International Theory (2014), 6:1, 28–67 © *Cambridge University Press, 2014*
doi:10.1017/S1752971913000304

Choosing anarchy: institutional alternatives and the global order

MOONHAWK KIM[1] and SCOTT WOLFORD[2]

[1] *Department of Political Science, University of Colorado at Boulder, Colorado, USA*
[2] *Department of Government, University of Texas at Austin, Texas, USA*

E-mail: moonhawk.kim@Colorado.edu

The international system may be anarchic, but anarchy is neither fixed nor inevitable. We analyze collective choices between anarchy, a system of inefficient self-enforcement, and external enforcement, where punishment is delegated to a third party at some upfront cost. In equilibrium, external enforcement (establishing governments) prevails when interaction density is high, the costs of integration are low, and violations are difficult to predict, but anarchy (drawing borders) prevails when at least one of these conditions fail. We explore the implications of this theory for the causal role of anarchy in international relations theory, the integration and disintegration of political units, and the limits and possibilities of cooperation through international institutions.

Keywords: anarchy; international institutions; order; institutional choice; cooperation

Why is the international system anarchic? In other words, why are there no common institutions with the means to enforce cooperation through the threat of punishment, like those present in domestic political systems? The rulers of states face basic problems of cooperation that are not fundamentally different from those that lead others to delegate the enforcement of rules to governments. However, international politics is characterized by a system in which rules are self-enforced by threats of costly retaliation. Why? Classic schools of thought like realism (Carr 1964; Waltz 1979), liberalism (Keohane and Nye 1977), constructivism (Wendt 1999), and even the English school (Bull 1977) take the self-enforcement of rules and agreements between states as given. While Wagner (2007) shows that this decentralized institutional order was a choice made to solve problems of cooperation, it remains unclear why other, more centralized alternatives were not chosen. Why, in other words, are some actors subject to a common government, while borders are drawn between others?

The answer to this question will have implications for the possibility of meaningful cooperation through international institutions, the formation of institutions like the territorial state and the drawing of borders, the traditional distinction between international and domestic politics, and variations in the extent to which parts of the international system are characterized by formal hierarchical relations. To explore these issues, we analyze a stylized model of institutional choice, cooperation, and punishment in the shadow of opportunistic incentives to cheat. We characterize political institutions according to one of two ideal types of punishment mechanism: (1) external enforcement or (2) self-enforcement. While the former punishes violators and offers limited compensation to victims through a set of common institutions, it requires upfront costs, whether material or subjective, to implement. The latter, on the other hand, requires no upfront investments but is costly *ex post*, relying on mutual threats of violence to punish violations of the rules. In essence, political actors face a choice between building a government and drawing a border in order to solve cooperation problems, and we provide a logic by which to understand when one institution may be chosen over the other.

Our theory suggests that anarchy prevails when either the frequency of interactions is low, when the costs of integrating under common institutions are prohibitively high, or when the probability of defection is too high or too low. On the other hand, when interactions are dense, violations are neither too likely nor too infrequent, *and* the costs of integration are not too high, players find it optimal to build common institutions and forfeit the right to use violence. Thus, the international system is anarchic because the interactions between states are relatively infrequent, given the costs entailed in yielding enforcement authority to a supra-state institution. Rather than preclude effective institutions (see Mearsheimer 1994), then, anarchy is a symptom of conditions under which common enforcement institutions are already infeasible. Thus, anarchy may be spurious to many of the international political outcomes it is often invoked to explain. Further, to the extent that states are able to secure their interests with threats of war, anarchy is an effective and self-enforcing institution in its own right. While anarchy is certainly ‘what states make of it’ (Wendt 1992), it is also a set of ‘congealed tastes’ (Riker 1980), in that states create anarchy – they *choose* it – when they select enforcement institutions.

After presenting the theoretical model and its implications, we discuss how it sheds light on processes of both integration and separation – that is, choices of self- or external enforcement – across cases as diverse as the writing of the Constitution of the United States, the formation and dissolution of Czechoslovakia, and the creation of the European Union. We then discuss the implications of our argument for the study of both international institutions and international relations more broadly. We also explore what roles international institutions can play, given that the

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

enforcement problem is solved by states' retention of their right to use violence, and what this implies about debates over the 'effectiveness' of institutions.

Anarchy and international politics

The assumption of anarchy sits at the core of virtually all contemporary approaches to international relations. Indeed, Waltz (1979) identifies the absence of world government as the primary factor distinguishing international from domestic politics. Where some theories posit the existence of hierarchy (e.g., Organski 1958; Lake 2009), their definitions can exist alongside the kind of formal, legal anarchy discussed here; anarchy is not the absence of order, but the absence of an accepted common authority with the legitimate right to use force. Anarchy is, in fact, a kind of order (see Hirshleifer 1995), one in which the enforcement of agreements is 'decentralized' (Wagner 2007), conducted by the parties to the agreements rather than some delegated third party. In the international system, this manifests itself in states reserving the right to use violence in defense of their interests, where in domestic political systems, individuals or groups agree to transfer the use of violence to the common authority of the state.[1]

International anarchy determines the possibility and depth of cooperation, because it implies that agreements must be self-enforcing in order to be successful. If states have no incentive to honor agreements, they can only be compelled to do so by threats from other states, which places limits on the depth of cooperation (Downs, Rocke and Barsoom 1996) and on the terms of viable peace agreements (Werner and Yuen 2005). Some scholars take the argument farther, arguing that anarchy precludes meaningful cooperation through international institutions (Mearsheimer 1994). However, we argue that anarchy is itself an institutional choice, a particular enforcement institution selected from a set of alternatives that represents some (albeit remarkably stable) 'congealed tastes' (Riker 1980). Just as former states forsake anarchy when they yield their sovereignty to join together in a new state, states can also *choose* anarchy when they break up into smaller ones, and the current map of the world reflects nearly 200 legally sovereign political units that have chosen to look out for themselves, as it were, in the maintenance of agreements and the protection of their interests.

[1] Lake's (2009) concept of 'relational hierarchy', in which one state yields autonomy in return for guarantees from another state, is a dyadic concept, while our notion of common institutions involves principals contracting out enforcement capacity to a common, third party agent that is distinct from the parties to the contract.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

In the few studies that view anarchy as order, the focus tends toward the development of the modern state or the survival of systems of anarchic relations. Spruyt (1994), for example, explores the role of centralization and territoriality in facilitating credible commitments between states, which explains their rise as the dominant mode of political organization in the early modern period. Wagner (2007) argues that the state emerged at the nexus of bargains between economic predators and their prey, as well as with competing predators, both internal and external. A system of territorial states emerged as a solution to problems of organized violence, which renders claims that anarchy is a either a cause of or permissive condition for war problematic (see Chapters 1–3). However, while each study draws a contrast between the world of territorial states and the conditions that preceded it, neither considers the explicit choice of external vs. self-enforcement in which we are interested.[2]

Anarchic relations are also mutable, as political actors can lose their ability to resort to punitive violence through destruction, conquest, or voluntary alienation. Hirshleifer (1995) shows that anarchic systems can remain stable, in the sense that no actors dominate or incorporate the others, when conflict is not too decisive and when threats of self-enforcement are not too costly. However, when conflict becomes more decisive, then anarchical relations can disappear into an empire created by conquest. States may also choose anarchy over common institutions voluntarily, as happened with the explosion of new sovereign states that emerged from the disintegration of the Soviet Union or South Sudan's recent separation from Sudan. Likewise, formally independent states can choose to eliminate anarchy, subjecting themselves to a common power as the original United States did in the moves from independence to the Articles of Confederation to the present Constitution of the United States.[3]

If anarchy is a choice, then it is made with some idea of its consequences relative to other institutional arrangements that are not chosen. If, for example, anarchy is chosen with an eye to underlying problems of cooperation, enforcement, and conflict, then taking it for granted in international relations theory may pose problems for judging its implications for the effects, and even the possibility, of other, more formal international institutions. Our goal, then, is to examine how and why two political units choose between decentralized self-enforcement of the rules and centralized

[2] For example, Spruyt's (1994) alternatives to the territorial state – the city-state and the city-league – still operated in nominal anarchy with respect to one another.

[3] Alesina and Spolaore (2003) discuss integration in the context of public goods provision across heterogeneous jurisdictions.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

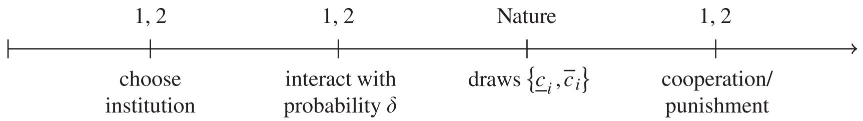
1, 2
1, 2
Nature
1, 2
choose institution
interact with probability δ
draws $\{\underline{c}_i, \overline{c}_i\}$
cooperation/ punishment

Figure 1 The timeline of the game.

enforcement through common institutions. We turn now to the specification of a model designed to answer this question.

A model of institutional choice

Like other theories of institutional choice, we take as our starting point a simple model of opportunistic behavior and inefficient noncooperative outcomes. We depart from previous work, however, by giving players a choice, before they attempt to cooperate, over how cheaters are to be punished. One option is an upfront investment in *external enforcement*, establishing an independent authority that imposes costs on violators and attempts to redress victims' grievances. The other is to pay nothing upfront and rely on *self-enforcement*, imposing costly punishments on one another in a process that is *ex post* inefficient. Put differently, external enforcement establishes a common authority that requires commitments of resources and sacrifices of autonomy from all parties, while self-enforcement relies on threats of war to punish violations, a system that foregoes upfront investments but is costly to all parties when punishment is invoked. Thus, the key difference between these institutions is the source of their inefficiency: *ex ante* for external enforcement, *ex post* for self-enforcement.

As shown in Figure 1, the game begins as players $i = \{1,2\}$ choose simultaneously whether to support external or self-enforcement, where choosing the former implies that each pays an upfront cost, $k > 0$. These *costs of integration* come in one of two forms. First, they can represent material investments, for example taxes or common pool contributions, required to establish an independent enforcement authority. Second, they may also be subjective, in that rulers or their citizens would prefer to retain the sense of common identity ensured by self-enforcement, regardless of potential efficiencies in the material costs of integration.[4] In other words,

[4] While we collapse both material and subjective factors into the same theoretical term (k), it is worth noting that subjective factors, particularly nationalisms or cultural affinities, can outweigh material factors, and vice versa. We could decompose the costs of integration into material k_m and subjective k_s components, such that the total costs are $k = k_m + k_s$, but since both material and subjective costs exert downward pressure on the attractiveness of external enforcement, so

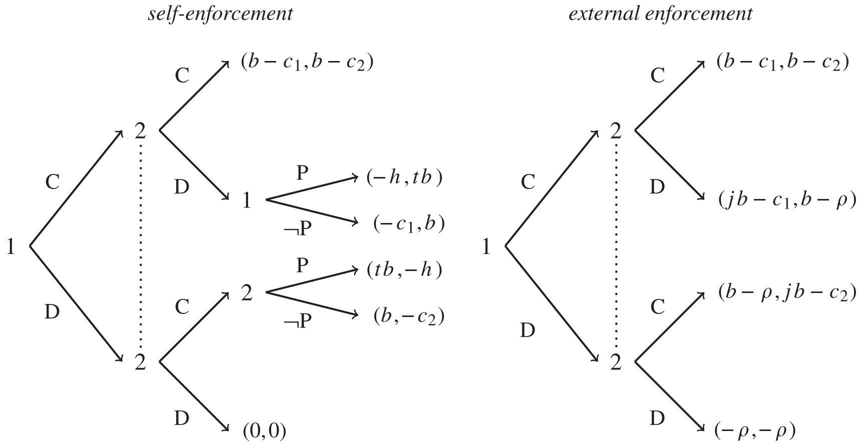

Figure 2 Cooperation/defection under self- and external enforcement.

the stronger a player's distaste for sharing resources or compromising with the other player, the higher the subjective costs of integration. Further, since external enforcement is costly upfront, we assume that it is implemented only in the event that both players choose it in the institutional choice phase; otherwise, if only one or neither player chooses external enforcement, self-enforcement is the equilibrium institution.

