What's Left Out and Why?

Informal Provisions in Formal International Law

Barbara Koremenos

University of Michigan

So far, the focus on formal law, including its details and their consequences, has dominated the political science and law literatures on international institutions. This is a tremendously important step in deepening our understanding of why and how international cooperation occurs. Scholars have shown repeatedly that the small details of formal international law, even the final clauses, matter. But this very focus has precluded what could be a very significant issue: Stuff left out! More specifically, I argue that important provisions may be left out not because they are redundant or nonessential, but by design. For instance, what made states leave out punishment provisions in the Nuclear Nonproliferation Treaty but articulate them in great detail in the series of International Coffee Agreements that began in the early 1960s?

In this paper, I present a theory of informal punishment provisions in formal international law as well as a research design for testing the theory that exploits a random sample of international agreements. This research design can be exported to study other potentially informal provisions within formal international law. The results on punishment provisions presented here provide further evidence that the details of law are chosen rationally, including the details of what's left out.

Few authors thus far have considered the role of informalism in international institutions. It should be noted that informalism is understood here as the absence of entire design elements, such as punishment provisions. This definition distinguishes this paper from earlier scholarship, like Schachter (1977), who lists characteristics of legally non-binding agreements; in other words, the agreement as a whole can be considered an informal one (Aust, 1986, follows the same definition). In particular, Schachter mentions that imprecise or overly general wording in international agreements are often taken as indicative of non-binding intention and a generally low level of legal obligation. Schachter's account squares well with Abbott and Snidal's (2000) conceptualization of soft law, which is characterized by weaker legal obligations, less precise wording, and less delegation than an ideal-type hard law agreement. In this sense, international agreements falling into the category of soft law are a form of informal international agreements. By contrast, informalism as defined here means that certain design provisions are intentionally left out of an agreement. The agreement itself can still be worded very precisely and hence create strong legal obligations; nonetheless, at the time the agreement was negotiated, some subset of the agreement was left out to be regulated informally in the future.

More recently, the literature has shifted emphasis to the relationship between formal and informal agreements. Cogan (2009) examines informal agreements in the selection of international bureaucrats: The heads of the International Monetary Fund and the World Bank, for instance, by convention are chosen from a European country and the United States, respectively. But even more generally, Cogan (2009, 211) argues, "tacit understandings that assign representation to certain states or groups of states are the norm, not the exception." Moreover, Cogan (2009, 212) provides a very clear rationale for why informal agreements enjoy such a prevalence in the international system: "informal agreements largely take account of, and reallocate authority to match, the differences in power and interests that pervade the international system when those differences cannot be acknowledged formally."

Many political scientists should be familiar with such an interpretation, yet to date there is very scant research on informalism in international institutions. While Downs and Rocke (1990) as well as Lipson (1991) have noted that some cooperation is optimally left informal, none of their

insights, to my knowledge, have been systematically tested or theoretically refined, perhaps because of the obstacles to quantifying what is informal. An exception is Stone's (2010) analysis of informal governance in international institutions. Stone explicitly considers the relationship between formal and informal governance and concludes that scholars "generally failed to connect the dots, because they have not appreciated that informal governance mechanisms exist primarily to serve the interests of powerful states, while formal rules are generally designed to protect the weak."

This paper is most closely connected to Stone's work, in that it examines the role of informal provisions in international agreements. This paper examines (the absence of) inducements to compliance in international agreements as an example of informalism in international institutions. Two questions guide the paper. First, in instances where we see no inducements, how can we distinguish between situations for which their absence is part of the informal design and those for which they are not intended at all? Second, under which conditions do we see informal design? The two questions are tightly connected, and to answer them, a combination of large-n quantitative analyses and detailed case studies is suggested. One result coming out of the analysis is that the omission of inducements to compliance need not imply meaningless agreements without enforcement power – a conclusion quite compatible with recent findings by Simmons (2009) that human rights agreements without formal enforcement mechanisms may still facilitate or instigate mechanisms for enforcement outside of the formal agreement.

Things Left Out – Useless or Hidden Uses?

Why are specific design provisions sometimes left out of international agreements? There are three potential explanations. First, a specific provision might not serve any purpose because the situation doesn't call for it. For instance, consider dispute resolution procedures. As Koremenos (2007) shows, including dispute resolution provisions into agreements is a deliberate choice in response to certain cooperation problems (see Koremenos and Betz, 2011a, for an extended

discussion). Put differently, the efficient design of international law implies that unnecessary design elements will be left out.

A second explanation is that other design features may fill their place, and hence one design element is substituted for another. Extremely precise wording, for instance, may render dispute resolution mechanisms for agreement interpretation obsolete (see Koremenos 2011b).

The third explanation is what this paper is about: Despite being useful for the underlying problem structure, states may deliberately leave out parts of an agreement intentionally in an attempt to gain, for instance, flexibility or to accommodate heterogeneous parties; this paper argues that in such cases, the potential for punishment is implicitly understood.

The Nuclear Non-Proliferation Treaty (NPT) is a striking example of this third category. The NPT does not include formal punishment provisions, yet participating states face severe enforcement problems, given the potentially devastating consequences but vast economic and political gains from the (mis)use of nuclear technology. Thus, explanations in the first category based on (the threat of) punishments not being necessary to the problem at hand are certainly not applicable. Similarly, the NPT does not exhibit any functional substitutes for enforcement provisions – the second category. Hence, the third category of explanations, informalism, is a viable candidate for explaining the lack of enforcement provisions in the NPT.

