

A Theory of Neutrality Rights in War*

Scott Wolford
University of Texas
swolford@austin.utexas.edu

February 19, 2014

Abstract

I analyze a model of war-fighting and war expansion in which a belligerent can violate the laws of neutrality to gain a military advantage and a third party decides whether to join the war against the belligerent. The third party is uncertain over the value of fighting, such that it wishes to intervene against an expansionist belligerent but to remain neutral if the belligerent is not expansionist. Absent codified neutrality rights, both unnecessary interventions against non-expansionist belligerents and regretted neutrality when the belligerent turns out to have been expansionist occur in equilibrium. When states can coordinate on an equilibrium in which only expansionist types violate neutrality, both unnecessary interventions and regretted neutrality are less likely. International law can thus produce self-enforcing cooperation in the absence of centralized enforcement. However, regretted neutrality is most common under full compliance; the best equilibrium for non-belligerent states turns out to be a separating equilibrium that identifies desirable interventions by drawing a line and inducing expansionist states to cross it. Violations of neutrality need not indicate that international law is ineffective; to the contrary, violations make it effective by solving a third party's information problem over the desirability of intervention.

*Draft version. Please do not cite without permission. Thanks to Amy Liu, Emily Ritter, Randy Stone, and Rachel Wellhausen for helpful comments and suggestions.

The Hague Convention of 1907 formalized centuries of customary law conferring special rights on neutral countries during wartime, declaring that “[t]he territory of neutral Powers is inviolable” (§V.1) and “[t]he fact of a neutral Power resisting, even by force, attempts to violate its neutrality cannot be regarded as a hostile act” (§V.10). Despite these provisions, as well as similar ones for neutral shipping (§XIII), compliance with the laws of neutrality varies both across and within wars. In August 1914, Germany violated Belgian neutrality in the process of attacking France, triggering British entry into the First World War—hardly unanticipated in Germany (Hastings 2013, Ch. 2)—and doing little to bolster faith in the effectiveness of international law (cf. Carr 1964). On the other hand, as Japanese ships approached Chemulpo on the Korean coast as part of a surprise attack on Russian assets in the region ten years earlier, the task force’s commander ordered that enemy warships not be fired upon in the presence of neutral ships from France, the United Kingdom, and the United States, lest those countries enter the war on Russia’s side (Connaughton 2004). Failing to destroy Russian ships threatened to undermine Japan’s opening gambit in the Russo-Japanese War. Nonetheless, despite issuing no declaration of war before the attack and arguably violating Korean neutrality at the same time (Hershey 1904), Japan was scrupulous about honoring the rights of the Western neutrals, even as their ability to project force in the region was limited.

These examples give rise to two questions. First, why do some belligerents honor the rights of neutral countries while others do not? Second, and more broadly, what can state choices over honoring and violating neutrality rights tell us about the effectiveness of international law, particularly in the high-stakes realm of interstate war? To answer these questions, I analyze a game-theoretic model of war-fighting strategy and war expansion in which a belligerent chooses to honor or violate the neutrality of another state, after which another party—who may or may

not be the aggrieved state—chooses whether to join the war against the belligerent. Though costly, intervening in the war increases the chances that the belligerent suffers defeat, thereby diminishing a potential future threat, but joining the war is only worth the cost if the targeted belligerent would have posed a substantial threat after winning today's war.

I build the theory around two main premises. First, international law establishes common beliefs and expectations about acceptable behavior, which helps states coordinate on particular sets of equilibrium strategies in response to violations (Morrow 2002, 2007). The effects of coordination emerge both directly in the equilibrium that states choose to play and, indirectly, in lowering the costs of intervention imposed by observers if neutrality is violated. Second, contrary to most extant work on international cooperation, I assume that states are concerned not with each other's willingness to honor international law, profiting or losing only from decisions over compliance (cf. Carrubba 2005, Downs, Rocke and Barsoom 1996), but with the scope of each other's foreign policy ambitions. Is a belligerent satiable or expansionist? Third parties would like to intervene against expansionist powers that pose a future threat, yet they would prefer to remain neutral if a belligerent's aims are truly limited, or satiable. I argue below that international law, as codified in rights of neutrality, can facilitate these judgments and, as a result, decisions over honoring and violating the rights of neutrals in war.

By focusing on some unique features of the laws of neutrality and the role of international law in creating shared expectations about actions and their consequences, the model sheds new light on both war expansion and the effects of international law. First, clearly stipulated neutrality rights can help states avoid unnecessary interventions and choose desirable ones, because they facilitate coordination of expectations on particular equilibria: specifically, those in which only expansionist states violate neutrality rights. Even when they do not produce full

compliance, laws that stipulate the rights of neutrals may still have a sizable effect by encouraging states to intervene against belligerents that violate them. By separating belligerents according to their ambitions—expansionist versus satiable—international law may “work” not because it creates compliance that would otherwise not exist, but because it sorts belligerents according to their willingness to provoke an expanded war. Full compliance, in fact, would lead to higher rates of regretted neutrality, while the separating equilibrium in which expansionist states *do* violate neutrality ensures that third parties can effectively tailor intervention decisions to belligerents’ likely ambitions. Thus, violations of neutrality rights need not indicate that international law is ineffective; to the contrary, they make international law effective by solving a third party’s information problem over the desirability of intervention.

International Law, Compliance, and Punishment

To the extent that international law is enforced, it is only because states find it in their interest to punish those others that violate the rules. This renders the proliferation of international law and compliance with it puzzling: why, in the absence of a centralized authority able to punish rule-breakers, do states comply at all? One set of answers focuses on the endogenous nature of the terms of cooperation: states only sign agreements with which they expect to comply, rendering high rates of compliance spurious to actions that states would already take absent legal obligations (Carrubba 2005, Downs, Rocke and Barsoom 1996, Kim and Wolford 2014). Additional explanations invoke the signaling value of ratification and compliance (Chapman and Chaudoin 2013, Simmons and Danner 2010), the realization of a natural propensity to comply (Chayes and Chayes 1993), the activation of interested domestic players (Conrad and Ritter

2013, Dai 2005), or in the role of explicit, codified rules in helping aggrieved states coordinate punishments against violators (Morrow 2002, Reinhardt 2001). On the other hand, compliance may simply be spurious to a lack of competing incentives (Valentino, Huth and Croco 2006).