After an institution is chosen, players interact in a future cooperative setting with probability $\delta \in [0,1]$. We take δ to be a measure of *interaction density* between the players; higher values indicate players that interact frequently, while lower values denote less frequent interactions.[5] Formally, this means that payoffs following the institutional choice phase are weighted by δ, such that player *i*'s expected value for the game conditional on institutional choice is $-k+\delta EU_i(\text{ee})$ for external enforcement and $0+\delta EU_i(\text{se})$ for self-enforcement.

As shown in Figure 2, the structure of interactions in the cooperation/punishment phase depends on the prior choice of enforcement institutions. In each case, the underlying cooperative problem is a simultaneous choice Prisoner's Dilemma, which we modify by allowing an immediate consequence

we can combine them in the present model without loss of generality. We are grateful to an anonymous reviewer for raising this point.

[5] While iterated cooperation games often use a term like δ to represent a discount factor or a shadow of the future (e.g., Axelrod 1984), others view it as a measure of the time between interactions (see Rubinstein 1982; Muthoo 1999; Powell 2004), short when δ is high and long when δ is low, and this latter intuition accords best with our conception of interaction density.

in the event of some noncooperative choices. If player i chooses to cooperate by playing C, it pays a cost $c_i > 0$ in order to transfer a benefit $b > 0$ to the other player, such that if both cooperate, each receives $b - c_i$. If both players defect by playing D, then neither pays a cost nor receives a benefit from the other, such that each receives zero. Finally, should no punishment occur after one player defects while the other cooperates, the violator receives b but pays no costs, while the victim pays $-c_i$ but receives no benefits. In standard Prisoner's Dilemma terminology, these payoffs are the 'temptation' and the 'sucker', respectively. While this basic structure is common across both institutions, they differ in what follows a player's defection.

Under self-enforcement, a cooperating player against which the other has chosen to defect has the opportunity to punish the other by playing P. If a player does not punish the other's defection, or ¬P, then the payoffs from either unilateral or mutual defection stand. Punishment, on the other hand, allows the victim to deny the violator some of the temptation payoff and limit the pain of being suckered. Specifically, the violator receives a fraction of the temptation $t \in [0,1)$, such that its payoff is tb. Next, the punisher recoups some – but not all – of the costs of cooperation, such that it receives $-h$, where $0 \leqslant h < c_i$. This ensures that, while punishment is costly, it does allow the victim to recover some of its lost utility; otherwise, punishment would never occur. When both players defect, they each receive zero.[6]

Note that this changes the standard Prisoner's Dilemma payoffs following unilateral defection, making them worse for the violator but better for the victim, though it has no effect on payoffs when both players have already defected. This ensures that the exercise of self-enforcement takes on the *ex post* inefficiency of war (cf. Fearon 1995), threats of which are the primary means by which states sustain cooperation and deter violations in the international system (Waltz 1979; Wagner 2007). Notably, players are equally 'powerful' in that they do not differ in their ability to level punishments against one another; while our theoretical claims do not depend on this assumption, we analyze an extension in the Appendix allowing for differential power and explore some additional implications.

External enforcement, on the other hand, delegates punishment to a third party and, with the material and subjective costs of integration paid up front, imposes a punishment $\rho > 0$ on the violator and compensates the victim with some share $j \in (0,1)$ of the nominal benefits of cooperation. Punishment can take the form of an imposed reparation, authorized

[6] Note that, if players were allowed to punish by paying h, neither would do so, since there would be no costs to save or gains to preserve, leading to payoffs (0,0) in any case. Therefore, we do not give players the option of punishment after both defect, though we have solved a model with these options, and the results are unchanged.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

retaliation, or direct action by a central authority, while compensation – though imperfect – might be the restoration of the status quo ante, receipt of reparations, or the fruits of authorized retaliation (see Reinhardt 2001). If player i defects while $-i$ cooperates, it receives $b-\rho$, the temptation less the punishment, while the victim receives $jb-c_{-i}$, its total compensation less the costs of cooperation. Finally, if both players defect, each receives only punishment, or $-\rho$.

Finally, each player enters the cooperation/punishment phase uncertain over the other's costs for cooperation. After an institution is chosen but before players choose to cooperate or defect, Nature randomly draws a cost for cooperation for each side that is high $\bar{c}_i$, with probability ϕ and low, $\underline{c}_i$, with probability $1-\phi$, then informs each player of its own type only. We assume that $\underline{c}_i < b < \bar{c}_i$, which ensures that there exist player-types for which the benefits of cooperation are greater than the costs $(\underline{c}_i)$, and those for which the costs of cooperation outweigh the benefits $(\bar{c}_i)$. Further, since players are unaware of their types until the beginning of the cooperation/punishment phase, the probability of drawing high costs, ϕ, indicates the extent of *commitment problems*, or the probability that cooperative agreements will be unenforceable with threats and require the invocation of punishments. We can also think of ϕ as an indicator of preference volatility, especially as it approaches one-half. Players are thus unable to perfectly anticipate both others' and their own incentives to comply with agreements (cf. Carrubba 2005), which ensures that enforcement institutions are chosen under some uncertainty as to how likely a player is to be suckered, as well as how likely it is to be punished.

While our model shares some important similarities with other work on institutional choice, it also exhibits some notable differences. First, like Alesina and Spolaore's (2003) model of the size of political jurisdictions, our theory integrates the inherent costs of heterogeneity under common institutions through our parameter representing the costs of integration, k. However, we also model explicitly the twin challenges of compliance and punishment both inside and outside common political institutions, which leads to additional insights over the choice of particular enforcement institutions. Second, where other studies of federalism and the rule of law also focus on how political actors weigh the costs and benefits of accepting external authority and exit (see, *inter alia*, Weingast 1997; de Figueiredo and Weingast 2005), we view the act of exit – or, in our case, refusing to accept an external authority – as an option that implies its own type of order: anarchy. For example, in de Figueiredo and Weingast (2005), exiting a federation may be costly, but nonparticipants neither gain subsequent benefits nor pay further costs, while in our model, opting out of external enforcement merely places the onus of maintaining cooperation on the players themselves. Thus, while

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

we abstract away from problems of integration pertaining to distributive conflict between center and units, we give players a richer 'exit' option, one that more closely resembles the 'anarchy' of international politics.

Analysis: choosing enforcement institutions

When will players invest in common enforcement institutions, and when will they choose anarchy? To answer these questions, we begin by identifying equilibrium behavior in each of the cooperation/punishment subgames, then work back through the sequence of play to analyze the players' *ex ante* choice over which institution will govern their future interaction. We then explain the choice of institution as a function of interaction density, δ; the extent of commitment problems, ϕ; the pain of punishment, ρ; and the costs of integration, k.

Before detailing equilibria in the cooperation/punishment subgames, it is worth noting some of their common strategic dynamics. First, regardless of enforcement institution, low-cost types $\underline{c}_i$ can be deterred from opportunistic defections with the threat of punishment, while high-cost types $\overline{c}_i$ cannot. Thus, punishment occurs on the equilibrium path with positive probability, which allows us to analyze tradeoffs between each enforcement institution.[7] Next, though players' initial choices take place under uncertainty about each other's future costs for cooperation, players reveal their types through their strategies, and the victim's punishment choice in the self-enforcement subgame occurs with full knowledge of the violator's type. Nonetheless, the victim's payoffs do not depend on the other player's type, rendering such posterior beliefs trivial; thus, we omit them from the discussion.

Now consider the cooperation/punishment subgame under self-enforcement, beginning with a victim's choice over punishing violations or letting them stand, which occurs with full knowledge of each player's previous move. If the victim of unilateral defection does not punish, it receives $-c_i$, but punishing allows it to secure $-h$, and since $h < c_i$ a victim is sure to punish a violator. In the equilibrium on which we focus, low-cost types would like to defect unilaterally, but they are deterred from it by the other player's threat of punishment; however, high-cost types cannot be deterred by the threat of punishment, and they are sure to defect in equilibrium.[8]

[7] By contrast, were we to apply this logic to the more common iterated Prisoner's Dilemma, analogous punishment strategies – essentially Grim Trigger and some pre-paid variant of Tit-for-Tat – external enforcement would never be chosen as long as both punishment strategies supported cooperation deterministically on the equilibrium path.

[8] We focus on an equilibrium in which player-types take unique actions – that is, a separating equilibrium. However, it is worth noting that, as in many games of asymmetric information, a pooling equilibrium also exists. In this equilibrium, all player-types defect, because neither player has an incentive to cooperate (and subsequently punish), yielding a payoff of $-h$, which is strictly

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Thus, as stated in Proposition 1, possible outcomes of this subgame are mutual cooperation, which occurs when both types have low costs of cooperation; punished unilateral violations, which occur when one player draws low costs and the other high costs; and mutual defection, which occurs when both players draw high costs for cooperation.

Proposition 1: The following are equilibrium outcomes of the cooperation/punishment subgame under self-enforcement, as long as $h \leqslant \overline{h}$ and $t \leqslant \overline{t}$. With probability $\phi\phi$, both players defect. With probability $2\phi(1-\phi)$, one player cooperates and punishes, while the other defects. With probability $(1-\phi)(1-\phi)$, both players cooperate. See Appendix for proof.

Note that we focus on a specific equilibrium of this subgame, that is one in which punishment is not so costly that players will never engage in it ($h \leqslant \overline{h}$) and sufficiently painful to deter the low-cost type from defecting ($t \leqslant \overline{t}$). (Otherwise, the results are trivial.) Before the cooperation/punishment phase, players know only the probability with which the costs of cooperation will be high or low, so i's expected utility for self-enforcement is

$$EU_i(\text{se}) = (1-\phi)(1-\phi)(b-\underline{c}_i) + (1-\phi)\phi(-h) + \phi(1-\phi)(tb)\,, \qquad (1)$$

where the outcomes that yield a nonzero payoff for either player are (a) mutual cooperation by low-cost types, which occurs with probability $(1-\phi)(1-\phi)$, (b) a low-cost player i punishing a unilateral violation, which occurs with probability $(1-\phi)\phi$, and (c) a high-cost player defecting and being punished, which occurs with probability $\phi(1-\phi)$.