To identify agreements in this third category – agreements where design provisions are to be expected, but left out nevertheless – we first need a theory of when these design elements are necessary to the problems being solved. In this paper specifically, we first need a theory of inducements to compliance in international agreements. The theoretical framework for this paper will be Rational Design, as developed in Koremenos, Lipson, and Snidal (2001) and extended in Koremenos (forthcoming). Rational Design is based on micro-foundations and grounded in the game-theoretic analysis of international cooperation. With a theory of inducements to compliance in hand, we then want to ask whether these inducements should be formal or informal.

The research design employed here is, first, to develop hypotheses about when punishment provisions are needed, and, second, to test these empirically with a random sample of international agreements coded for both the underlying cooperation problems (like enforcement problems that give states incentives to defect) as well as design elements (like formal punishment provisions). The model predicts whether such provisions should be included in any specific agreement, based on its underlying characteristics and characteristics of the parties to it. Third, and most important, if the empirical model 'overpredicts,' (i.e., theoretically an agreement should include inducements to compliance, but empirically it does not have any), we potentially have a case of informalism, which then can be subjected to further inquiry – both through systematic comparisons of agreements that include the formal punishment provisions and those which should but do not and through detailed case studies.

The data to conduct these evaluations come from Koremenos' Continent of International Law (COIL) research project, which includes a random sample of 234 agreements drawn from four issue areas. COIL will be described in more detail below.

A Theory of Punishment Provisions in International Agreements

In general, we can identify three broad categories of inducements to compliance: assistance, rewards, and punishments. To assess their relative empirical prevalence, I will draw on the Continent of International Law (COIL) project, which in addition to theoretical development features data collection on a random of sample of international agreements. COIL is a multi-year project that assembled data on 234 agreements from the United Nations Treaty Series (UNTS), a database that includes all agreements registered or filed with the UN Secretariat since 1946 as well as many agreements registered with League of Nations. Importantly, such registration is a prerequisite to invoke an agreement before any organ of the United Nations, which creates large incentives for states to register agreements.

COIL, following UNTS definitions, focuses on four major issue areas: economics, the environment, human rights, and security. Conditional on these four issue areas, 234

agreements were drawn randomly. This choice was in part motivated by the extant literature, which typically compares agreements within specific issue areas, as in Mitchell's (2002-2011) database of International Environmental Agreements or the Alliance Treaty Obligations and Provisions data set of Leeds *et al.* (2002). More significantly, the choice was motivated by COIL's theoretical premise that issue areas are comparable once one looks at the set of underlying cooperation problems that brought states to the negotiating table. The data set contains 103 economics agreements, 43 environmental agreements, 41 human rights agreements, and 47 security-related agreements.

For the purposes of COIL, every agreement in the UNTS is considered an international agreement unless it is excluded by one of the following five criteria. First, an agreement's sole goal must not be simply to establish the procedures of the negotiations of other agreements. Second, an agreement must involve at least two states. Thus, agreements between a single state and an international organization were excluded. (This constitutes the biggest category of excluded agreements.) Third, agreements must prescribe, proscribe, or authorize behavior that is observable at least in principle. Thus, agreements not specific enough to include (at least potentially) objective criteria for determining performance are excluded – few, if any, agreements were excluded based on this criterion. Fourth, agreements must not solely implement the provisions of already existing agreements. Fifth, and following UNTS definitions, agreements that simply extend existing agreements through time are not counted as separate international agreements. A complete description of the sample and the sampling as well as coding procedures can be found in Koremenos (forthcoming).

The empirical part of COIL is the wealth of data collected, covering hundreds of details of each international agreement in the sample. The questions cover, for instance, the underlying cooperation problem(s); the main prescriptions, proscriptions, and authorizations; the existence of preambles, annexes, or appendices; precision; voting rules; membership criteria as well as any mention of nonstate actors; dispute resolution; monitoring; and flexibility provisions like duration, escape clauses, withdrawal clauses, and reservations. For the purposes of this paper, the most important variables pertain to whether an agreement has any inducements to

compliance, and if so, which ones – assistance, rewards, or punishments? (An agreement may, of course, have more than one type of inducement to compliance.) Table 1 displays descriptive statistics on this question.

Following the managerial school of Chayes and Chayes (1993), assistance to parties who might otherwise have difficulty complying should be one of the most important and most effective aids to foster compliance. Yet, and from this perspective surprisingly, in the COIL dataset only six agreements provide for such assistance; moreover, three of these agreements also include punishment provisions.[1] This low occurrence of assistance might be a consequence of informalism itself, or a manifestation of states' bleak perspective on assistance as a remedy to non-compliance. In either case, the data offers too little variation in order to build a meaningful model of this type of inducement to compliance. The same applies to rewards, with only one agreement in the sample including rewards for compliant behavior.[2]

In contrast, punishment provisions occur frequently in international agreements. In most cases, punishments are conducted by agreement members themselves or by bodies created by the agreement. However, punishments may also be delegated to already existing international institutions. In the context of punishment provisions, the most relevant such institution is the United Nations Security Council (UNSC). The UNSC is a source of punishments in multiple

---

[1] The six agreements mentioning assistance are the "International Convention on oil pollution preparedness, response and cooperation, 1990" (UNTS 32194), the "Montreal Protocol" (UNTS 26369), the "Agreement on the conservation of cetaceans of the Black Sea, Mediterranean Sea and contiguous Atlantic area" (UNTS 38466), the "International Coffee Agreement" from 1994 (UNTS 31252), the "Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction" (UNTS 33757), and the "United Nations Convention against Transnational Organized Crime" (UNTS 39574). The latter three agreements include punishment provisions in addition to assistance.