Whatever accounts for compliance, states often make decisions over violating international law long after acceding to the protocols that govern their action. This is particularly true with respect to the laws of neutrality during war. The rights and obligations of neutral countries constituted a set of customary rules and practices long before their codification as part of The Hague Convention (Bull 1977, Ch. 2), where states acceded to particular laws under “a veil of ignorance of which wars [would] be fought in the future” (Morrow 2002, pp. S54-S55). Once embroiled in a war, states consider the value of compliance by calculating costs and benefits associated with threats of reciprocal punishment (Morrow 2007, Wallace 2012) and, in the case of violating neutrality, the possibility of new belligerents joining the war. Faced with these immediate wartime decisions, the coordination value of having stipulated the rights and privileges of neutrals comes to the fore, especially when they specify what constitutes a violation of neutrality and how neutrals or their guarantors may respond.

The laws of neutrality can have two effects on behavior. First, codified rules can help potential violators and potential punishers agree on which equilibrium among many possible equilibria is likely to be played (Morrow 2002). In particular, clear neutrality rights may confirm for a belligerent what a third party will view as a violation and increase its expectation that violation will result in punishment; if players do not agree on both of these issues, then deterrence of violations is impossible.¹ Second, clarity of the right to retaliation or resistance by force also

¹But see Valentino, Huth and Croco (2006) for evidence that signing relevant conventions does not appear to have an impact on the targeting of civilians in interstate war.

coordinates expectations among other actors, be they other states or a third party's domestic constituency, about whether intervention is justified or legitimate.² Specifically, violations of neutrality may make it easier to rally domestic support, as happened in Britain following the violation of Belgian neutrality in 1914 (Hastings 2013, pp. 85-102), or reassure others about an intervener's own intentions (cf. Voeten 2005), having the direct material effect of lowering the costs of intervention. The model below analyzes both dynamics, but it departs from extant work in its characterization of the desirability of punishing violations.

Coordinating punishment is especially relevant in explaining compliance through threats of reciprocity, which support laws limiting violence in war or the treatment of prisoners of war (Morrow 2001, 2002, 2007, Posner 2003). These laws govern the actions of belligerents with respect to their co-belligerents. However, explanations that rest on reciprocity share an assumption that does not track well substantively with violations of neutrality laws, because they govern the behavior of belligerents with respect to non-belligerents. Theories concerned with reciprocity assume that punishment is inherently costly for those who would carry it out; put differently, states would have no incentive to punish each other if no violations of the rules had occurred (Carrubba 2005, Downs, Rocke and Barsoom 1996, Kim and Wolford 2014). In many instances of cooperation supported by threats of reciprocal punishment, such as violations of trade agreements (Reinhardt 2001) or the abuse of prisoners of war (Morrow 2007, Wallace 2012), this is surely reasonable. However, when states consider how to respond to violations of neutrality rights, this assumption may not be appropriate, because punishment is not based on reciprocity. In fact, when it comes to fighting some belligerents, it may be *desirable*.

²For a similar argument about the relationship between the United Nations Security Council, intervention decisions, and international support, see Hurd (2005), Voeten (2005), and Chapman (2011).

To see how, consider the anticipated consequences of the German and Japanese violations of neutrality discussed above: an expansion of the war either by or on behalf of aggrieved states. Expansion of the war need not be mere punishment, though, parties may view joining the war as attractive whether or not laws were violated. When states join ongoing wars, they do so mindful of the costs of fighting, their ability to affect the outcome, the issues at stake, and—very often—their beliefs about future threats posed by the belligerents (see [Altfeld and Bueno de Mesquita 1979](#), [Waltz 1979](#), [Wolford 2014](#)). Potential interveners, then, must try to gauge whether a belligerent is “generally aggressive” or simply “dissatisfied with the particular state it attacked” ([Powell 1999](#), pp. 193-194). In other words, observers are concerned over the scope of a belligerent’s ambitions; is it satiable, and therefore likely to stop after victory, or is it expansionist, and therefore unlikely to be satisfied after today’s war? This poses a dilemma, in that defeating an expansionist belligerent alongside others in today’s war may be easier than defeating it alone in the future ([Powell 1999](#), Ch. 5), but doing so is costly, and intervention is wasteful if the belligerent turns out to have been satiable. Put differently, with full information, some potential interveners would want to join the war against an expansionist belligerent *regardless* of its respect for the rights of neutrality.

Contrast this with the logic of reciprocal punishment that sustains limitations on violence in war, where neither side has an incentive to punish the other—that is, violate laws they hope to uphold—unless the other has first violated restrictions ([Morrow 2002, 2007](#)). Punishment may be sometimes necessary to preserve deterrence, but it is known to be generally undesirable in and of itself. However, when joining a war might provide other benefits that other options do not, such as forestalling the adverse shift in the distribution of power that would follow the victory of an expansionist belligerent, then intervention may be inherently desirable. Put dif-