Next, if players have chosen external enforcement, there are also four possible outcomes, corresponding again to mutual cooperation, unilateral defection (with automatic punishment), and mutual defection. As before, we focus on equilibrium strategies in which player-types take unique actions, such that outcomes are determined by the combination of player-types: low-cost types cooperate, confident that they will receive the benefits of cooperation through the other player's cooperation or compensation from the external enforcer, while high-cost types defect, preferring the certainty of punishment to the payment of any costs of cooperation in equilibrium. Proposition 2 characterizes these outcomes.

worse than mutual defection. However, since we are interested in effective threats of punishment and uncertainty over preferences, we focus on a separating equilibrium (for both cooperation/punishment subgames) in which all outcomes – including mutual defection – are possible and players act according to type.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Proposition 2: The following are equilibrium outcomes of the cooperation/punishment subgame under external enforcement, as long as $\underline{\rho} \leqslant \rho \leqslant \bar{\rho}$. With probability $\phi\phi$, both players defect. With probability $2\phi(1-\phi)$, one player cooperates while the other defects. With probability $(1-\phi)(1-\phi)$, both players cooperate. See Appendix for proof.

This equilibrium requires punishments that are neither too weak, lest even low-cost types defect, nor too strong, lest high-cost types be deterred from defection. Formally, the constraint is $\underline{\rho} \leqslant \rho \leqslant \bar{\rho}$, or $\underline{c}_i - jb\phi \leqslant \rho \leqslant \bar{c}_i - jb\phi$. We focus on this equilibrium for two reasons. First, it is the most interesting and substantively plausible equilibrium; few enforcement schemes are or can be perfect. Second, since the underlying rate of noncooperative behavior is constant across both enforcement institutions, it helps us rule out any explanation for external enforcement that derives from the desire to reduce conflict; rather, we show that players may choose external enforcement for a number of reasons independent of potential reductions in conflict. To be sure, a reduction in observed violations may be a plausible consequence of choosing external enforcement, but it need not be the case in order to explain why such institutions exist.

As before, players are uncertain over the costs of cooperation before entering the subgame, so we can characterize player *i*'s *ex ante* payoff for the cooperation/punishment subgame under external enforcement as

$$\begin{aligned} EU_i(\mathrm{ee}) = {} & \phi\phi(-\rho) + \phi(1-\phi)(b-\rho) + (1-\phi)\phi(jb - \underline{c}_i) \\ & + (1-\phi)(1-\phi)(b - \underline{c}_i), \end{aligned} \tag{2}$$

where player *i* is punished for any time it defects but receives some limited compensation, *jb*, should it be the victim of unilateral defection.

With expected payoffs for the cooperation/punishment subgame under each enforcement institution established, we now turn to an analysis of the conditions under which players choose either external or self-enforcement. Recall that the institutional choice is made simultaneously, such that the option with upfront costs – external enforcement – occurs if and only if both players prefer it to self-enforcement. Since payoffs are symmetric, we present only player *i*'s choice. It chooses external enforcement if and only if $-k + \delta EU_{\mathrm{i}}(\mathrm{ee}) > \delta EU(\mathrm{se})$, where δ again represents interaction density, or the probability that players interact in a future cooperative setting. External enforcement entails an upfront cost, *k*, balanced against a greater probability of receiving (some of) the benefits of cooperation than would be possible under self-enforcement. When will both players be willing to pay the costs of integration?

Proposition 3: Given $h \leqslant \bar{h}$ and $\underline{\rho} \leqslant \rho \leqslant \bar{\rho}$, external enforcement is the equilibrium institution if interaction density is sufficiently high, or $\delta > \underline{\delta}$.

Otherwise, self-enforcement is the equilibrium institution. See Appendix for proof.

External enforcement thus requires sufficiently high interaction density, or

$$\delta > \frac{k}{\phi((1-\phi)(b(j-t+1)+h-\underline{c}_i)-\rho)} \equiv \underline{\delta}. \tag{3}$$

Put differently, players must be confident enough in the occurrence of future cooperative interactions that paying up front for external enforcement is not wasteful. If, on the other hand, players have opportunities to exploit each other only infrequently, then they will forego paying the immediate costs of establishing external enforcement, trusting instead in the relatively low risk of employing *ex post* inefficient punishments like war.[9]

Interaction density, however, is not sufficient to explain external enforcement. For the constraint defined in (Equation 3) to bind – that is, for $\delta > \underline{\delta}$ to be satisfied for plausible values – two other conditions must be satisfied. First, the costs of integration, whether material or subjective, must be sufficiently low, or

$$k < \phi((1-\phi)(b(j-t+1)+h-\underline{c}_1)-\rho) \equiv \overline{k}.$$

Formally, this ensures that $\underline{\delta} < 1$, but substantively it guarantees that the costs of integration are not so large as to make external enforcement unattractive for even the highest interaction densities. Second, the pain of punishment, which both players expect to endure with some probability, must not be too large, or

$$\rho < (1-\phi)(b(j-t+1)+h-\underline{c}_1) \equiv \hat{\rho}.$$

Otherwise, if punishment were more painful, players would refuse to trade some possibility of bringing punishment upon themselves in return for some compensation when they have been the victim of unilateral defection.

Next, by taking comparative statics over the threshold $\underline{\delta}$, we can also specify the effects of commitment problems on institutional choice. As shown in Figure 3, the probability of drawing high costs for cooperation has a nonlinear effect on the attractiveness of external enforcement. To see this formally, note that the first derivative of $\underline{\delta}$ with respect to ϕ is

$$\frac{\partial\underline{\delta}}{\partial\phi} = \frac{k((1-2\phi)\underline{c}_1-b(1-2\phi)(j-t+1)+2h\phi-h+\rho)}{\phi^2((1-\phi)\underline{c}_1-b(1-\phi)(j-t+1)+h(\phi-1)+\rho)^2}$$

[9] Notably, interaction density need not be 'high' in any absolute sense – merely high enough to justify the costs of integration (k), which when low enough, can make the threshold over δ easy to meet even as ostensibly 'low' interaction densities.

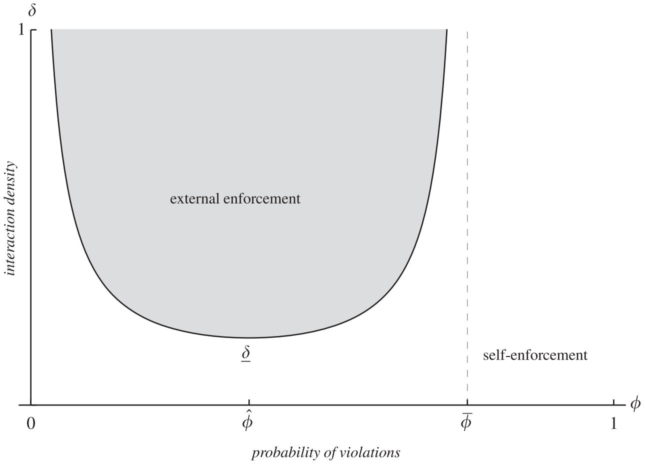

Figure 3 Equilibrium institutions by interaction density and probability of violations.

the sign of which depends on ϕ. When commitment problems are sufficiently unlikely, or

$$\phi < \frac{1}{2}\left(1 - \frac{\rho}{b(j-t+1)+h-\underline{c}_1}\right) \equiv \hat{\phi},$$

the derivative is negative, meaning that increases in the probability of a violation lower $\underline{\delta}$ and make the constraint easier to satisfy. On the other hand, once ϕ passes $\hat{\phi}$, then the derivative is positive, and as the probability of violations increases, the constraint $\underline{\delta}$ becomes more difficult to satisfy. Conveniently, $k < \overline{k}$ also becomes easier to meet as ϕ increases through low values and harder to meet as it increases through high values.

Thus, external enforcement institutions are difficult to sustain when the probability of violations is low, because the upfront costs are unlikely to be worth paying. Players would rather save the costs of integration today and gamble on the low odds of needing to invoke punishment in the future. However, when the probability of violations is sufficiently high, then mutual punishment becomes so likely that that the costs of integration are again wasteful, and players will opt instead for self-enforcement. By contrast, the attractiveness of external enforcement peaks when the probability of violation approaches $\hat{\phi}$, where the costs of cooperation are relatively more volatile. Therefore, when violations are neither too unlikely to justify the costs of integration nor too likely to justify the costs, external enforcement becomes the preferred enforcement institution. Anarchic systems of self-enforcement, on the other hand, can be consistent with either high *or* low rates of rule violations and punitive violence.

Next, by re-solving the constraint $\rho < \hat{\rho}$ in terms of ϕ, we can derive the precise upper bound on the probability of violations that can support external enforcement. Specifically, external enforcement can only occur in equilibrium when

$$\phi < 1 - \frac{\rho}{b(j - t + 1) + h - \underline{c}_i} \equiv \overline{\phi},$$

where $\underline{\delta}$ approaches $\overline{\phi}$ only asymptotically, as shown in Figure 3. As before, while players would like to see others punished for violations, an increase in the probability of future commitment problems applies both to other players *and* themselves, increasing the probability of receiving $-\rho$ after mutual defection. Even with high interaction density and tolerable punishments, when bringing external punishment upon oneself is too likely, players will opt for anarchy – where they are punished only to the extent that other players can impose it on them – rather than submit to an external authority likely to punish them in the future.

Finally, consider the role of j, which represents the share of the benefits of cooperation that the victim of unilateral defection receives when a violator is punished, whether because of reparations or the fruits of authorized retaliation. Each of the critical constraints defined above – $\underline{\delta}$, $\overline{k}$, and $\overline{\phi}$ – becomes easier to satisfy as j increases. Thus, the more effectively external enforcement institutions can compensate victims while also punishing violators, the more likely are players to accept external enforcement.

The choice between external and self-enforcement is a complex function of interaction density, the costs and benefits of institutional alternatives, and the expected probability of violations. When players interact frequently, the costs of building common institutions are not prohibitive, and violations are either too unlikely nor too likely, they will opt for external enforcement, paying upfront to save themselves the *ex post* costs of punishing violators in the future. However, if interactions are too infrequent, the costs of integration too high, or violations are very likely or very unlikely, players will opt instead to live in an anarchic system, saving today's investment in common institutions to mete out future punishments themselves.

Explaining integration and separation

To demonstrate the logic of our model, we examine in this section three prominent historical cases. Each illustrates how the expectations of the model correspond to empirical phenomena in political leaders' choice to (1) integrate themselves and delegate enforcement to an external entity or (2) separate themselves and reserve the right to carry out their own enforcement. First, we examine the transition from the Articles of Confederation to the US Constitution. Second, the integration and the dissolution of Czechoslovakia provide an example of political entities that have peacefully gone through both the integration and the separation processes. Following the Czechoslovak case, we briefly discuss civil wars and secessions more generally. Third, we turn to the case of most significant integration in the current era – the European Union (EU). Rather than tracing the whole of the history of European integration, we analyze the most critical step, the Maastricht Treaty, which created the EU.

In our theory, states as sovereign political entities are endogenous to the choice of external enforcement. Therefore, the ontological units in our discussion below are political leaders and entities under their rule. Krasner (1999) makes a similar move in theorizing the determinants of sovereignty. In each case, we begin with the new institution that political leaders choose to create. Then we identify the previous institutional context and the initial parameter values – ϕ, δ, and k – and trace through how these values change over time, shaping political leaders' incentives to choose the new institution.

From the Articles of Confederation to the US Constitution

Despite its current status as a major global power, the United States did not begin life as a highly integrated political entity. The 13 original states negotiated an initial agreement to form a confederation – the Articles of Confederation and Perpetual Union – while fighting for independence from Great Britain. The extent of integration among the entities in both this initial agreement and the strengthened subsequent agreement – the United States Constitution – is consistent with the predictions of our model. The historical nature of this example, unlike the subsequent cases, also suggests that our model is not tied to any particular historical configuration of the international system.