[2] This agreement is the "Convention on the prohibition of the development, production and stockpiling of bacteriological (biological) and toxin weapons and on their destruction" (UNTS 14860), and as fn 3 below demonstrates, is also coded as having punishment.

agreements in the sample, via states having the right to complain formally to the UNSC, and consequently these agreements are coded as having formal punishment provisions.[3]

The third column of Table 1 demonstrates the vast occurrence of punishment provisions in international agreements, especially when viewed in proportion to the other two inducements to compliance. The results square well with Thompson (2009), who argues that the presence of collective punishments and sanctions is a crucial factor in explaining compliance with international agreements. At the same time, a closer look at Table 1 reveals that the incidence of punishment provisions varies vastly across issue areas. Nearly half of all agreements in the issue area of economics and more than a third of human rights agreements contain punishment provisions; this share is much lower for environmental agreements, with less than 9%, and security agreements, with 17%.

Overall, given the low occurrence of assistance and rewards, the analysis in the following will be restricted to punishment provisions in international agreements as the sole inducements to compliance.

---

[3] For example, punishments were delegated to the United Nations Security Council in The "Convention on the prevention and punishment of the crime of genocide" (UNTS 1021), the "Convention on the prohibition of the development, production and stockpiling of bacteriological (biological) and toxin weapons and on their destruction" (UNTS 14860), the "Treaty on the prohibition of the emplacement of nuclear weapons and other weapons of mass destruction on the sea-bed and the ocean floor and in the subsoil thereof" (UNTS 13678), and the "Convention on the prohibition of military or any other hostile use of environmental modification techniques" (UNTS 17119) among others. The "International Convention on the Suppression and Punishment of the Crime of Apartheid" (UNTS 14861) mentions the United Nations Security Council, but does not equip it with enforcement power; rather, the Convention references the Security Council as a source of decisions that members to the Convention are required to implement.

Table 1: Agreements with Inducements to Compliance

	assistance	rewards	punishments
economics	1 (1%)	0 (0%)	47 (46%)
environment	3 (7%)	0 (0%)	4 (9%)
human rights	1 (2%)	0 (0%)	18 (44%)
security	1 (2%)	1 (2%)	8 (17%)
total	6 (3%)	1 (0%)	77 (33%)

When would we expect to see such punishment provisions incorporated into an international agreement? The Rational Design and the Legalization (Goldstein et. al 2001) literature offer some theoretical conjectures. Rational Design (Koremenos, Lipson, and Snidal, 2001) starts from the observation that international institutions display dramatic variation in their design. To understand this variation, Rational Design relies on game-theoretic insights, which implies taking preferences and constraints seriously, in order to relate cooperation problems to specific design provisions. The premise is that states and other actors on the international scene design institutions through purposeful, rational interactions. As a consequence, Rational Design offers a systematic account of the relationship between certain design features and cooperation problems. Cooperation problems, in turn, reflect the constellation of preferences and constraints in any given situation; Koremenos, Lipson, and Snidal (2001) as well as Koremenos (forthcoming) provide a more extensive treatment.

A crucial part of COIL was the coding and rigorous definition of such cooperation problems. The instrument used for the coding procedure identifies more than a dozen different cooperation problems, one or more of which may be present in any international agreement. Several of these are theoretically connected to the inclusion of punishment provisions in international agreements. (Importantly, a critical part of the COIL coding endeavor was the separation of coding for the cooperation problems (often the independent variables in many analyses) and the design elements (the dependent variables). Two independent sets of coders were employed.

A prime condition calling for punishment provisions is the existence of an enforcement problem: when the incentives to defect are large, states want to insure themselves against being the 'sucker' by being able to punish the defector. In an abstract sense, Enforcement problems correspond to the celebrated Prisoners' Dilemma. Both parties would do better by cooperating, but each individually has an incentive to cheat – and this strictly dominating strategy of cheating results in the outcome of both parties defecting, which constitutes the worst possible outcome for both. One way to address such enforcement problems is to impose severe and credible sanctions on defectors. Lowering the noncooperation payoff of the original game, defections thus become less attractive for each party, and hence mutual defections are less likely to occur.[4] Expecting the maintenance of cooperation, states are thus willing to sign onto agreements that would be infeasible in the absence of punishment provisions. Hence, a first conjecture is that, *other things equal, the presence of enforcement problems results in the inclusion of punishment provisions.*

Domestic commitment problems constitute another factor that should lead, other things equal, to the inclusion of punishment provisions in international agreements. A commitment problem arises if an actor's current optimal plan for the future is no longer optimal once this future arrives; in other words, the current plan is inconsistent over time. The most prominent example of such commitment problems pertains to a central bank's monetary policy (Barro and Gordon, 1983), but a simpler example is constituted by dieting: Promising not to eat any chocolate cake the next day might constitute an optimal plan right after a feast that involved huge quantities of chocolate cake; however, once the next day arrives, chocolate cake might appear much more luring, and the commitment to not eat any becomes much less appealing. Absent any mechanisms to tie one's hands, therefore, the commitment to not eat any chocolate cake is not