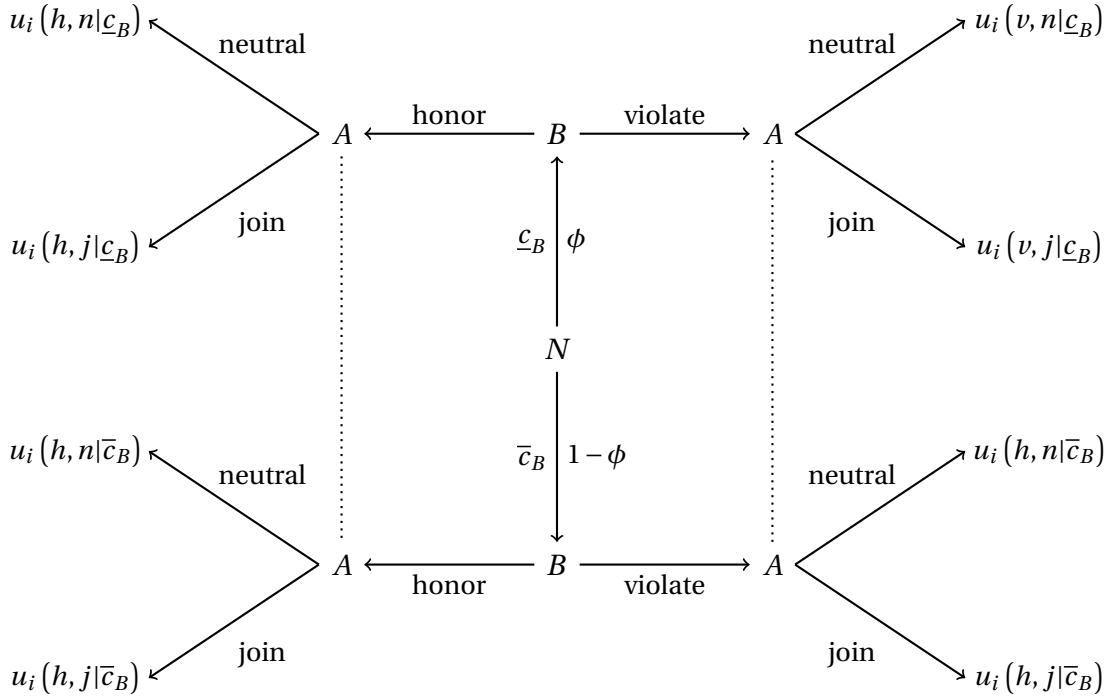
ferently, ensuring that a trading partner follows agreed-upon tariff rules is a rather different goal than preventing the victory of a threatening expansionist state. As such, conceiving of the laws of neutrality as enforced by threats of reciprocal punishment may not be as useful as examining how codified neutrality rights might help solve third parties' informational problems about the desirability of joining the war.

Understand how the laws of neutrality affect the behavior of states requires a focus on the coordinating role that international law plays in (a) clarifying what constitutes acceptable and unacceptable behaviors and (b) establishing shared expectations over what happens in the event of violations. Below, I introduce a model of war-fighting strategy and potential war expansion, where a belligerent must choose whether to violate neutrality rights before a third party state chooses whether to intervene in the war against it. International law can have two potential effects, both rooted in the coordination of expectations. First, it can help players coordinate on particular equilibria when multiple equilibria exist. Second, it may lower the costs of intervention in the event of violations by coordinating the expectations of additional states and an intervener's own domestic constituency, which could lead to more credible threats of punishment. Neutrality rights must also work in the shadow of the third party's uncertainty, as it begins the game unaware of the desirability of intervention. I then explore how the law of neutrality affects patterns of both compliance with neutrals' rights and the expansion of war.

The Model

Suppose that a belligerent state (B) is engaged in a war in which it can increase its chances of victory by violating the rights of some neutral state. It might invade the neutral along the way

Figure 1: A model of war-fighting strategy and war expansion



to its main target, as Germany did in the First World War, or it might launch an attack on its main target that puts at risk a neutral's assets or military forces, as Japan opted not to do in the opening hours of the Russo-Japanese War. Next, suppose that some non-belligerent state (A), which may or may not be the state whose neutrality B can exploit, can choose whether to remain neutral or intervene against the belligerent. While A can observe B 's actions, it is uncertain over the desirability of intervention; if the belligerent is expansionist, placing a high value on winning today's war because it plans further aggrandizement, state A would like to intervene and help defeat the belligerent now, but if B is not expansionist (here, satiable), A would rather save the costs of intervening and remain out of the war.

As shown in Figure 1, Nature first chooses B 's type, expansionist with probability ϕ and satiable with probability $1 - \phi$, and reveals it only to B ; thus, B is the "informed" player, while

A is the “uninformed” player. When expansionist, B assesses the benefits of violating neutrality rights more highly than when it is satiable, because its interests encompass not only today’s war but also the position in which it will be placed for future expansion. On the other hand, when satiable, B is interested only in defeating today’s opponent, and violating neutrality rights offers no additional benefits, since it does not plan to turn against future enemies after today’s war. Germany, for example, coveted the Low Countries in World War II not merely for their utility in the invasion of France but also their potential role in securing the Continent and threatening the United Kingdom; if Germany wished only to defeat France, on the other hand, then violating neutrality would have carried less, though perhaps not negligible, inherent value.

To formalize the scope of B ’s ambitions, I define its type as the costs it pays (relative to the benefits) for violating neutrality rights ($c_B > 0$). Therefore, when $c_B = \underline{c}_B$, B is expansionist, and when $c_B = \bar{c}_B$, it is satiable, such that $0 < \underline{c}_B < \bar{c}_B$. The scope of B ’s ambitions also shapes A ’s incentives to intervene in the war, such that it would prefer, all else equal, to remain neutral if B is satiable and poses no future threat but to intervene if B is expansionist and is worth fighting today. However, since A knows only the probability distribution from which B ’s type is drawn, it begins the game uncertain over the desirability of intervention, and it uses the type distribution to inform its prior beliefs ϕ_0 over the belligerent’s type, such that it believes at the beginning of the game that B is expansionist with probability $\phi_0 = \phi$.