The US Constitution departed significantly from the Articles of Confederation in many important ways.[10] Most critically, the constituent states no longer possessed a *de facto* veto over the federal government's decisions. Decision making became generally more centralized, including the federal government's authority to enforce laws with force, which clearly delegated enforcement power to a third party above the former colonies. The centralization, however, was not maximal. Although the Constitution vested greater power in the executive than before, power was separated among three branches of government. Moreover, the Constitution divided the responsibility for national defense between the federal and state governments.

[10] The following discussion draws from Hoffman (2006, 55–80) unless otherwise cited.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

What explains the 13 former colonies' decision to jettison the confederation arrangement and delegate significant enforcement authority to a third party – the federal government? Hoffman (2006) argues that institutional design can foster trusting and cooperative relationships among political actors, and where the Articles engendered distrust among the 13 states, the Constitution reduced it. More explicitly, 'Whereas the Articles of Confederation presumed that the states had to defend their interests against one another, the assumption underlying the federal Constitution was that the states did not have to be quite so vigilant' (Hoffman 2006, 71). What his explanation fails to account for, however, is why the assumption among states about one another changed over time to prompt the institutional change. Our explanation highlights how this process occurred.

Two key factors dominated the years during which the colonies negotiated and ratified the Articles of Confederation: very high probabilities of violation by one another (high ϕ, or low uncertainty) and relatively low interaction density (low δ). First, states were highly suspicious of one another's intentions. Virginia congressman Richard Henry Lee wrote the following in his letter to the governor of the state: '[o]ne thing is certain, that among the middle and southern states Virginia has many enemies, arising from jealousy and envy of her wisdom, vigor, and extent of territory' (quoted in Burnett 1923; Hoffman 2006, 57). Second, during the years of bargaining over the Articles of Confederation, Britain dominated the colonies' attention, thereby reducing the interaction density among the colonies, as evidenced by the sidetracking of negotiations over the Articles during the early stages of the war.

The institutional design of the Articles of Confederation reflected these underlying dynamics. As expected by the high probabilities of violation and low interaction density, states were not integrated in a meaningful sense. The confederation entailed insignificant delegation of power to the federal government and constituted a minimum necessary cooperation among the constituent and sovereign states. The responsibilities of the federal government under the Articles – consisting solely of the Continental Congress – mainly included foreign policy, diplomacy and commanding the federal military. Internally, however, states possessed a *de facto* veto over congressional decisions (Hendrickson 2003, 150–51). Congress also lacked authority to tax or to conscript citizens, stripping it of any consequential enforcement power.

States' behavior under the Articles was consistent with the high probabilities of violation and the lack of meaningful interaction among them. Violations among states were rampant, as they disputed boundaries in western lands and tariffs on trade between the states. In fact, James Madison identified numerous violations in his paper 'Vices of the Political

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

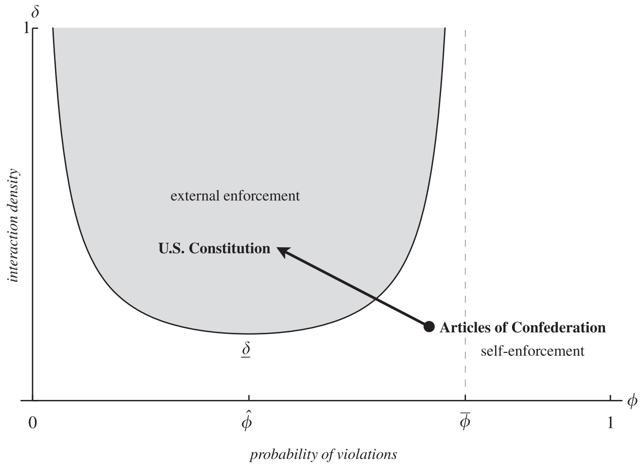

Figure 4 Institutional change from the Articles of Confederation to the US Constitution.

System of the United States', which preceded *The Federalist* papers.[11] Given the weak institutional design that reflected the underlying commitment problem among the states, the congress was incapable of punishing the violations. The severity of the violations grew to the point where even the Anti-Federalists became concerned with 'the growing irrelevance of congress in managing common concerns and resolving differences' (Hendrickson 2003, 211).

Although Hoffman implies that these problems compelled the states to strengthen the union, our model anticipates that these violations were neither sufficient nor necessary for delegation of enforcement to an external entity. In the absence of moderate probabilities of violation and high interaction density – along with low costs of integration – political actors will not seek integration. Conversely, precisely when rampant violations are absent – reflecting the middling probabilities of violation – political actors will choose to delegate enforcement to an external entity, as long as the other conditions are also satisfied.

Our explanation highlights the ramifications of the Revolutionary War for the American states. The states moderated their likelihood of violations (lower ϕ) and increased their interactions (higher δ) among one another in the aftermath of the war. Figure 4 summarizes the changes. First, states became less likely to violate their commitments with other states once

[11] Hendrickson (2003: Ch. 24) summarizes the eight such vices pertaining to inter-state and state-federal relations.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

the war was over, because threats to their well-being came not from one another but from within the states and outside the union. Onuf (1983, 173–85) describes the ‘anarchic’ situation in the United States after the war where internal rebellions and concerns about a counterrevolution sponsored by Great Britain dominated. While these perils did not eliminate the longstanding conflict between large and small states within the union, they reduced the salience of the ongoing disagreement between the two groups.

Second, the United States’ emergence as an international actor compelled the states to increase the interaction with one another. The United States signed national treaties with European states – Great Britain, France, and Holland – the most important being the Treaty of Peace, which among other provisions, required Americans to pay back pre-war debts to British creditors and to prevent persecution of loyalists (Hendrickson 2003, 213). John Jay warned that if any one state violated the provisions of the treaty, ‘“any part of the Community” could “bring on the whole” the calamities of a war’ (Hendrickson 2003, 214). Fears also existed that if any part of the United States was in violation of the treaty, the whole country would be economically shut out of Great Britain. Thus, the foreign relations of the United States tied the behavior and policies of each state with those of all the others, which in turn increased the interaction density among the states.

The decreased probability of violations and increased interaction density moved the American states from the lower right corner of Figure 4 to the middle. Whereas before the changes the states chose to reserve the authority to self-enforce, after the changes they had incentives to negotiate and create institutions for strong external enforcement. These exogenous changes account for why the states became more ‘trusting’ in Hoffman’s terms, the change for which his explanation fails to account. Moreover, the institutional solution that the states reached during the Constitutional Convention – the Great Compromise, basing representation in the House on states’ population and guaranteeing equal representation in the Senate – subsequently lowered the costs of integration (k) among the states, which helped consolidate and stabilize the institutional settlement reached in 1787.

The formation and dissolution of Czechoslovakia

Next, the history of Czechoslovakia provides unique examples of both integration between *and* separation into political entities. The unified state formed in 1918 and, after significant political turmoil throughout much of the century, peacefully dissolved into two independent states at the end of 1992. Both the initial formation of the state after the First World War and its ultimate

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

dissolution after the Cold War are consistent with the predictions of our model. Changes in the interaction density and the probability of violations resulting from broader political and economic changes occurring in Europe shaped political leaders' choices with respect to integration and separation.

Formation. When Czech and Slovak leaders agreed in May 1918 to create an independent Czecho-Slovak state (Kraus and Stanger 2000, 2), the former sought independence from the Austrian Empire, while the latter sought independence from Kingdom of Hungary. In October of the same year, Thomás Masaryk, the Czech leader, issued the Czechoslovak Declaration of Independence in Washington, DC. The new state that emerged was fully unified with the central government located in Prague. The parliamentary system was based on the Austrian electoral system (Leff 1988, 48). Suffrage was compulsory as well as universal, and seats were allocated according to modified proportional representation. Although the system resulted in a high level of fragmentation with 50 parties contesting parliamentary seats in the First Republic's first four elections (Leff 1988), the 'twenty-year record of political stability and liberal democracy was a notable achievement in the history of democratic constitutional experiments' (Kraus and Stanger 2000, 4).

Why did the two political entities choose to integrate and create a single sovereign state? Leff (1988) implicitly points to the role that Masaryk played in pushing for unified independence by the two states. While pursuing independence from Austria, he also advocated an alliance with Slovakia, which would allow the combined Slav majority of nine million people to vastly outnumber the three-million strong German minority (Leff 1988, 35). As the Czech leader, Masaryk did play a critical role. However, a single leader would have been unlikely to successfully push for integration in the absence of favorable structural conditions. Our explanation highlights the conditions that provided the underlying incentives for the two entities to pursue integration.

For most of their history, Czech and Slovakia remained separate. Most recently, before the creation of the First Republic, Czech Lands – consisting of Bohemia and Moravia – were a part of the Austrian Empire and Slovakia was a part of the Kingdom of Hungary, as it had been for most of the second millennium. While conflicts existed between the encompassing empire and kingdom, the component parts – Czech lands and Slovakia – were not likely to go to war against each other. The formation of the dual monarchy between Austria and Hungary in 1867 lowered the probability of violations (ϕ) even more. Moreover, both the subjective and the material costs of integration (k) were not high. Despite the presence of significant ethnic minorities – Germans in the Czech lands and Magyars in Slovakia – both

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

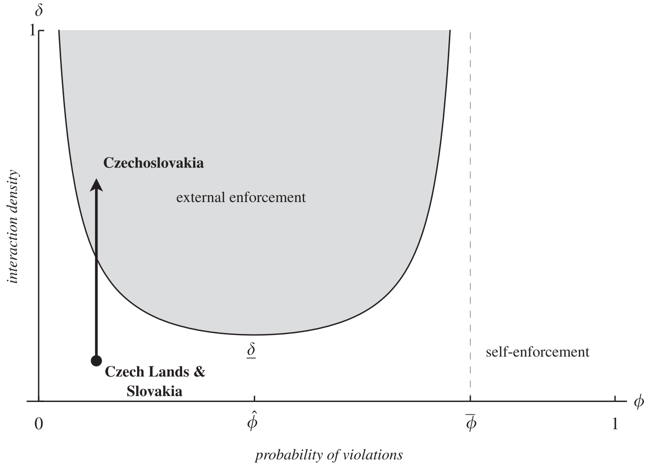

Figure 5 Integration of Czech lands and Slovakia, 1918.

entities consisted of Slavs, lowering any mutual suspicion and difficulties in integration. The two entities were also key economic components of the empire. A large income gap existed between the two, but 'Hungarian economic policy had made Slovakia the most industrially developed sector of its territory' (Leff 1988, 12), thereby making the material costs of integration relatively low.