---

[4] Note that this is different from repeated play, where given a sufficiently-long shadow of the future, defections can be prevented by grim trigger strategies. Rather, imposing sanctions on defectors corresponds to changing the payoffs of the underlying game, thus turning a Prisoners' Dilemma into a game where the dominant strategy is to cooperate (and, when using Nash Equilibrium as the solution concept, mutual cooperation arises as the equilibrium outcome, instead of mutual defection as in the original game).

credible. In the realm of politics, commitment problems often arise out of domestic politics. In states that have not solved the credible commitment problem through domestic institutions, changes in the political support base of a government, for instance, may make it harder to comply with international commitments and thus trigger violations. Punishment provisions, whose negative consequences offset such political pressures, may then deter violations.

That states intentionally try to solve such commitment problems through 'tougher' agreements is also recognized in the legalization literature. As Goldstein et al. (2000: 393) put it, "Governments and domestic groups may also deliberately employ international legalization as a means to bind themselves or their successors in the future. In other words, international legalization may have the aim of imposing constraints on domestic political behavior." Therefore, a second conjecture is that, *other things equal, the presence of commitment problems results in the inclusion of punishment provisions.*

A third cooperation problem potentially associated with the inclusion of punishment provisions is uncertainty about behavior, which captures situations where the actions of other states are hard or even impossible to observe – for instance, because they take place at the domestic level, a level to which a third party or other states have no access. If actions by other states are hard to observe, this creates incentives to defect since defections might go unnoticed. For any given probability that a defection is detected,[5] punishment provisions thus lower the incentive to defect; therefore, as the probability that a defection is detected decreases due to uncertainty about behavior, punishment provisions become more attractive. Hence, a fourth conjecture is that, *other things equal, the presence of uncertainty about behavior results in the inclusion of punishment provisions.*

Overall, this yields three conjectures linking cooperation problems and punishment provisions. Punishment provisions are more likely to be included in international agreements in the

---

[5] This probability can, in turn, be affected by the inclusion of monitoring provisions. Thus, a close relationship exists between monitoring and punishment provisions; see Betz and Koremenos (2011b).

presence of enforcement problems, commitment problems, and under uncertainty about behavior.

Punishment provisions also become more attractive in multilateral agreements. In bilateral agreements, no coordination is necessary to punish a defector – Axelrod (1984)'s celebrated insights on Tit-for-Tat as a strategy in two-actor games impressively underscore this point. As Oye (1985, 19) points out, as the number of actors increases, the likelihood of including a state that is "too weak (domestically) to detect, react, or implement a strategy of reciprocity, that cannot distinguish reliably between cooperation and defection by other states, or that departs from even minimal standards of rationality" is increasing dramatically as well. Moreover, if punishments cannot be targeted to single defectors, but apply equally to all participants of the agreement, strategies of reciprocity become impossible to implement if states do not want to risk the breakdown of cooperation. Thus, 'spontaneous' punishments are harder to achieve in multilateral than in bilateral agreements, and coordinated, often even centralized punishment mechanisms are required in multilateral agreements. The fourth conjecture, hence, is that, *other things equal, bilateral agreements are less likely to include punishment provisions than multilateral agreements.*

Finally, issue areas may differ systematically in their propensity to have punishment provisions included given issues of renegotiation-proofness: Is the delivery of the punishment in the interest of all states other than the defector ex post? Punishment possibilities in economic agreements, for instance, are often renegotiation-proof when they are based on market incentives. An example is the series of International Coffee Agreements, where the signatories managed to include renegotiation-proof punishment procedures into their agreements (Koremenos, 2002). This is crucial for the credibility and therefore effectiveness of punishment provisions – if punishment provisions are subject to renegotiation (maybe because conducting the punishment is costly to the punishing party as well), they serve little purpose, and in particular can hardly deter violations. For instance, it is domestically and militarily costly to punish another state for human rights violations, especially since there is no material benefit to

offset such costs. Thus while states may commit to such punishments, ex post, there is an incentive not to follow through.

Therefore, to take into account such systematic differences not already captured with the variables mentioned already as well as the conditional random sampling, the empirical model will include dummies for three of the four issue areas included in COIL with the expectation that punishment in economics agreements are most amenable to renegotiation-proofness whereas human rights agreements are least amenable. (These expectations are theoretically-based, but not with the backing of a game-theoretic model like the conjectures from Rational Design.)[6]

(One issue is whether renegotiation-proofness as an explanation is more fitting at the stage where formal versus informal punishments are chosen.)

Identifying Informalism

Based on these conjectures, we can build an empirical model of punishment provisions; all of the data come from the Continent of International Law dataset. Table 2 shows coefficient estimates and heteroskedasticity-consistent standard errors from a probit regression, with the presence of formal punishment provisions as the dependent variable; the regressors are based on the conjectures mentioned above, and all are coded as dichotomous variables. For details on the coding procedures, see Koremenos (forthcoming).

Table 2: Probit Results

Dependent Variable: Presence of Formal Punishment Provisions	Coefficient	Std. Error
uncertainty behavior	0.419	0.332
enforcement problem	0.983***	0.229
commitment problem	1.381***	0.282

---

[6] Future drafts will incorporate just such a model.

bilateral agreement	-1.003***	0.298
economics	0.942***	0.362
environment	-1.032**	0.468
human rights	0.162	0.407
constant	-1.076***	0.373
N	234	
Log Pseudo-Likelihood	-86.116	

Coefficients Estimates and Heteroskedasticity-Consistent Standard Errors.