Next, B chooses whether to violate (v) or honor (h) the rights of a neutral state. This can be either A or some other neutral whose fate A can observe; A ’s payoffs are sufficiently general to cover both cases. If B violates neutrality rights, it receives a military boost, such that it experiences an increased chance of victory and a position useful for future expansion. Following B ’s decision, A decides whether to join in the war (j) against the belligerent, which is costly up

front but reduces B 's chances of victory, or to remain neutral (n). Thus, the game can end in one of four ways: (a) B honors neutral rights and A remains neutral (h, n); (b) B honors neutral rights and A intervenes (h, j); (c) B violates neutral rights and A remains neutral (v, n); and (d) A violates neutral rights and A intervenes (v, j).

After A 's decision, each player receives a payoff that reflects the expected outcome of the war, less the costs of violating neutrality c_B or intervening in the war ($c_A > 0$). To capture the possibility that codified laws of neutrality may lower the costs of intervention in the event of a violation, I modify A 's costs of joining the war by the scalar $d^v \in (0, 1]$, where $v = 1$ if B violates and zero otherwise, such that its costs of intervention are $d^v c_A$. By either activating domestic constituencies or drawing the support of other states, A 's costs for intervening therefore may—but need not—be reduced in the presence of codified neutrality rights.

If B wins the war, it receives 1, and it receives zero if it loses; it also pays the upfront cost c_B if it violates neutrality.³ For its part, A receives zero—its best outcome—if B loses the war. However, should B win, A receives the reduced-form payoff $u_A(c_B) \leq 0$, which decreases in the scope of B 's ambitions, such that $u_A(\underline{c}_B) < u_A(\bar{c}_B) \leq 0$. Thus, A 's worst outcome is victory by an expansionist B , while victory by a satiable B is at least as unfavorable as either type of B suffering defeat. Finally, B 's probability of winning the war, $p_B \in (0, 1)$, is

$$p_B \equiv \rho \lambda^v + b - ja,$$

where $\rho \in (0, 1)$ is an underlying probability of victory modified by each side's military capabilities ($b > 0$, $a > 0$) and a military boost $\lambda > 1$ that B receives only if it violates neutrality (i.e., if

³Since B has already decided to fight, all other costs of war are sunk and excluded from the analysis.

$v = 1$).⁴ As such, B 's probability of victory increases in its own capabilities (b) and decreases in A 's (a), where $j = 1$ if A intervenes against B and $j = 0$ if A remains neutral, removing A 's contribution from the equation.⁵

We can now state each player's payoff function. For a generic type of B ,

$$u_B = \begin{cases} (\rho + b) & \text{if } h, n \\ (\rho + b - a) & \text{if } h, j \\ (\rho\lambda + b) - c_B & \text{if } v, n \\ (\rho\lambda + b - a) - c_B & \text{if } v, n, \end{cases}$$

and for A , again with respect to an arbitrary type of B ,

$$u_A = \begin{cases} (\rho + b) u_A(c_B) & \text{if } h, n \\ (\rho + b - a) u_A(c_B) - c_A & \text{if } h, j \\ (\rho\lambda + b) u_A(c_B) & \text{if } v, n \\ (\rho\lambda + b - a) u_A(c_B) - dc_A & \text{if } v, n \end{cases}$$

These utility functions capture the essentials of the strategic problem. B is tempted to violate neutral rights, especially if the costs of doing so are not compounded by A 's intervention, while A is uncertain over the value of paying the costs of intervention, which may be wasted if B turns out to have been satiable. This also highlights the key distinction between the present model

⁴For other uses of the "difference" form of the contest success function, see Che and Gale (2000) and Jia, Skaperdas and Vaidya (2013).

⁵To ensure that $p_B \in (0, 1)$, I impose $(ja - b)/\lambda^v < \rho < (1 + ja - b)/\lambda^v$ throughout the analysis.

and other treatments of international law: if only *A* knew *B*'s type, it would wish to intervene in the war regardless of whether or not *B* complies with neutrality rights. To the extent that *A* does “punish” violations in this model, the violation that triggers the punishment may be spurious to its effects on *A*'s beliefs over *B*'s type. Nonetheless, as I show in the next section, the international law of neutrality can have a substantial effect on the behavior of states even when they do not value inherently each other's compliance.

Analysis

Like many signaling games, the one analyzed here has multiple equilibria. For some combinations of parameters, there are several equally rational ways to play the game, leading to different sets of outcomes, depending on what players believe about each other's chosen strategies. I distinguish three types of equilibria by their effects on *A*'s beliefs about *B*'s type. First, in a *separating* equilibrium, each type of *B* takes a unique action that distinguishes itself from the other type, which resolves *A*'s uncertainty. Second, in a *pooling* equilibrium, player-types take the same action, preventing *A* from updating its beliefs. Finally, in a *semi-separating* equilibrium, one player-type takes a unique action but another randomizes, allowing *A* to only partially update its beliefs. I then use the possible effects of international law—(a) coordinated expectations of punishment and (b) lowered costs of intervention—to explain why states might play one equilibrium as opposed to others.

For laws of neutrality to have some effect on state behavior, their presence must be linked to rates of and responses to violations of neutrality that do not occur in their absence. Determining what would happen in the absence of neutrality rights is difficult, though, if their effect

is to coordinate expectations on a particular equilibrium. Suppose, for example, that codified neutrality rights can lead states to coordinate on expectations that A will intervene if B violates, but that states can develop those expectations absent the law; in this case, neutrality rights can have no effect. This would be particularly relevant in situations where only one equilibrium, especially one based on strong expectations of punishment, happens to exist. However, where multiple equilibria exist, particularly ones in which there is no strong expectation of punishment, then coordination can meaningfully alter state behavior by aiding the selection of equilibria with different rates of and responses to violations. Therefore, I focus on a baseline case in which punishment is rare and, on the whole, ineffective: a semi-separating equilibrium in which the expansionist B always violates neutral rights and the satiable B , expecting intervention to be sufficiently unlikely, violates probabilistically.⁶

Three factors make the opportunistic-violation equilibrium attractive as a baseline. First, both expansionist and satiable types violate neutrality rights, the latter doing so opportunistically due to a weak threat of punishment. Second, it exists alongside other equilibria, some involving more credible threats of punishment, where the coordinating effects of international law can most plausibly change behavior by facilitating the play of other sets of strategies. Finally, as stated in Proposition 1, this equilibrium exists when A 's ex ante expectation about the probability of expansionism is relatively low; truly expansionist states are believed to be rare, which tracks well substantively with the empirical record.