The Czech lands' and Slovakia's choice to integrate was simultaneously a choice to separate themselves from their respective imperial masters. While we do not focus on the latter choice, it likely reflected increasing probabilities of violation between ethnic groups and political entities within empires that resulted in separatist movement across the region. What encouraged the two entities' decision to integrate, however, reflected the increasing density of interaction (δ) between the two, as indicated in Figure 5. Coinciding well with the creation of the Austro-Hungarian Empire, Kraus and Stanger maintain that '[b]y the second half of the nineteenth century, ethnically related Czechs and Slovaks increased their contacts and found a common denominator in their quests of self-determination' (Kraus and Stanger 2000, 2). There were increased contacts at the societal level on both sides. On the Czech side, groups such as Československá jednota, consisting of Czech intelligentsia and middle class, maintained contacts with Slovak intellectuals (Leff 1988, 33). These groups also worked to increase the broader Czech population's awareness of Slovaks' circumstances. Leff also identifies a 'network of social and cultural contacts [between Czechs and Slovaks] that grew up in the prewar period' (p. 36). For example, Slovak Catholics came to see their Czech counterparts

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

as potential allies. More generally, ethnic separatism was burgeoning all over Europe. This likely boosted the expectation that Czechs and Slovaks held about the density of future interactions among co-ethnics in the region. Masaryk clearly pushed his agenda of unified independence of Czechoslovakia. However, in the absence of the increased interaction density that occurred under Austro-Hungarian Empire and intensified during the early years of the 20th century, his strategy would not have been successful.

Dissolution. As stable as the First Republic was during the interwar years, the unified state's subsequent decades were tumultuous.[12] The initial downfall occurred in Munich in October 1938, where the Western powers acceded to Nazi Germany's plan to divide the Czechoslovak territories. In 1939, Czech lands became a German protectorate and Slovakia became a separate state under German sponsorship. After the war in 1945, the two entities reemerged as a unified state (the Second Republic), but its existence as a democratic entity only lasted 3 years. Communists engineered a government takeover in February 1948, henceforth officially referred to as the 'Glorious February Revolution'. The major attempt at political and economic reform in 1968 – the Prague Spring – ended with an invasion by the Soviet Union. However, one aspect of the reform survived. Addressing the longstanding Slovakian discontent and nationalism, on January 1, 1969, Czechoslovakia became a federation. Last, the Velvet Revolution, which started at the end of 1989, brought democratization to the state. External factors decidedly shaped these developments over six decades. The changes within the state did not reflect choices that leaders made on behalf of their political entities but rather reflected exogenous constraints that dictated the institutional features.

By contrast, the separation of the state into Czech Republic and Slovakia manifested the internal political leaders' calculations and choices under changing external circumstances. After various undertakings throughout 1992 to maintain the unity of the state, on January 1, 1993, the country peacefully separated into two sovereign entities. They dismantled the common federal government they shared and each national government that co-existed with the federal government became responsible for governing its respective national entity.

Why did Czechoslovakia break up into two independent states, abandoning external enforcement for self-enforcement? Many analyses locate the cause at a very fundamental level – from the onset of the First

[12] The brief summary here draws from Leff (1997).

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Republic, Slovaks were dissatisfied with the unification. More specifically, 'in spite of extensive efforts by politicians and intellectuals in the interwar period and partly also after the Second World War, the idea of a common Czechoslovak state did not put down deep roots in Slovak soil' for structural, cultural and psychological reasons (Musil 1995, 2). As indicated by the separation of the state during the Second World War and the federalization of it in 1969, the issue of dissolution did *not* emerge for the first time after the Velvet Revolution. In other words, the unified state always struggled with subjective costs of integration – k – that were not sufficiently low. What remains unexplained, however, is the timing of the severance, for which deep structural factors are unable to account. The insights from our model highlight the role of exogenous changes that changed political leaders' incentives with respect to maintaining the unified state.

The Velvet Revolution started the transition process that was similar to post-communist transitions in other states but also different in an important respect. Compared to Hungary or Poland, the newly named Czech and Slovak Federative Republic was ethnically more heterogeneous, and the first election reflected that reality (Leff 1997, 99). Moreover, the political institutions were divided. In addition to each republic's own legislature, 'the Federal Assembly was subdivided into chambers where each republic's representatives had veto power over important constitutional issues' (Leff 1997). This federative structure reflected each side's higher certainty about the other's unwillingness to cooperate (ϕ). The compartmentalization also reduced the interaction density (δ) among political leaders and the societies. At the societal level, the interaction density had decreased as well. Kučera and Pavlik note that '[i]nternal migration was very intensive especially after the war, while subsequently a process of a certain closing and isolation of both populations is evident…' (p. 36). In short, the probability of violations was higher than under the First Republic and the interaction density somewhat lower. Figure 6 shows the starting point of post-Velvet Revolution Czechoslovakia.

The critical factor over the subsequent few years, we argue, was the increase in both sides' probability of violations. Žák (1995) supports this argument most tersely: 'It is certainly possible to give the political elite credit for not considering the use of force' (p. 267). That the use of force was not out of the realm of possibility in pursuing the breakup indicates both sides' loss of confidence in the other. More fundamentally, the challenges of post-communist transition – political liberalization and economic reform – accentuated preexisting differences. In order to be successful in the reemerging democracy, leaders of the two political entities had to cater to divergent societal interests. 'While in the Czech Lands the premise of

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

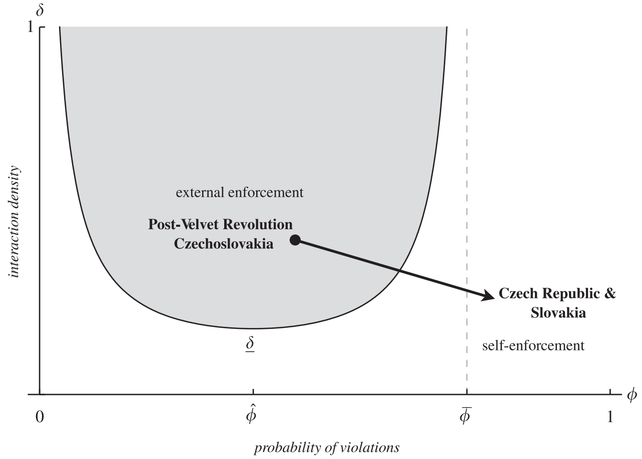

Figure 6 Dissolution of Czechoslovakia, 1993.

success was anti-communism, in Slovakia it was national and social issues' (Žák 1995, 266). Prime Minister Václav Klaus of the Czech Republic sought 'wholesale economic reform and tighter federation' (Leff 1997, 131), while Slovak Prime Minister Vladimír Meciar wanted a decentralized confederation to satisfy Slovak nationalism. The divergent interests implied that the federation could not function without violating one or the other republic's interests. Ultimately, '[i]n Czechoslovakia there were few politicians on the federal and republic level who were willing (and able) to share power' (Žák 1995, 266). In summary, during the early 1990s, the probability of violations among the two republics steadily increased as their interests diverged. The interaction density also steadily decreased as the leaders consolidated power within their respective republics. This movement, as summarized in Figure 6, ultimately resulted in leaders' incentives for separation of the sovereign state into two independent units.

Civil wars, secessionist movements, and failed states

The separation of Czechoslovakia was peaceful, but for many other extant states, the move to self-enforcement is more violent, resulting in civil wars, secessionist movements and failed states. While we do not consider a particular case, our theory can shed light on these violent transitions as well. Existing arrangements for external enforcement – that is the state monopolizing the legitimate use of force – will disintegrate when interaction density falls, the costs of maintaining integration increase or the probability of violations becomes more extreme. Interaction density can fall when

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

relevant populations (e.g. two ethnic groups) naturally partition themselves over time into distinct geographic areas (Schelling 1978). A substantial disruption in the existing transportation and communication technology can also decrease interaction density. Costs of maintaining integration can increase, for example, when the state is no longer able to project its coercive power to all of the extant territory and/or population. This likely happens when a group of political actors occupy previously irrelevant and remote part of the territory, consistent with the role of 'rough terrain' in civil wars (Fearon and Laitin 2003) and failed states. Lastly, actors' underlying preferences for violating cooperative arrangements can exogenously change over time, either due to changing internal or external situations that political leaders confront.

Evolving integration of the EU

The EU is the archetype of integration among sovereign states in today's world. While Europe's integration does not include complete rejection of sovereignty at the national level or full delegation of authority to the supranational level, the EU does embody the highest level of integration states have achieved in the modern era and provides a useful case to illustrate the workings of our theory. At the same time, the continuous efforts at integration by European states over the nearly six decades make difficult identification of exogenous factors influencing the ongoing integration. Continuing the approach we follow above, however, we take states' previous institutional choices as constituting an exogenous context for their subsequent institutional choices. After briefly discussing the overall pattern, we focus on the Treaty on European Union, commonly referred to as the Maastricht Treaty.

Background. European states entered the post-World War II era with a higher interaction density than other regions around the world. The close geographical proximity among the states provided the foundation for this, but the history of major wars on the continent further reinforced this density of interaction. In this context, the most critical factor that contributed to the initial integration efforts by the European states was the moderation in the violation probability, especially between France and Germany. Whereas the probability was high through the war, it diminished with the defeat of Germany, though it did not guarantee that Germany would remain fully cooperative in the future. Uncertainty over Germany's underlying preferences (middling ϕ) prevailed in the region.[13]

[13] Other regions provide useful contrasts. The Middle East is likely characterized by a high violations probability and North America by a low violations probability.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Technological innovations over time reinforced greater interaction density (δ) and lowered material integration costs (k). Lowering of the costs of transportation and communication increased interactions not only among the rulers but among the citizens of each member state of the EU. Building on the ubiquitous network of railroads on the continent, the rise of various budget airlines around Europe and the concomitant rise in intra-continental passenger travel demonstrated this change most explicitly.

Moreover, socio-political innovations also lowered the integration costs. EU member states not only created a centralized dispute settlement institution – the European Court of Justice (ECJ) – but also devised a means for this supranational body to enforce its rulings. Rather than a central EU institution, national courts and institutions proceeded to execute and enforce the rulings by the ECJ (e.g., Burley and Mattli 1993; Alter 1998). This innovation enabled a relatively low-cost implementation of centralized enforcement by the ECJ.

The Maastricht Treaty. The integration of European states proceeded throughout the post-World War II era, notwithstanding the Eurosclerosis that slowed down the process in the 1970s and early 1980s. The entry into force of the Maastricht Treaty in 1993 was a crowning achievement in that the treaty dramatically increased the level of integration among member states, primarily by creating a single currency for the members choosing to do so. Relinquishing national currencies and delegating the monetary authority in a central institution constituted a significant political integration. While the single currency was the principal component of Maastricht Treaty, the other aspects of the treaty laid the foundation for further political integration. In addition to the 'pillar' on the European Communities, the treaty included a pillar on Common Foreign and Security Policy and a pillar on Cooperation in the Fields of Justice and Home Affairs. Accordingly, this treaty is substantively different from the previous and the subsequent treaties of European integration.

What explains EU member states' decision to negotiate and ratify the Maastricht Treaty? Moravcsik (1998) argues that societal preferences in powerful member states were the primary factors that explain EU member states' choice to create a single currency, a key aspect of the treaty. While he demonstrates that actors long held a preference for moving to a single currency, the argument leaves unexplained what led societal actors to develop that preference and the other parts of the Maastricht Treaty. By contrast, our model highlights two separate factors: the higher interaction density resulting from the previously successful attempts at integration and the increased probability of violation among member states resulting from the inchoate single market. These factors explain the other components of

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

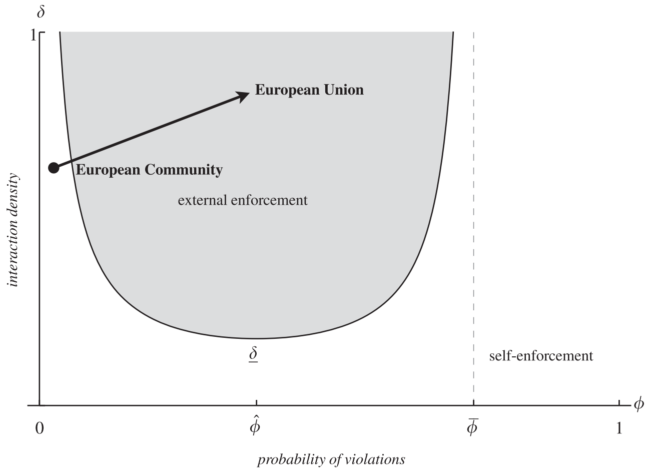

Figure 7 Treaty on the European Union (TEU), 1993.

the treaty as well as the origin of economic actors' preferences for a single currency.

