	351.711	227.7867	221.815
BIC	362.438	246.8774	262.656
Constant	-0.344	-.964	-2.087***
	(0.209)	(0.566)	(0.750)
Observations	264	195	171

A.3. Full model with the alternative preference heterogeneity indicator

	Full Model
Unanimity	-1.412***
	(0.430)
Power Heterogeneity	0.705**
	(0.297)
Preference Heterogeneity	-0.693
	(2.600)
IO size	1.75e-07
	(1.50e-07)
Enforcement mechanism	-0.211
	(0.468)
Issue Economy	1.187**
	(0.474)
Issue Finance	1.283**
	(0.642)
Issue Security	0.491
	(0.663)
Issue Environment	0.642
	(0.513)
Mulitple Issue	1.119*
	(0.622)
Post VCLT	0.431
	(0.378)
Constant	-0.996
	(1.689)
Observations	175

A.4. Baseline model with IOs without Vrules prescribed included

	Model 1
Majoritarian (vs. Unanimity)	0.797***
	(0.261)
Heterogeneity Indicators	
Material Power Heterogeneity	
UN Voting Heterogeneity	
IO Size	
Additional Control Variables	
Enforcement Mechanism	
Issue Trade	
Issue Security	
Issue Finance	
Issue Environment	
Multiple Issues	
Post VCLT	
Constant	-0.344
	(0.210)
AIC	365.4116
BIC	372.6011
Observations	269

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

A.5. Full model with alternative voting rule indicator

	Full Model
Binary Voting Rule Indicator (R&W)	1.233**
	(0.490)
Power Heterogeneity	0.600**
	(0.295)
Preference Heterogeneity	0.287
	(2.602)
IO Size	2.66e-07*
	(1.57e-07)
Enforcement Mechanism	0.0429
	(0.483)
Issue Economy	0.576
	(0.457)
Issue Finance	1.092*
	(0.654)
Issue Security	0.0396
	(0.644)
Issue Environment	0.415
	(0.532)
Multiple Issue	0.782
	(0.670)
Post VCLT	0.0824
	(0.372)
Constant	-2.682
	(1.703)
Observations	160

Standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

A.6. Full model with additional heterogeneity controls

	Full Model
Unanimity vs. Voting Rules	-1.616**
	(0.573)
Power Heterogeneity	0.547
	(0.415)
Preference Heterogeneity	-1.734
	(0.954)
IO Size	0.312
	(0.245)
GDP Growth Heterogeneity	0.00425
	(0.217)
Regime Type Heterogeneity	0.0103
	(0.0234)
Enforcement Mechanism	0.118
	(0.638)
Trade	1.514*
	(0.595)
Security	0.225
	(0.834)
Finance	1.207
	(0.772)
Environment	0.161
	(0.708)
Multiple Issues	1.472
	(0.882)
Post VCLT	0.831
	(0.574)
Constant	-0.799
	(0.981)
Observations	116

Standard errors in parentheses
*** $p<0.001$, ** $p<0.01$, * $p<0.05$