Proposition 1 (Opportunistic Defection). *When $\phi \leq \min \{ \underline{\phi}, \bar{\phi} \}$, the following strategies and be-*

⁶There also exists an equilibrium in which both types of B violate while A remains neutral. However, that A would *never* find intervention attractive is empirically implausible, and I focus instead on the opportunistic violations equilibrium as the most satisfying example of a world without codified neutrality rights. It is also possible to support a pooling equilibrium in which both types of B violate while A is sure to intervene, regardless of B 's action, but only with implausible out-of-equilibrium beliefs—specifically, A would have to believe B to be expansionist if it opted to honor neutrality rights. As such, I reject this equilibrium as implausible.

liefs constitute a Perfect Bayesian Equilibrium. If expansionist, B violates. If satiable, B violates with probability g^ and honors with probability $1 - g^*$. If B honors, A believes $\phi' = 0$ and does not intervene. If B violates, A believes $\phi' = \chi$ and intervenes with probability y^* , remaining neutral with probability $1 - y^*$.*

Proposition 1 outlines the strategies and beliefs that make up this semi-separating equilibrium. While the expansionist type is sure to violate, the satiable type does so probabilistically, randomizing between mimicking the expansionist and revealing itself by honoring neutrality, which renders A indifferent over joining the war in the event of a violation. As a result, A knows that B is satiable when it honors neutrality, and A is happy to remain neutral. When A observes a violation, though, it knows only that an expansionist B would certainly have violated and a satiable B might have done so. A raises its estimate of the probability that B is expansionist, though it cannot rule out the possibility of a wasteful intervention. As such, it responds by randomizing between intervention and neutrality in order to ensure that the satiable type of B is itself indifferent over honoring and violating.

This is nearly the worst-case scenario for the international laws of neutrality. The expansionist type of B violates neutral rights because the threat of intervention is insufficient to deter it, and even the satiable B is occasionally tempted by the promise of a military advantage as well. For its part, state A finds it difficult to judge the desirability of intervention by observing B's behavior, leading to both wasted interventions, where A joins the war against B even when it is the satiable type, and regretted neutrality, where A remains on the sidelines only for B to turn out to be expansionist. International law can have one or both of two effects where, absent codified neutral rights, this equilibrium exists. First, it may coordinate expectations on an-

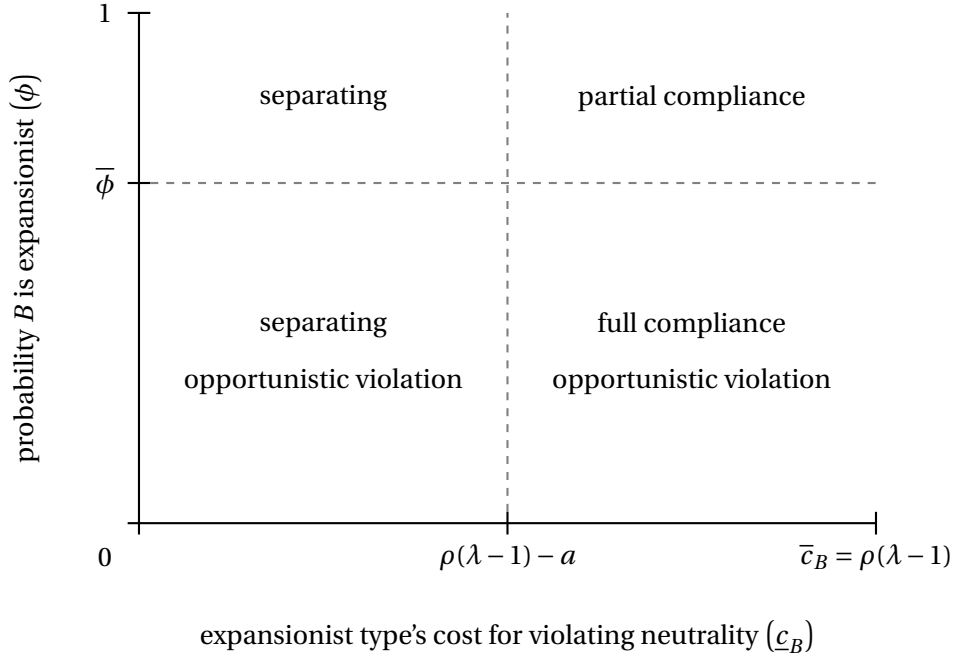
other equilibrium with different rates of violation and responses to it by facilitating equilibrium selection. Second, by altering A 's costs for intervention, it may eliminate the opportunistic-violations equilibrium and see it replaced by another one.

To fix ideas, I consider the first effect in isolation before considering both effects together. Isolating the equilibrium selection mechanism requires that I restrict the costs of intervention following violation such that $d = 1$, which means that A pays the same costs for joining the war whether or not B violates or honors. Here, international law's only effect can be to move the players from one equilibrium to another. Next, I relax this assumption and allow $d < 1$, which means that A pays lower costs for joining the war if B violates neutrality, thanks to the loosening of domestic constraints or the facilitation of diplomatic support from other states. Both factors can have meaningful effects on state behavior, and in the sections that follow I explain these effects in terms of moving states away from the opportunistic violations equilibrium and into different patterns of behavior with different rates of and responses to violations.