The background against which states signed the Maastricht Treaty was shaped by the extant institutions and preceding treaties among the member states. The two main institutions were the European Monetary System (EMS) created in 1979 and the Single European Act (SEA), which states signed in 1986 and entered into force in 1987. Through the Exchange Rate Mechanism (ERM) of the EMS, European Community members limited fluctuations in their exchange rates (Nugent 2003, 296). Much of the SEA increased the efficiency of decision-making within the European institutions (Nugent 2003, 58–59), and its central substantive provision provided for the completion of the European Community's internal market by 1992. These conditions reflected high interaction density (δ) among the member states and relatively low probability of violations (ϕ). Figure 7 shows this as the starting point of the institutional change: the European Community.

In this context, violations consisted of states' failure to maintain their exchange rate commitments within the ERM and of actions that impinged on other member states' well-being as the integration of the national markets increased. The EMS experienced significant problems both at the outset and subsequent to the SEA. Contrasting economic conditions and monetary policies of states in the European Community rendered cooperation difficult, and the states had to adjust and expand the ranges of acceptable exchange rates (Oatley 2012, 242–44). Despite the recurring conflicts, no means of effective punishment existed at the European Community level (Nugent 2003, 305–06). In addition, states with high

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

levels of social protection became wary of those with lower levels of protection, particularly over the possibility of 'social dumping' in which firms might relocate to member states with lower levels of social protection (Nugent 2003, 60).

These increased violations reflected two deeper changes. First, the interaction density (δ) had increased among states. Obviously, the density was always high among western European states throughout modern history. The movement towards a single market, however, further heightened the interaction among the states.[14] Second, states' probability of violations (ϕ) grew, increasing states' uncertainty about one another. The end of the Cold War altered the dynamics in Europe. Even without the unification of Germany – which Moravcsik (1998) characterizes as the conventional wisdom for explaining the Maastricht Treaty and dismisses – states would have become more uncertain about the overall cohesiveness of the community and one another's proclivities to cooperate.

While the EU case fits well with the expectations from our model, it simultaneously demonstrates the uniqueness of the conditions fostering integration at the global level. The density of interaction is shaped by the institutional evolution of the EU itself as well as geography and history. Combined with this density is the probability of violation that hovers the middling range at critical junctures – the end of the Second World War and the end of the Cold War. These are difficult conditions to reproduce in other parts of the world, and precisely due to this impediment, anarchy persists at the global level. We turn to this issue in the next section.

Anarchy in international politics

A key implication of our theory is that anarchy is not an exogenous, structural constraint but an endogenous, institutional choice made by rulers that wish to rely on self-enforcement of cooperative arrangements rather than delegate enforcement to an external entity. The choice of anarchy often *appears* to be a structural constraint, because the underlying conditions shaping rulers' choices have endured. The conditions, however, are not immutable, and over time rulers can choose different enforcement arrangements on which to rely. Like any institution, anarchy itself represents a set of 'congealed tastes' (Riker 1980). Nonetheless, our theory points to three potential sources of change that can lead to integration of

[14] The United Kingdom has always been a skeptical participant in the European integration project. While the explanation for the remoteness can be numerous, the geographical distance between the islands and the continent as the factor constraining interaction density and thus integration is conceivable.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

states: increases in interaction density, moderation in the probability of violations or decreases in the costs of integration.

Scholars have long presumed the existence of anarchy at the international level as an immutable condition that states take for granted (e.g., Waltz 1979; Axelrod and Keohane 1985; Powell 1994). Milner (1991), on the other hand, challenges the conception and assumption of anarchy in international relations and maintains that the distinction between the domestic and the international realms is exaggerated. Contrary to Milner, we argue that political actors' reliance on external enforcement of cooperative arrangements qualitatively differs from their reliance on self-enforcement (anarchy). Similar to Milner, however, we maintain that the distinction between the two is neither fundamental nor exogenous. Rather, anarchy at the global level – the absence of a centralized enforcement institution – is the result of rulers' collective choices. Indeed, the conditions necessary for centralized institutions are quite demanding. As discussed above, the density of interaction among political entities needs to be sufficiently high, the probability of violations can be neither too high nor too low and the costs of integration need to be sufficiently low. In the discussion that follows, we discuss why the conditions at the global level lead political entities to choose self-enforcement rather than external enforcement.

Density of interactions

Despite the relatively high degree of economic interaction that characterizes the late-20th and early-21st century, the frequency of interaction among political entities around the world is very low, especially compared with that among individuals at the local level. Two historical characteristics of war – the self-enforcement of cooperative arrangements among political entities – reinforce this observation. First, wars have been extremely rare (King and Zeng 2001). Although all political entities may consistently have had low probability of violation throughout history, another, likelier, possibility is that the density of interaction among entities has been too low to necessitate frequent reliance on self-enforcement. Second, the wars that have taken place have been more likely to be among neighboring political entities (Vasquez 2009). However, even among contiguous political entities at the global level – with the density of interaction presumably higher than that among more distant ones – war has been a rare event in absolute terms.

One of the main constraints that keep the interaction infrequent at the global level is technology. Technological progress has not been sufficiently extraordinary to overcome the barrier that geographical distance between political entities poses for frequent and dense interaction. Technological

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

innovations – railroads, steamships and jetliners – have led to a dramatic fall in the costs of transportation and facilitated interactions. Changes in communication technologies – telegraph, telephone, satellite communications, and the internet – have also contributed to increased interactions. However, the costs of interaction continue to remain substantially higher for actors at the global level than for actors at the local level. Data on domestic vs. international travel are consistent with this distinction. Every year between 2006 and 2010, at least 30 times more individuals in the United States traveled within the country than traveled abroad to other countries.[15] Although the United States is larger than average in its geographical size, the contrast highlights the different levels of interaction that take place within and across political entities.

Lastly, political entities' choice of maintaining anarchy in the world due to low interaction density exhibits negative feedback. At the local level, political actors have a high level of interaction, choose to invest in external enforcement and create a state. Once states come into being as coherent and unified entities, however, they further increase interactions among actors within them at the cost of interactions among actors across states. Economists have identified the 'border effect' (Engel and Rogers 1996) – cities within a state trading more with other cities in the same state than with cities same distance away but in a different state – which exemplifies this dynamic.[16] More generally, Waltz (1970) argues the following:

> In the late 19th and early 20th centuries, the external sector loomed large. Not only was the level of external transactions high in comparison with internal production, but also the internal order was characterized by a low level of governmental activity. Even if the interdependence of nations has increased in the meantime, the progress of internal integration and the increased intervention of governments in their domestic economies means that for most states the internal sector now looms larger than it once did (p. 208).

In short, once a set of political entities make a choice to have external enforcement at their respective locality but maintain self-enforcement at among themselves, path dependent dynamics sustain low interaction densities at the global level but high interaction densities within localities.

[15] United Nations World Tourism Organization. Compendium of Tourism Statistics: Data 2006–10.

[16] The original finding was in the context of US–Canada trade, and some economists have disputed the existence of this phenomenon. But others have confirmed it in a different, least likely context – the EU (Nitsch 2003).

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Probability of violations

The role of the probability of violations, or commitment problems, is not immediately intuitive. Entities do not pursue integration and delegation to external enforcement when doing so might be easiest – when the violation probability is low – or when doing so might be most needed – when the violation probability is high. Instead, they choose external enforcement when their probability of violations is relatively uncertain, which is consistent with institutionalist arguments highlighting the role of uncertainty in creating and designing institutions (e.g., Abbott and Snidal 2000; Koremenos *et al.* 2001).

As Figure 3 shows, however, given low interaction density – which is the case at the global level as discussed above – any probability of violations is consistent with actors choosing self-enforcement over external enforcement. Thus, while this parameter does not independently account for why political entities choose anarchy at the global level, examining its implications is helpful for understanding the origins and role of anarchy in international politics.

The probability of violations is the probability of actors being the high-cost type, which finds the costs of cooperation prohibitively high. The distinction between the high-cost and the low-cost type already exists in the international relations literature – as between status quo and revisionist powers (e.g., Schweller 1996). A world with high probabilities of violation is one of relatively more revisionist powers, and a world of less likely violations has relatively more status quo states. Identifying political entities' propensity for violations is difficult. However, given their choice of anarchy, the frequency of self-enforcement (i.e. war) we observe should reflect the underlying distribution of the two types. The rarity of war in modern history, even among neighboring political entities, suggests that the world has typically contained more of the status quo types, those that incur low costs for cooperation. This is also consistent with the pithy observation by Henkin (1979) that 'almost all nations observe almost all principles of international law and almost all of their obligations almost all the time,' as well as the notion that the costs of war generally provide strong incentives for political entities to avoid it (Fearon 1995).

This interpretation of the world reverses the conventional understanding of anarchy as a cause of war. The prevalent assertion in the literature is that anarchy causes or at least provides a permissive environment for war. Our theory suggests that political entities have chosen anarchy or self-enforcement as the mode of organizing global politics because they anticipate the need for self-enforcement to be sufficiently low. In other words, the likelihood of war affects existence of anarchy, rather than the opposite.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Although the low probability of high-cost types in the world appears to have been generally the case, the distribution of types can certainly vary across regions of the world and over time. Contrary to the world that we appear to occupy, an alternate type of anarchy is possible. If the world consisted of relatively more revisionist types, implying a high probability of violations, political entities would also choose anarchy. In this case, however, the calculation would be that an investment in external enforcement would insufficiently compensate for the rampant violations – many of which would be committed by the entity in question, who wishes not to bring punishment on itself. In this anarchy, life is likely to be indeed nasty, brutish, and short, with frequent or constant violence resulting from self-enforcement. As Wendt (1992) has forcefully stated, 'anarchy is what states make of it', but the distribution of actors' preferences in the world rather than some deep normative structure shapes how they behave under anarchy.

One last implication of this parameter relates to democracy promotion. Since the end of the Cold War, promoting democratization has been an important aspect of foreign policy by the United States and the EU (e.g., Cox *et al.* 2000). The policy, rooted in the findings of the democratic peace theory, is aimed at increasing the probability that states will choose to cooperate at the international level. If the democratic peace thesis is correct, then democracy promotion has an interesting implication for international integration in the context of our theory. If democratization in fact makes states more cooperative at the international level – that is reduce their probability of violations – then democratized states will be *less* likely to choose external enforcement – that is international integration. Because states will be more cooperative, they will need to rely less on external enforcement. Greater democratization at the domestic level thus implies greater disintegration at the international level. While this implication decidedly breaks the Kantian tripod of democracy, economic interdependence and international organizations leading to perpetual peace at the international level (see Russett and Oneal 2001), the ultimate result may be just as pacific.[17]

Costs of integration

In addition to the low density of interaction among rulers, the costs of integration are exceedingly high at the international level. On the material side, these costs are best conceived as total taxes political entities would

[17] This development would also resolve the 'trilemma' that Rodrik (2000) identifies among integrated national economies, nation state and mass politics.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

need to collect to devise an arrangement that can effectively punish a defector for deviating from cooperation. The necessary taxes for punishing individuals at the local level differ crucially from those for punishing political entities at the global level. Punishing an individual for violating an existing social contract – through imprisonment or execution – is relatively inexpensive. By contrast, punishing a political entity as an organized corporate actor would be far more difficult (cf. Ritter and Wolford 2012). Thus, the upfront investment to create a global external enforcement institution would be substantially higher than investing in a similar institution at the local level.