***significant at 1%, ** significant at 5%, * significant at 10%.

The conjectures delineated above perform quite well empirically. Except for the coefficient on uncertainty about behavior, all coefficients are large in substantive terms and statistically significant at the 1% level.[7] The marginal effects of the occurrence of each cooperation problem and of being a bilateral agreement are shown in Table 3. The first four columns show the marginal effects for an agreement in each of the four issue areas, where all variables but the respective issue area dummy are zero. The fifth column averages these marginal effects over issue areas, using the relative frequency of each issue area to weigh the issue-area specific marginal effects. What is immediately obvious from the table is that the marginal effects are much weaker for environmental agreements than for any of the other issue areas. Table 3 also shows that the marginal effects of each cooperation problem are large in substantive terms: on average, the presence of a commitment problem, for instance, raises the probability that punishment provisions are included by 42 percentage points. Enforcement problems rank second in terms of the magnitude of the marginal effect, on average resulting in a 30 percentage point increase in the probability that punishment provisions are included. Overall, judging from the significance and size of these marginal effects, the conjectures outlined above perform extremely well.

---

[7] Interestingly, the lack of significance on the variable, Uncertainty about Behavior, is quite consistent with Koremenos and Betz (forthcoming) that this cooperation problem is best solved with an institutional mechanism that is primarily about information clarification, not punishment.

Table 3: Marginal Effects

	economics	environment	human rights	security	total
uncertainty behavior	0.166	0.028	0.130	0.115	0.124
enforcement problem	0.355	0.113	0.347	0.322	0.303
commitment problem	0.447	0.216	0.499	0.479	0.420
bilateral agreement	-0.319	-0.017	-0.153	-0.122	-0.195

Marginal Effects for each issue (first four columns), and averaged over all issue areas, weighing by relative frequency of each issue (fifth column). Marginal effects calculated for hypothetical multilateral agreement without any of cooperation problems.

Judging by the plot of the Receiver Operating Characteristics (ROC), the model also seems to perform well for predictive purposes. The ROC is better suited to assess the goodness of fit in models with binary dependent variables than the commonly used number of observations that are correctly predicted, as the ROC does not depend on an arbitrarily chosen cutoff value. Instead, the plot of the ROC shows the fraction of agreements with punishment provisions that were correctly predicted versus the fraction of agreements without punishment provisions that were incorrectly predicted as a function of the cutoff value. For instance, at a cutoff value of 1, no agreement is predicted to have punishment provisions, hence all of those with punishment provisions are incorrectly predicted, while all of those without punishment provisions are correctly predicted. Hence, the ROC assumes a value of (0,0) at a cutoff value of 1. If the model has no predictive power, the ROC will display a straight line, and the area beneath it will be 0.5; in contrast, if the model fits perfectly, the area under the curve will be 1. For the model presented above, the area under the ROC curve is 0.903, which indicates an astonishingly good fit. The model therefore seems to perform well for predictive purposes, and agreements without punishment provisions that were predicted to have them consequently can be viewed as 'outliers,' that is, candidates for informal punishment provisions.

Specifically, based on the coefficient estimates, we can predict the probability that each

agreement in the sample should, theoretically, include punishment provisions. The predicted probabilities range from 0.0% to 96.9%, with a mean of 32.9% and a standard deviation of 0.327. Using 0.5 as cut-off, 84.6% of the agreements were predicted correctly – i.e., were predicted to have punishment provisions and in fact had them formally incorporated, or were predicted to not have any punishment provisions and in fact did not have any. Table 4 shows those fourteen agreements that were predicted to have punishment provisions with a probability of at least 50%, yet don't have any – i.e., these are the candidates for informal punishment provisions. This set of agreements, which I will refer to as *misclassified agreements* in the following, will now be subjected to further inquiry.

Table 4: Misclassified Agreements

COIL Code	Agreement Name	Pred. Prob.
HR 8	International Covenant on Civil and Political Rights.	0.812
HR 10	Convention for the Protection of Human Rights and Fundamental Freedoms.	0.812
HR 11	Convention on the Elimination of All Forms of Discrimination against Women.	0.812
HR 18	Protocol relating to refugee seamen.	0.527
HR 22	International Convention on the Suppression and Punishment of the Crime of Apartheid.	0.812
HR 25	Convention (with Final Protocol) concerning the reciprocal grant of assistance to distressed persons.	0.527
HR 2-5	International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families.	0.687
AC 57	Agreement with respect to quality wheat.	0.802
EN 48	Convention on Nature Protection and Wild Life Preservation in the Western Hemisphere.	0.750
IN 16	Exchange of letters constituting an agreement relating to investment guaranties.	0.596
SE 2-18	Treaty on Collective Security.	0.901
FN 2-13	United Nations Convention on Independent Guarantees and Stand-by Letters of Credit.	0.894
FN 2-17	International Convention for the Suppression of the Financing of Terrorism (with annex).	0.612
FN 2-208	Agreement on economic and financial cooperation between the Kingdom of Spain and the Republic of Argentina.	0.596

Agreements without punishments, yet predicted to include punishments with probability of at least 0.5.