Equilibrium Selection

Figure 2 plots the existence of several equilibria as a function of A 's prior belief that B is expansionist ($\phi_0 = \phi$) and the expansionist type's costs for violating neutrality (\underline{c}_B). Moving up along the vertical axis, A begins the game increasingly certain that B is expansionist as ϕ approaches one, beginning from certainty that B is satiable when $\phi = 0$. The horizontal axis tracks the extent of B 's possible expansionism, such that near the origin the difference between \underline{c}_B and \bar{c}_B is at its largest, but moving right along the horizontal axis, the difference between expansionist and satiable types begins to disappear. This allows for a characterization of likely outcomes

Figure 2: The equilibrium space when $d = 1$



based on A 's beliefs about the desirability of intervention, which depend on both ϕ and $\bar{c}_B - \underline{c}_B$, and how B 's strategy in the shadow of international law affects those beliefs.

As stated in Proposition 1, the opportunistic defection equilibrium exists when A believes it relatively unlikely that B is expansionist ($\phi \leq \bar{\phi}$). Contrast this with $\phi > \bar{\phi}$, where A is sufficiently sure of B 's expansionism that opportunistic defections no longer occur. Given A 's credible threat to intervene in the event of a violation, only expansionist types violate. When the expansionist type is very expansionist, i.e. when $\underline{c}_B < \rho(\lambda - 1) - a$, there exists a separating equilibrium in which A only intervenes if B violates and B only violates if it is expansionist. On the other hand, when the expansionist type is not wholly undeterrable ($\underline{c}_B \geq \rho(\lambda - 1) - a$), we see a semi-separating partial compliance equilibrium in which the expansionist B sometimes, but not always, honors neutrality rather than face certain intervention. In each case, A 's threat to

intervene against an expansionist is sufficiently strong ex ante that international law can have no meaningful effect; since these equilibria are unique in this part of the parameter space, there is no possibility of opportunistic violations against which to judge the effects of international law.⁷ Therefore, in what follows, I focus strictly on cases where $\phi \leq \bar{\phi}$.

If codified neutrality rights are to have an effect on equilibrium selection, they can do so only when the opportunistic violation equilibrium exists alongside others in which states coordinate their expectations on the belief that violations will be met with intervention. Figure 2 shows that the baseline equilibrium exists alongside one of two alternatives, both of which are associated with beliefs in credible threats of intervention: (a) a separating equilibrium, like the one discussed above, and (b) a full compliance equilibrium in which neither expansionist nor satiable types of B violate neutral rights.

Proposition 2 (Full Compliance). *When $\phi \leq \bar{\phi}$ and $\underline{c}_B > \rho(\lambda - 1) - a$, the following strategies and beliefs constitute a Perfect Bayesian Equilibrium. Both types of B honor. If B honors, A believes $\phi' = \phi$ and does not intervene. If B violates, A believes $\phi'' = 1$ and intervenes.*

Proposition 2 characterizes the full compliance equilibrium, which involves both a perfectly credible threat of intervention in the event of violation and state B honoring neutral rights regardless of its type. Specifically, if B violates, A will believe it to be expansionist, at which point it is happy to intervene and try to prevent B 's from posing a threat in the future. In this quadrant of the parameter space, where expansionism is not too likely and the expansionist type is not too expansionist—that is, not as large a threat as it would be if \underline{c}_B were lower—then total compliance with the laws of neutrality can be achieved simply by coordinating expectations.

⁷See Proposition 3 and its proof for a characterization of the separating equilibrium and the appendix for a Proposition and proof of the partial compliance equilibrium.

In the standard approach to international law, this is the best possible equilibrium, in that the laws of neutrality are never violated; A makes a threat of intervention that ensures compliance, and the threat is never called in along the equilibrium path. However, since A is concerned that expansionist types pose a future threat, it foregoes desirable interventions—those in which it intervenes against an expansionist B . Since expansionism is relatively unlikely and the difference $\bar{c}_B - \underline{c}_B$ between satiable and expansionist types is small, A tolerates some instances of regretted neutrality in return for saving the costs of intervention on the equilibrium path. While A engages in no wasteful interventions on the equilibrium path, it nonetheless foregoes desirable interventions, allowing expansionist states to pose a future threat. In this sense, the threat of intervention supporting full compliance may be *too* strong.

Proposition 3 (Separating). *When $\underline{c}_B < \rho(\lambda - 1) - a$, the following strategies and beliefs constitute a Perfect Bayesian Equilibrium. B violates if expansionist and honors if satiable. If B violates, A believes $\phi' = 1$ and intervenes. If B honors, A believes $\phi' = 0$ and does not intervene.*

Proposition 3 characterizes the equilibrium that exists simultaneously with opportunistic violation when $\underline{c}_B < \rho(\lambda - 1) - a$, or when the difference between expansionist and satiable types is relatively large—alternatively, when the expansionist type poses a more extreme threat. In this part of the parameter space, A also wields a credible threat of intervening upon a violation of neutrality; however, in this separating equilibrium, the satiable type of B remains content to honor the laws of neutrality and avoid intervention, while only the expansionist type violates. A 's threat to intervene after violations is perfectly credible, just as it is in the full compliance equilibrium, the difference here being that the expansionist type is so bent on reaping the military gains of violating neutrality that it cannot be deterred from doing so.

While it cannot be said that international law “works” here in the sense of deterring as many violations as it does in the full compliance equilibrium, it does have an important, though largely unexplored, effect: it forces a separation of belligerent types that allows *A* to identify *desirable* interventions. In fact, *A* is sure in this case to avoid both regretted interventions *and* regretted neutrality. If *A* valued only *B*’s compliance with laws of neutrality—that is, if intervention were only useful as a punishment with the aim of deterrence—then we would rightly interpret the law of neutrality as doing a rather poor job here. Yet when compliance is less important than choosing intervention wisely, punishment is spurious to the knowledge *A* gains from the very fact that *B* cannot be deterred: *B* is expansionist and, as a result, worth intervening against in order to cut off a future threat at the knees.