Technological innovations influence the material costs of integration by affecting political entities' capacity to use coercion against others. The emergence of political leaders' ability to apply effective coercion – internally within organized political entities as well as externally toward other entities – was a critical condition for the development of the modern state (e. g., Poggi 1978; Tilly 1992). In other words, the rise of new technologies facilitating coercion and punishment over particular territories and populations enabled leaders to establish the modern state. Despite all the modern technologies, however, political entities' ability to apply effective coercion at the global level is limited. The amount of time necessary even today for deploying aircraft carriers to a particular destination exemplifies this constraint. Even at the top published speed of 20 knots, a US aircraft carrier reaching Strait of Gibraltar from Mayport, Florida, takes 7.6 days. Reaching the Cape of Good Hope takes 24.4 days (O'Rourke 2012, 22).

Moreover, technological constraints that hinder high-density interactions at the global level also make effective monitoring at that level costly. In closer proximity, violations are easier to monitor and detect. At the global level, where the scale is larger, monitoring is more difficult. Accordingly, creating an institution at the global level that can effectively monitor violations to facilitate enforcement would be costly.

The costs of integration need not be restricted to such material ones; integration may also carry subjective costs, and these can be exorbitantly high at the global level. Although neighboring states may have similar cultures, languages, and ethnic compositions, which lower the subjective costs of integration, the same is not true as the distance between states increase (Alesina and Spolaore 2003). Sociocultural differences pose high barriers against leaders choosing to integrate their political entities. While the difference has somewhat subsided in recent years, something as seemingly universal as trade agreements were relatively rare in Asia unlike in all the other regions of the world. The conventional wisdom accounting for this difference has been the high degree of heterogeneity in the region (Duffield 2003).

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

In addition to highlighting these more material costs of integration, our theory calls into question a prevalent argument about the difficulty of international integration. Scholars frequently discuss 'sovereignty costs' as a major source of impediment to greater integration at the international level (e.g., Abbott and Snidal 2000; Moravcsik 2000). However, sovereignty costs need not pose an exogenous constraint on political entities' calculations of how much integration to pursue. Instead, the combination of high costs of integration and low interaction density makes political leaders hesitant about investing in external enforcement at the global level. As Krasner (1999) shows, when politically expedient, rulers have readily chosen to incur 'sovereignty costs' to incorporate external entities into the internal authority structure.

Conclusion

We have provided a rationale by which to understand the anarchic character of the international system. Specifically, political entities prefer to retain the right to use violence in defense of their interests rather than delegate enforcement to a common institution when the density of their cooperative interactions is low enough that the costs of integration are otherwise prohibitive. Where changes in these factors occur locally, we observe both the integration and disintegration of states, as well as variation in the extent to which political entities delegate enforcement functions to common institutions. Anarchy is, in a very real sense, a choice, an enforcement regime chosen in place of the centralized schemes that rulers manage inside states.

With the choice of anarchy explained, what can it then tell us about international relations in general? Most notably, it casts doubt on the utility of anarchy as an explanatory factor of international outcomes, because it is endogenous to other, underlying features of the political environment. Specifically, the volatility of preferences and frequency of violations, the density of interactions, and the costs of building institutions all serve to explain both the ease with which players cooperate *and* the enforcement institutions they choose – self- or external enforcement – to support cooperation. Anarchy, then, is not immutable structure (cf. Waltz 1979), but a symptom of 'congealed tastes' (Riker 1980), one defined by the basic features of cooperative interactions defined above.

Our theory posits a political structure characterized by varying degrees and frequencies of interaction, benefits and costs from cooperation, and delegation problems, and we can use combinations of these factors to explain the occurrence and absence of institutions with robust enforcement power. For example, given the relatively few interactions between

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

political entities at the global level – compared with the density of interactions between, say, neighbors who live on the same street – we should expect both anarchy and few 'strong' institutions to which these entities have granted the right to use violence to punish violations of cooperative agreements. Put differently, the costs of yielding sovereignty to a supra-national body able to use force are prohibitive given the rarity with which states require the use of force to enforce the rules on one another (i.e. war).

Therefore, the apparently shallow nature of international cooperation is not explained by anarchy, because they are both results of factors that encourage self-enforcement and discourage external enforcement. This, of course, is not to say that shallow agreements and anarchy are signs of underlying cooperation problems, because both outcomes are themselves choices made by rulers in a given context of interactions, preference volatility, and costs of cooperation. The logic does, however, suggest where analysts might look to observe the formation of 'deeper' forms of cooperation like political integration and alienation of the right to use force – regions or periods in which the technology of cooperation or the density of interactions produce a need to govern interactions that were previously either infrequent or unimportant.

Further, to the extent that the international system is anarchic, then we should also expect that the most common and most effective international institutions will be those that seek to solve coordination problems through the provision of focal solutions and common knowledge (see, e.g., Voeten 2005). On the other hand, institutions promising to centralize the enforcement of rules through a common institution responsible for violence – perhaps like the League of Nations in its ideal form – are not only unlikely to succeed but also simply unattractive to rulers facing the conditions that make anarchy attractive. Put differently, by choosing anarchy, rulers have already decided that common enforcement is too expensive, and this should perhaps give analysts pause in judging what levels or types of cooperation are feasible in international relations in the first place.

References

Abbott, Kenneth W., and Duncan Snidal. 2000. "Hard and Soft Law in International Governance." *International Organization* 54(3): 421–56.

Alesina, Alberto, and Enrico Spolaore. 2003. *The Size of Nations*. Cambridge: MIT Press.

Alter, Karen J. 1998. "Who are the 'Masters of the Treaty'? European Governments and the European Court of Justice." *International Organization* 52(1): 121–47.

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

Axelrod, Robert, and Robert O. Keohane. 1985. "Achieving Cooperation Under Anarchy: Strategies and Institutions." *World Politics* 38(1): 226–54.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Bull, Hedley. 1977. *The Anarchical Society: A Study of Order in World Politics*. New York: Columbia University Press.

Burley, Anne-Marie, and Walter Mattli. 1993. "Europe Before the Court: A Political Theory of Legal Integration." *International Organization* 47(1): 41–76.

Burnett, Edmund C. ed. 1923. *Letters of the Members of the Continental Congress, 8 vols*. Washington, DC: Carnegie Institution of Washington.

Carr, Edward Hallett. 1964. *The Twenty Years' Crisis: An Introduction to the Study of International Relations* 2nd ed. New York: Harper Torchbooks.

Carrubba, Clifford J. 2005. "Courts and Compliance in International Regulatory Regimes." *Journal of Politics* 67(3): 669–89.

Cox, Michael, G. John Ikenberry, and Takashi Inoguchi, eds. 2000. *American Democracy Promotion: Impulses, Strategies, and Impacts*. Oxford: Oxford University Press.

de Figueiredo, Rui Jose Pacheco Jr., and Barry R. Weingast. 2005. "Self-Enforcing Federalism." *Journal of Law, Economics, and Organization* 21(1): 103–35.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. "Is the Good News About Compliance Good News About Cooperation?" *International Organization* 50(3): 379–406.

Duffield, John S. 2003. "Asia-Pacific Security Institutions in Comparative Perspective." In *International Relations Theory and the Asia-Pacific,* edited by G. John Ikenberry, and Michael Mastanduno, 243–70, Chapter 7. New York: Columbia University Press.

Engel, Charles, and John H. Rogers. 1996. "How Wide is the Border?" *American Economic Review* 86(5): 1112–125.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3): 379–414.

Fearon, James D, and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1): 75–90.

Hendrickson, David C. 2003. *Peace Pact: The Lost World of the American Founding*. Lawrence, KS: University of Kansas Press.

Henkin, Louis. 1979. *How Nations Behave: Law and Foreign Policy*. New York: Columbia University Press.

Hirshleifer, Jack. 1995. "Anarchy and its Breakdown." *Journal of Political Economy* 103(1): 26–52.

Hoffman, Aaron M. 2006. *Building Trust: Overcoming Suspicion in International Conflict*. Albany, NY: State University of New York Press.

Keohane, Robert O, and Joseph S. Nye. 1977. *Power and Interdependence*. Boston: Little Brown.

King, Gary, and Langche Zeng. 2001. "Explaining Rare Events in International Relations." *International Organization* 55(3): 693–715.

Koremenos, Barbara, Charles Lipson, and Dunan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(4): 761–99.

Krasner, Stephen D. 1999. *Sovereignty: Organized Hypocrisy*. Princeton: Princeton University Press.

Kraus, Michael, and Allison Stanger. 2000. "The Past as Prologue." In *Irreconcilable Differences? Explaining Czechoslovakia's Dissolution*, edited by Michael Kraus, and Allison Stanger, 1–6. New York: Rowan and Littlefield.

Kučera, Milan, and Zdeněk Pavlik. 1995. "Czech and Slovak Demography." In *The End of Czechoslovakia*, edited by Jiří Musil, 15–39, Chapter 2. New York: Central European University Press.

Lake, David A. 2009. *Hierarchy in International Relations*. Ithaca: Cornell University Press.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Leff, Carol Skalnik. 1988. *National Conflict in Czechoslovakia: The Making and Remaking of a State, 1918–87*. Princeton: Princeton University Press.

Leff, Carol Skalnik. 1997. *The Czech and Slovak Republics: Nation Versus State*. Boulder, CO: Westview Press.

Mearsheimer, John J. 1994. "The False Promise of International Institutions." *International Security* 19(3): 5–49.

Milner, Helen. 1991. "The Assumption of Anarchy in International Relations Theory: A Critique." *Review of International Studies* 17(1): 67–85.

Moravcsik, Andrew. 1998. *The Choice for Europe: Social Purpose and State Power from Messina to Maastricht*. Ithaca, NY: Cornell University Press.

Moravcsik, Andrew. 2000. "The Origins of Human Rights Regimes: Democratic Delegation in Postwar Europe." *International Organization* 54(2): 217–52.

Musil, Jiří. 1995. "Introduction." In *The End of Czechoslovakia,* edited by Jiří Musil, 1–11, Chapter 1. New York: Central European University Press.

Muthoo, Abhinay. 1999. *Bargaining Theory with Applications*. Cambridge: Cambridge University Press.

Nitsch, Volker. 2003. "National Borders and International Trade: Evidence from the European Union." *Canadian Journal of Economics* 33(4): 1091–105.

Nugent, Neil. 2003. *Government and Politics of the European Union*, 5th ed. Durham, NC: Duke University Press.

Oatley, Thomas H. 2012. *International Political Economy*, 5th ed. New York: Longman.

Onuf, Peter S. 1983. *The Origins of the Federal Republic: Jurisdictional Controversies in the United States, 1775–81*. Philadelphia, PA: University of Pennsylvania Press.