Explaining Informalism

Domestic Enforcement Mechanisms

A first look at Table 4 reveals that human rights agreements dominate the category of misclassified agreements: they often lack formal punishment provisions when we expect to see them. Simmons (2009) offers an explanation for this pattern that is based in domestic political factors. She argues that human rights treaties become meaningful, and thereby exert compliance pressures, by empowering domestic (or transnational) individuals, groups, or political entities; these actors, then, use commitments made by governments to pressure them into obeying higher human rights standards. Moreover, international agreements may help local actors define their agendas more clearly, agree on a common set of priorities, obtain additional options for litigation, and thereby gain a more effective bargaining position vis-à-vis the government. The presence of such mechanisms renders punishment provisions in some international agreements obsolete and, as Simmons convincingly shows, especially so in human rights treaties. Not coincidentally, two of the misclassified agreements in Table 3, the Convention on the Elimination of All Forms of Discrimination against Women and the International Covenant on Civil and Political Rights, are studied in depth in Simmons' book. It will be illuminating to study the remaining agreements in Table 3 from the theoretical perspectives that Simmons and Dai offer; in particular, it will be worthwhile to see whether their respective theories apply to agreements outside the issue area of human rights as well. The Convention on Nature Protection and Wild Life could be an especially promising case, given the large number of environmental NGOs that could operate within the mechanisms that Dai and Simmons identify.

Heterogeneity among Participants

One might suspect that misclassified agreements tend to be compromised of more heterogeneous sets of states. Heterogeneity among parties to an agreement makes compromise harder to achieve. Decentralized (that is, not formally and explicitly defined)

punishment provisions thus become more attractive. Oates (1999) makes a similar point in the context of fiscal federalism: when preferences are diverse, delegating decision-making authority and fiscal autonomy to single actors becomes more attractive, relative to agreeing on centralized, uniform mechanisms. Ehrlich and Posner (1974) connect this mechanism to law-making: if compromise is hard to achieve, it becomes more attractive to leave out specifics. This argument is in line with Koremenos (2011b) who, building on Ehrlich and Posner, argues that international law-makers design law efficiently when choosing the precision or vagueness of the substantive terms. In particular, heterogeneity among participants makes it harder to agree on precise rules. Such rules could be substantive ones, as in Koremenos (2011b), or they could be procedural, such as the punishment provisions examined here.

It is important that the theoretically-based argument here is that heterogeneity among participants affects the design of punishment provisions (that is, whether they are formal or informal); in contrast, heterogeneity is assumed not to affect the likelihood that punishments provisions are required (and hence it is left out of the probit regression shown in Table 1). Some readers might object to this and argue that homogeneous groups of states are less likely to require punishment provisions in their agreements. Such a statement, however, jumps to a conclusion too quickly and implicitly assumes that states that are more alike also face fundamentally different cooperation problems than states that differ in their characteristics. There is no theoretical reason to jump to such a conclusion, however, with the possible exception of the cooperation problem Uncertainty about Preferences. Consider the prime example of an enforcement problem, the stylized Prisoners' Dilemma. Here, the two actors are as alike as they can be in a game-theoretic sense: each player faces the same problem and the same payoff structure, such that the game is symmetric. Yet, this homogeneity does not rule out the existence of an enforcement problem, and hence the usefulness of punishment provisions.

To assess the relationship between preference heterogeneity and informal punishment provisions, as a first cut I rely on Gartzke (2006)'s Affinity of Nations index. The index[8] infers

---

[8] The index corresponds to the variable s2uni in Gartzke's data.

preference homogeneity from voting patterns in the United Nations General Assembly for any two states. In Gartzke's data, the most homogeneous voting pattern is indicated by 1 and the least one by -1. Since I want to measure preference heterogeneity, I invert this scale, such that -1 indicates the most homogeneous and 1 the most heterogeneous dyad. For bilateral agreements, this measure is obtained directly from Gartzke's data set. For multilateral agreements, I first create a data set with non-directed dyads for any two participants in the multilateral agreement. For instance, in an agreement with three participants, there are three dyads: between state 1 and state 2, between state 1 and state 3, and between state 2 and state 3 (the reverse dyads are redundant). Exploiting the "weakest link assumption," I then use the affinity index for the most heterogeneous dyad within the multilateral agreement to measure preference heterogeneity within the agreement. This assumption implies that the most heterogeneous dyad is crucial for the decision whether to use formal versus informal punishments. Gartzke's data codes only for years after 1946 and contains several missing values, such that I do not have heterogeneity measures for all agreements in the sample; 30 values are missing, one of them pertaining to a misclassified agreement. This is the "Convention on Nature Protection and Wild Life Preservation in the Western Hemisphere," which was signed in 1940 before any UN voting took place.