Since *A* would intervene if it knew *B* to be expansionist even if it honored neutrality rights, punishment and intervention are not about reciprocity and the avoidance of temptation, and this presents a rather different view of the effects and effectiveness of international law than in more common accounts. In the full compliance equilibrium, threats of intervention work in the sense that expansionist states that are not *too* aggressive opt to honor neutrality rights; however, this does come at a cost to state *A*, who forgoes intervention when it observes rights being honored, despite the fact that it might be allowing a future threat to grow under its nose. When the expansionist *B* is so aggressive as to be undeterrable, we observe the separating equilibrium, where codified laws “work” precisely because they are violated: only if the expansionist *B* violates neutral rights can *A* tailor its strategy most effectively by doing precisely what it most wishes to do against an expansionist belligerent by joining the war against it.

The Costs of Intervention

The second possible effect of international law involves the coordination of expectations not between states in the game but between A and other actors in the system. While the formers affect the equilibrium likely to be played, the latter has exogenous effects on the payoffs to particular actions, in this case A 's costs for intervention in the event that B violates international law. If states or domestic constituencies are more likely to support an intervention—or to refrain from opposing it—after a clear violation of codified neutrality rights, then the costs of intervention may be lower if B violates than if it honors. In this section, I capture this possibility by relaxing that assumption that $d = 1$, allowing instead for the costs of intervention to see a modest drop following B 's choice to violate, such that $\underline{d} < d < 1$.⁸ Figure 3 shows that a reduction in the costs of intervention changes outcomes substantially, eliminating some instances of multiple equilibria and rendering separating and full compliance equilibria unique in particular parts of the parameter space.

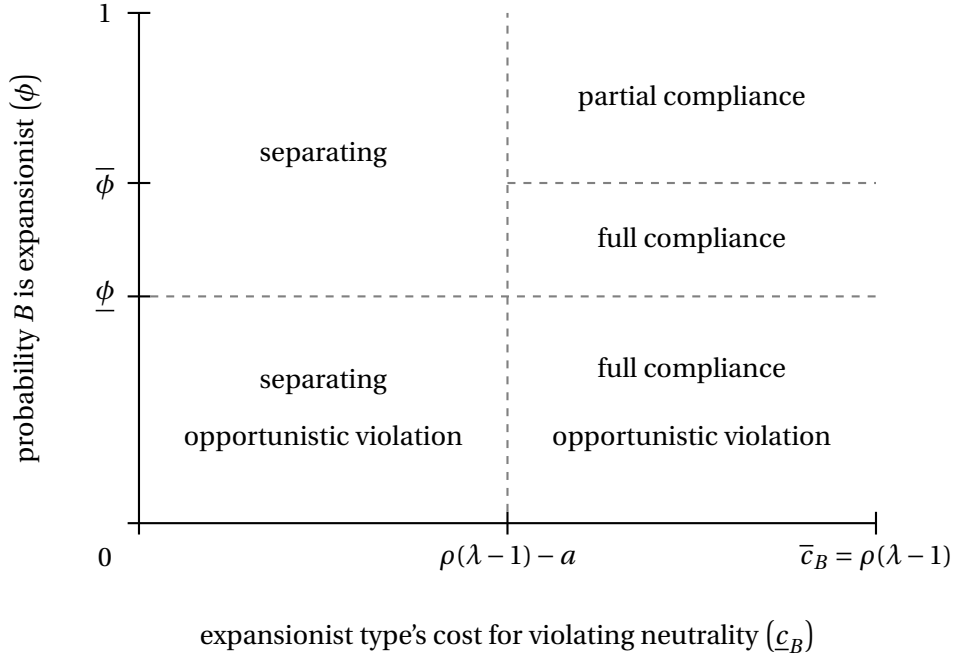
As before, let the area below $\bar{\phi}$ denote a world in which the opportunistic violation equilibrium exists without codified neutrality rights, and the area above continues to involve threats of intervention that are credible absent codified law. While assuming $\underline{d} < d < 1$ introduces no new equilibria, it does change the parameter values for which particular equilibria exist, by introducing a new threshold ($\underline{\phi}$) below which opportunistic violation is a plausible outcome.⁹

Comparing Figure 3 to Figure 2, the difference is stark. Thanks to the decreased costs of intervention after a violation, A can now wield sufficiently credible threats to intervene for an even

⁸Specifically, I assume that $\underline{d} < d < 1$, where $\underline{d} = u_A(\bar{c}_B) / u_A(\underline{c}_B)$, which ensures that the costs of intervention cannot be lowered so much that A is willing to intervene against a satiable type that violates but not an expansionist type that does not violate.

⁹As shown in the proof of Proposition 1, $\underline{\phi} = \bar{\phi}$ when $d = 1$, and it falls below $\bar{\phi}$ when $d < 1$.

Figure 3: The equilibrium space when $\underline{d} < d < 1$



lower probability that B is expansionist; as a result, the opportunistic violations equilibrium no longer exists when $\underline{\phi} < \phi \leq \bar{\phi}$, replaced with either the full compliance or the separating equilibrium, depending on where \underline{c}_B falls relative to $\rho(\lambda - 1) - a$.

This lowered cost of intervention facilitated may reduce the relevance of the equilibrium selection effect, but, notably, the behavioral predictions remain the same: codified neutrality rights can move states from a world of opportunistic violations to either separating behavior or full compliance. Further, across both figures, rates of regretted neutrally remain highest for low values of ϕ , where expansionism is not believed overly likely, and high values of \underline{c}_B , where the expansionist type is, in a relative sense, not too expansionist, where the full compliance equilibrium exists. However, equilibria that exist uniquely, as they do when $\underline{\phi} < \phi \leq \bar{\phi}$, are more robust to shocks in the strategic-informational environment, because there is no problem

of multiple equilibria. Therefore, while international law can affect both compliance and war expansion behavior merely by coordinating expectations on particular equilibria, it can do so in a more robust way when it also enables states to intervene against violators at lower cost.