Organski, Abramo Fimo Kenneth 1958. *World Politics*. New York: Knopf.

O'Rourke, Ronald. 2012. *Navy Nuclear Aircraft Carrier (CVN) Homeporting at Mayport: Background and Issues for Congress. CRS Report for Congress R40248*. Washington, DC: Congressional Research Service.

Poggi, Gianfranco. 1978. *The Development of the Modern State: A Sociological Introduction*. Stanford, CA: Stanford University Press.

Powell, Robert. 1994. "Anarchy in International Relations Theory: The Neorealist-Neoliberal Debate." *International Organization* 48(2): 313–44.

Powell, Robert. 2004. "Bargaining and Learning While Fighting." *American Journal of Political Science* 48(2): 344–61.

Reinhardt, Eric. 2001. "Adjudication without Enforcement in GATT Disputes." *Journal of Conflict Resolution* 45(2): 174–95.

Riker, William H. 1980. "Implications from the Disequilibrium of Majority Rule for the Study of Institutions." *American Political Science Review* 74(2): 432–46.

Ritter, Emily H., and Scott Wolford. 2012. "Bargaining and the Effectiveness of International Criminal Regimes." *Journal of Theoretical Politics* 24(2): 149–71.

Rodrik, Dani. 2000. "How Far Will International Economic Integration Go?" *Journal of Economic Perspectives* 14(1): 177–86.

Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica* 53(1): 97–109.

Russett, Bruce, and John Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: W.W. Norton & Company.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. New York: W.W. Norton & Company.

Schweller, Randall L. 1996. "Neorealism's Status Quo Bias: What Security Dilemma?" *Security Studies* 5(3): 90–121.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Spruyt, Hendrik. 1994. “Institutional Selection in International Relations: State Anarchy as Order.” *International Organization* 48(4): 527–57.

Tilly, Charles. 1992. *Coercion, Capital, and European States, AD 990–1992* Revised Edition. Cambridge, MA: Blackwell.

Vasquez, John A. 2009. *The War Puzzle Revisited*. Cambridge: Cambridge University Press.

Voeten, Erik. 2005. “The Political Origins of the UN Security Council’s Ability to Legitimize the Use of Force.” *International Organization* 59(3): 527–57.

Wagner, R. Harrison. 2007. *War and the State: The Theory of International Politics*. Ann Arbor: University of Michigan Press.

Waltz, Kenneth N. 1970. “The Myth of National Interdependence.” In *International Cooperation*, edited by Charles P. Kindleberger, 205–23. Cambridge, MA: MIT Press.

Waltz, Kenneth N. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.

Weingast, Barry R. 1997. “The Political Foundations of Democracy and the Rule of Law.” *American Political Science Review* 91(2): 245–63.

Wendt, Alexander. 1992. “Anarchy is What States Make of it: The Social Construction of Power Politics.” *International Organization* 46(2): 391–425.

Wendt, Alexander. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.

Werner, Suzanne, and Amy Yuen. 2005. “Making and Keeping Peace.” *International Organization* 59(2): 261–92.

Žák, Václav. 1995. “The Velvet Divorce–Institutional Foundations.” In *The End of Czechoslovakia,* edited by Jiří Musil, 245–68, Chapter 13. New York: Central European University Press.

Appendix

Proofs

Proof of Proposition 1: Begin with an aggrieved player i’s choice over punishing or tolerating a violation. It punishes when $u_i(P) \geqslant u_i(\neg P) \Leftrightarrow -h \geqslant -c_i$, which is sure to be true, since $c_{\mathrm{i}} > h$; further, this is a best response for any beliefs i might hold over $-i$’s type.

In the proposed equilibrium, high-cost types defect and low-cost types cooperate. To show that these choices are in equilibrium, we characterize optimal behavior for each player-type. When player i is high cost, or $c_i = \overline{c}_i$, defection is optimal when

$$EU_i(D) > EU_i(C) \Leftrightarrow \phi(0) + (1-\phi)(tb) > \phi(-h) + (1-\phi)(b-\overline{c}_i),$$

which is sure to be true, since $\overline{c}_i > b$ and $h > 0$, ensuring that the expected utility for cooperation is negative. When player i is low cost, or $c_i = \underline{c}_i$, cooperation is optimal when

$$EU_i(C) \geqslant EU_i(D) \Leftrightarrow \phi(-h) + (1-\phi)(b-\underline{c}_i) > \phi(0) + (1-\phi)(tb),$$

which is true when $h \leqslant ((1-\phi)(b-\underline{c}_i) - tb)/\phi \equiv \overline{h}$ and $t < 1 - \underline{c}_i/b - h\phi/(b(1-\phi)) \equiv \overline{t}$, as stipulated by the equilibrium.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press

Therefore, the separating equilibrium exists when $h \leqslant \overline{h}$ and $t \leqslant \bar{t}$, and the *ex ante* probability that (a) both players defect is $\phi\phi$, (b) 1 defects while 2 cooperates is $\phi(1-\phi)$, (c) 1 cooperates while 2 defects is $(1-\phi)\phi$, and (d) both cooperate is $(1-\phi)(1-\phi)$.

Proof of Proposition 2: In the proposed equilibrium, high-cost types defect and low-cost types cooperate. To show that these choices are in equilibrium we characterize optimal behavior for each player-type. When player *i* is high cost, or $c_i = \bar{c}_i$, defection is optimal when

$$EU_i(D) > EU_i(C) \Leftrightarrow \phi(-\rho) + (1-\phi)(b-\rho) \geqslant \phi(bj - \bar{c}_i) + (1-\phi)(b - \bar{c}_i),$$

or when $\rho \leqslant \bar{c}_i - bj\phi \equiv \bar{\rho}$. When player *i* is low cost, or $c_i = \underline{c}_i$, cooperation is optimal when

$$EU_i(C) \geqslant EU_i(D) \Leftrightarrow \phi(bj - \underline{c}_i) + (1-\phi)(b - \underline{c}_i) \geqslant \phi(-\rho) + (1-\phi)(b-\rho),$$

or when $\rho \geqslant \underline{c}_i - bj\phi \equiv \underline{\rho}$.

Therefore, the separating equilibrium exists when $\underline{\rho} \leqslant \rho \leqslant \bar{\rho}$, and the *ex ante* probability that (a) both players defect is $\phi\phi$, (b) 1 defects while 2 cooperates is $\phi(1-\phi)$, (c) 1 cooperates while 2 defects is $(1-\phi)\phi$, and (d) both cooperate is $(1-\phi)(1-\phi)$.

Proof of Proposition 3: Recall, first, that external enforcement is the equilibrium institution if both players prefer it to self-enforcement in expectation and, second, that its existence also requires $\rho \geqslant \underline{c}_i - bj\phi \equiv \underline{\rho}$. Therefore, using player *i*'s expected utility for each enforcement institution as defined in equations (1) and (2), external enforcement is the equilibrium institution when $-k + \delta EU_i(\text{ee}) > \delta EU_i(\text{se})$, or

$$\begin{gathered} -k + \delta[\phi\phi(-\rho) + \phi(1-\phi)(b-\rho) + (1-\phi)\phi(jb-\underline{c}_i) + (1-\phi)(1-\phi)(b-\underline{c}_i)] \\ > (1-\phi)(1-\phi)(b-\underline{c}_i) + (1-\phi)\phi(-h) + \phi(1-\phi)(tb) \end{gathered}$$

This inequality is true when

$$\delta > \frac{k}{\phi((1-\phi)(b(j-t+1)+h-\underline{c}_i)-\rho)} \equiv \underline{\delta},$$

but in order for the constraint to be satisfied at possible values of δ, that is for $0 < \underline{\delta} < 1$, the following two constraints must also bind:

$$k < \phi((1-\phi)(b(j-t+1)+h-\underline{c}_1)-\rho) \equiv \overline{k}$$

and

$$\rho < (1-\phi)(b(j-t+1)+h-\underline{c}_1) \equiv \hat{\rho}.$$

Therefore, when three constraints are satisfied – $\underline{\rho} \leqslant \rho \leqslant \min\{\bar{\rho}, \hat{\rho}\}$, $\delta > \underline{\delta}$, and $k < \overline{k}$ – external enforcement is the equilibrium institution. If at least one constraint is not satisfied, then self-enforcement is the equilibrium institution.

Supplemental analysis of differential power

In the baseline version of the self-enforcement subgame analyzed above, both players are equally 'powerful' in that their payoffs for punishing the other player are equal. However, power is often distributed unequally. In this section, we explore how differing levels of power affect our main claims and show that, while differential power can change the ease with which the critical constraints supporting external enforcement are satisfied – that is, making them more restrictive – it does not change the fundamental answer to our basic question. Differential power can make external enforcement less likely, but when it occurs, it still does so for the same reasons: high interaction density, low costs of integration, and less painful punishments; in other words, the 'when' changes, but the 'why' does not.

Suppose that, under self-enforcement, punishing another player's defection is costly, such that player *i* pays a cost, h_i, when it plays P. When h_i is low, player *i* is strong – that is, it has a higher payoff for inefficiency outcomes like war – and when h_i is high, player *i* is weak. When $h_i \leqslant c_i$, as it in the main model, threats to punish unilateral defections are credible, but when $h_i > c_i$, they are not. Thus, aggrieved players punish unilateral defections when the former is true, and they do not punish them when the latter is true. In cases of mutual defection, players do not punish, since $-h_i < 0$, but the results in this section are robust to a more complicated model in which players would have some incentive to punish, and therefore receive $-h_i$, following mutual defection. This alters player *i*'s expected utility for self-enforcement, such that it receives

$$EU_i(\text{se}) = (1-\phi)(1-\phi)(b-\underline{c}_i) + (1-\phi)\phi(-h_i) + \phi(1-\phi)(tb),$$

as opposed to $(1-\phi)(1-\phi)(b-\underline{c}_i) + (1-\phi)\phi(-h) + \phi(1-\phi)(tb)$ in the baseline model.

How do varying levels of power affect the incentive to choose external enforcement? Straightforwardly, powerful players do better under self-enforcement than weak players, making them on average less willing to submit to external enforcement. We can show this by solving $-k+\delta\ EU_i(\text{ee}) > EU_i(\text{se})$, which – as above – yields a set of three constraints that, if satisfied, produce external enforcement:

$$\delta > \frac{k}{\phi((1-\phi)(-\underline{c}_i + b(j-t+1) + h_i) - \rho)},$$

$k < \phi((1-\phi)(-\underline{c}_i + b(j-t+1) + h_i) - \rho)$, and $\rho < (1-\phi)(-\underline{c}_i + b(j-t+1) + h_i)$. Substantively, each constraint becomes easier to satisfy as h_i increases – that is, as a player becomes weaker – but the mechanisms behind the choice of external enforcement remain the same, as evidenced by the fact that the constraints emerge in terms of the same parameters, in the same direction.

Thus, while external and self-enforcement equilibria exist for the same reasons when we take power into account, it is nonetheless true that weaker players prefer external enforcement for a wider range of the parameter space. However, since external enforcement cannot occur without each player's consent, strong players remain the gatekeepers; even when weak players prefer external enforcement, strong players will veto it unless they also prefer it, which requires relatively higher interaction densities, lower costs of integration, and less-painful external punishment.

https://doi.org/10.1017/S1752971913000304 Published online by Cambridge University Press