Table 5: Preference Heterogeneity

	Obs.	Mean	Std. Error
Affinity of Nations			
misclassified	13	-0.553	0.124
all other	191	-0.639	0.025
total	204	-0.634	0.025
difference		0.086	0.102
Vanhanen			
misclassified	8	62.34	9.633
all other	120	28.93	3.656
total	128	31.02	3.546
difference		33.40	14.40
Freedom House			
misclassified	9	4.444	0.568
all other	183	1.568	0.209
total	192	1.703	0.206
difference		2.876	0.954

The top panel of Table 5 shows the mean heterogeneity measure from Gartzke and its standard error for the misclassified agreements and for all other agreements. *There is a difference in terms of heterogeneity between correctly classified and misclassified agreements and it is in the expected direction; but the difference of 0.086 is small in size and statistically not significant* (the p-value is 0.398).[9]

Another way to measure heterogeneity among participants is their level of democracy. The literature provides several such measures, and two of them are considered in the following: the Vanhanen democracy index (a 100 point scale) and the Freedom House democracy index (a

---

[9] Interestingly, however, if punishments that are delegated to the UNSC are not coded as punishments, the difference in terms of preference heterogeneity is substantial, with a value of 0.20, and statistically significant at the 5% level.

seven point scale).[10] The middle and bottom panel of Table 6 display the results from the respective measures. To measure heterogeneity, for bilateral agreements the difference in the respective democracy scores between the two participants is used. For multilateral agreements, first a data set with all dyads in the multilateral agreement was created. The difference in the democracy indices for the most dissimilar dyad then determines the heterogeneity measure for the agreement (the `weakest link' assumption). *Both measures of heterogeneity are significantly higher for misclassified agreements: informal punishment provisions are more likely to incur in agreements composed of heterogeneous sets of states. (Interestingly, the relationship does not hold when using Polity scores.)* Potentially, this informalism allows discretationary punishment strategies. It is easier to see violations by a like-minded government as not as serious and to dish out less harsh punishments accordingly if an explicit punishment is not formally mandated by the treaty text.

US Involvement

Another factor potentially influencing the choice between informal and formal punishment provisions is the involvement of the United States. The effect, however, could go in either direction. On the one hand, one might suspect that agreements with US participation are less likely to include formal punishment provisions. The US government is often particularly wary of becoming subject to international law, and as the most powerful actor in the international system, should be most, and often is, concerned about preserving its autonomy (Koh, 1997, 2615ff.; Tierney, 2008, 284). Hence, in order to maintain relative freedom in conduct, the United States should require such provisions to be left out. This view squares well with Stone (2010)'s insight cited earlier that it is formal law that protects the weak and informal law that serves the powerful. It also matches with Koh (2003, 1484)'s observation that while the United States might be more reluctant to sign onto international agreements, it is doing well in complying with its treaty obligations – and consequently is more reluctant to negotiate and ratify agreements that impose severe obligations that then are harder to obey.

---

[10] All of the data was drawn from a data set collected by Pippa Norris, which is available at http://www.pippanorris.com/. Polity scores were not significantly correlated with misclassified agreements.

On the other hand, there is some anecdotal evidence that the United States prefers formal law in many settings. An example is the series of arms control with the Soviet Union and now START with Russia; these agreements are replete with very detailed and specific verification schemes. Similarly, some argue that the United States refuses to ratify the Comprehensive Test Ban Treaty until the verification scheme is more specific (Medalia, 2008).

Finally, the United States may allow for punishment provisions, yet exempt itself via reservations; Koremenos and Betz (2011) observe such a pattern for dispute settlement procedures, and it is not out of question that a similar mechanism is at work with respect to punishment provisions.

The data provide weak evidence in favor of the second perspective. Out of the 14 misclassified agreements, 7 are without US involvement; out of the remaining 220 agreements, 151 are with US involvement. Formal punishment provisions are somewhat more likely under US involvement, but the difference is not statistically significant. It is, nevertheless, an interesting empirical result, as it contrasts with Stone (2011)'s claim that strong states should push for informal law.

A less crude and more context-specific measure for power besides a US dummy can also be obtained. In economic agreements, GDP is likely to be the most relevant variable, while in security agreements, military capabilities are the most relevant factor. Absent better measures, I will preliminarily rely on GDP also to measure power differences for environmental and human rights agreements. As it turns out, including this context-specific measure of power differences together with issue area dummies and a dummy for US involvement into a probit regression, where the dummy for misclassified agreements is the dependent variable, produces coefficient signs of the expected sign, but they are statistically insignificant:[11] US involvement is associated with no informal punishment provisions, whereas power differences are associated with informal punishment provisions. That is, *while power differentials in general tend to result in informalism, as Stone predicted, US involvement has the opposite effect – the US seems to push for*

---

[11] Leaving out issue area dummies or the US dummy does not increase the significance of the context-specific power measure.

formalism. However, none of the coefficient estimates reaches statistical significance (the p-value on the context-specific power measure is 0.243, the p-value on US involvement is 0.227).

Conclusion

This analysis offers a new research design that allows one to test hypotheses about informality across a set of international agreements, thereby adding significantly to other large-n work on informalism that rely on specific (albeit rich and detailed) case studies. It also offers a list of agreements that can be subjected to case studies that can shed light on whether informality is indeed occurring in the misclassified agreements and what it looks like.

Consider, for example, one of the misclassified agreements: The International Convention for the Suppression of the Financing of Terrorism. There is good evidence that this agreement is being enforced, despite the lack of a formal enforcement (punishment) provision. Specifically, when it was needed, a Security Council Resolution (Resolution 1373) was adopted that was based on the agreement and effectively enforces it – even though the UNSC is never mentioned in the agreement text itself. Thus it is not a large leap to say punishment was implicit despite being left out formally. This example also suggest the UNSC may play an even more important role for informal enforcement than what one would expect based on an evaluation of the treaty texts.

References

Abbott, Kenneth W., and Duncan Snidal (2000). "Hard and Soft Law in International Governance," *International Organization*, Vol. 54, No. 3, pp. 421-456.

Aust, Anthony (1986). "The Theory and Practice of Informal International Agreements," *International and Comparative Law Quarterly*, Vol. 35, No. 4, pp. 787-812.

Axelrod, Robert (1984). *The Evolution of Cooperation* (New York, NY: Basic Books).

Barro, Robert J. and David B. Gordon (1983). "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy*, Vol. 91, No. 4, pp. 589-610.

Betz, Timm and Barbara Koremenos (2011). "Information and Enforcement in International Agreements – Monitoring and Dispute Resolution Provisions," Working Paper (Ann Arbor, MI: University of Michigan, Department of Political Science).

Chayes, Abram, and Antonia Handler Chayes (1993). "On Compliance," *International Organization*, Vol. 47, No. 2, pp. 175-205.

Cogan, Jacob Katz (2009). „Representation and Power in International Organization: The Operational Constitution and its Critics," American Journal of International Law, Vol. 103, No. 2, pp. 209-263.

Downs, George W. and David M. Rocke (1997). *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations* (Princeton, NJ: Princeton University Press).

Ehrlich, Isaac and Richard A. Posner (1974). "An Economic Analysis of Legal Rulemaking," *Journal of Legal Studies*, Vol. 3, No. 1, pp. 257-286.

Gartzke, Erik (2006). "The Affinity of Nations Index, 1946-2002, Version 4.0, March 2006," available at http://dss.ucsd.edu/~egartzke/htmlpages/data.html.

Goldstein, Judith O., Miles Kahler, Robert O. Keohane, and Anne-Marie Slaughter (2000). "Introduction: Legalization and World Politics," *International Organization*, Vol. 54, No. 3, pp. 385-399.

Hafner-Burton, Emilie (2009). *Forced to be Good: Why Trade Agreements Boost Human Rights* (Ithaca, NY: Cornell University Press).

Koh, Harold Hongju (1997). "Why Do Nations Obey International Law?" *Yale Law Journal*, Vol. 106, No. 8, pp. 2599-2659.

Koh, Harold Hongju (2003). "On American Exceptionalism," *Stanford Law Review*, Vol. 55, No. 5, pp. 1479-1527.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal (2001). "The Rational Design of International Institutions," *International Organization*, Vol. 55, No. 4, pp. 761-799.

Koremenos, Barbara (2002). "Can Cooperation Survive Changes in Bargaining Power? The Case of Coffee," *Journal of Legal Studies*, Vol. 31, No. 1, pp. S259-S283.

Koremenos, Barbara (2007). "If Only Half of International Agreements Have Dispute Resolution Provisions, Which Half Needs Explaining?" *Journal of Legal Studies*, Vol. 36, No. 1, pp. 189-212.

Koremenos, Barbara and Timm Betz (2011). "The Design of Dispute Settlement Procedures in International Agreements." In Jeffrey L. Dunnoff and Mark A. Pollack (eds.). *International Law and International Relations: Synthesizing Insights from Interdisciplinary Scholarship*. Cambridge: Cambridge University Press, 2012.

Koremenos, Barbara (2011a). "Institutionalists and International Law," Working Paper (Ann Arbor, MI: University of Michigan, Department of Political Science).

Koremenos, Barbara (2011b). "An Economic Analysis of International Rulemaking," Working Paper (Ann Arbor, MI: University of Michigan, Department of Political Science).

Koremenos, Barbara (forthcoming). "The Continent of International Law," *Journal of Conflict Resolution*.

Leeds, Brett, Jeffrey Ritter, Sara McLaughlin Mitchell, and Andrew Long (2002). "Alliance Treaty Obligations and Provisions," *International Interactions,* Vol. 28, No. 3, pp. 237-260.

Lipson, Charles (1991). "Why are Some International Agreements Informal?" *International Organization,* Vol. 45, No. 4, pp. 495-538.

Medalia, Jonathan (2008). "Comprehensive Nuclear-Test-Ban Treaty? Issues and Arguments," *CRS Report for* Congress (Washington, DC).

Mitchell, Ronald B. (2002-2011). "*International Environmental Agreements Database Project (Version 2010.3),*" available at: http://iea.uoregon.edu/.

Oates, Wallace E. (1999). "An Essay on Fiscal Federalism," Journal of Economic Literature, Vol. 37, No. 3, pp. 1120-1149.

Oye, Kenneth A. (1986). "Explaining Cooperation under Anarchy: Hypotheses and Strategies," in Kenneth A. Oye (ed.). *Cooperation under Anarchy*, pp. 1-24 (Princeton, NJ: Princeton University Press).

Schachter, Oscar (1977). "The Twilight Existence of Nonbinding International Agreements," *American Journal of International Law*, Vol. 71, No. 2, pp. 296-304.

Simmons, Beth A. (2009). *Mobilizing for Human Rights. International Law and Domestic Politics* (Cambridge, MA: Cambridge University Press).

Stone, Randall W. (2010). "Controlling Institutions: International Organizations and the Political Economy," Paper Presented at the Annual Meeting of the American Political Science Association, September 2010.

Thompson, Alexander (2009). "The Rational Enforcement of International Law: Solving the Sanctioners' Dilemma," *International Theory*, Vol. 1, No.2, pp. 307-321.

Tierney, Michael J. (2008). "Delegation Success and Policy Failure: Collective Delegation and the Search for Iraqi Weapons of Mass Destruction," Law and Contemporary Problems, Vol. 71, pp. 283-31