Summary and Discussion

In this section, I present an informal summary of the theory's key insights into (a) the conditions under which international law affects state behavior, (b) the mechanisms behind these effects, and (c) implications for rates of compliance and the expansion of war. Despite varying substantive contexts, the central insight behind each implication stems from the same strategic tension: the most prominent punishment for the violation of neutrality—joining the war against an offending state—is sometimes desirable, yet third parties are often uncertain over whether a belligerent is expansionist, and therefore worth fighting today, or satiable, and therefore not worth fighting. To the extent that international laws of neutrality affect state behavior, it is because they shape expectations over the likely response to violations and because they help separate types according to the scope of belligerents' foreign policy ambitions, and this latter dynamic produces some of the theory's most novel insights.

First, as noted at the beginning of the analysis, *codified neutrality rights have their greatest impact on state behavior when expansionism is believed to be rare*, because coordination is easiest when expansionism is believed very likely. When *A* already has a strong prior belief that *B* is expansionist, then it can make a sufficiently credible threat to join the war that the satiable *B* is deterred from violating opportunistically, and there is very little scope for international law to alter state behavior. On the other hand, when *A* begins the game less certain that it faces an ex-

pansionist *B*, then there is room for international law to alter behavior, because an equilibrium in which satiable types opportunistically violate thanks to weak threats of intervention exists. Whether law works by simply selecting one equilibrium from among many or replacing the opportunistic violations equilibrium with a full compliance or separating equilibrium, it does so only when states cannot already coordinate expectations on credible threats of punishment.

This raises an important point about how to judge whether international law shapes state behavior, especially when its primary effect is to coordinate expectations amongst multiple equilibria. If states can already coordinate on particular responses to violations, as they can when the belligerent in question is already believed to be expansionist, then there is little scope for international law to shape behavior. Setting aside for a moment the specific behavioral predictions associated with them, looking for the effects of neutrality rights requires that we identify cases where expansionism is not believed likely *ex ante*—a difficult task to be sure. However, sampling on cases involving both likely and unlikely expansionism runs the risk of over-attributing both the deterrent and coordinating effects on neutrality rights.

Second, in addition to identifying when international law is likely to be effective, the theory also has implications for what those effects are likely to be. While increasing the credibility of threats to intervene can encourage compliance by satiable states, there is also a significant informational effect, particularly when states can coordinate on the separating equilibrium, because *violations of neutrality rights facilitate desirable interventions, because only expansionist states are likely to violate the law*. As a result, violations of neutrality rights, like Germany's in 1914, may represent a failure of deterrence but *not* of international law itself. In fact, full compliance entails instances of regretted neutrality, where *A* is sufficiently doubtful of *B*'s expansionism and its threat to intervene following violation is so credible that expansionist states

mask their type, pooling with satiable states and avoiding interventions that, all else equal, *A* would wish to engage in. A separating equilibrium, on the other hand, occurs when the threat of intervention cannot deter an expansionist *B* from violating but can deter a satiable type; here, the law effectively solves the information problem with which *A* begins the game, allowing it to avoid wasteful interventions and engage in those that are most desirable.

Because the laws of neutrality can be “effective” in two different ways, deterring some violations but encouraging desirable interventions against belligerents it does not deter, then compliance is a problematic standard by which to judge its role in shaping state behavior. Rates of compliance can be a useful explanatory variable when cooperation is sustained by threats of reciprocal punishment (see [Morrow 2002](#), p. S44), but when punishment is spurious to desirable interventions, a lack of compliance does not indicate that international law has no effect on state behavior. To assess this model empirically, it is necessary, first, to identify cases where codified neutrality rights can have an effect, such as a low *ex ante* probability that a belligerent is expansionist. Second, the analyst must identify differences between the opportunistic violations equilibrium and its alternatives. In this case, the most prominent difference is a reduced rate of unnecessary interventions, because both violations and regretted neutrality still occur in equilibrium. Identifying regretted interventions is difficult, particularly because war can create regret over its political and material costs regardless of its underlying aims ([Arena n.d.](#), [Fearon 1995](#)), but other outcomes such as rates of compliance and violation may not be sufficient to distinguish laws with and without effect on the behavior of states.

The model also suggests that, despite their inherent advantages in joining the war, *more powerful third-party states should be more prone to regretted neutrality than weaker states*. At first glance, this runs counter to intuition: powerful states should be better able to recover the

costs of war, and as such should have no excuse for failing to intervene. However, this misses the point that the very credibility of a powerful state's threat to intervene, and the extent to which it can affect B 's military fortunes, alters an expansionist B 's incentives to violate in the first place. We can see this by examining Figures 2 and 3, where the threshold that divides full compliance from separating equilibria is $\underline{c}_B = \rho(\lambda - 1) - a$, which falls as A 's military power (a) increases. As a result, full compliance exists for values of \underline{c}_B where separating would otherwise occur if a were lower. Therefore, it is the very credibility of A 's threat to intervene that forces the expansionist type to pool with the satiable type, leaving A 's information problem unsolved and forcing it, thanks its prior doubts that B is expansionist (low ϕ), to remain neutral and run the risk that it will allow an expansionist state to go unchecked in today's war. Weaker third parties should thus appear better able than the strong to tailor their decisions to join wars to their desire to balance against future threats. This is, in fact, consistent with the violation of Belgian neutrality in the First World War; Germany, convinced that any British military contribution to the war would be marginal, was more than willing to countenance its intervention in order to realize its plans of sweeping into France through Belgium (Hastings 2013, p. 101).

Finally, the model has several implications for our understanding of the expansion of interstate war, because *deterrence of violations is achieved by the threat that neutrality rights will facilitate the expansion of the war*. Notably, since these interventions occur on the equilibrium path when the separating equilibrium exists, we should not expect this relationship to be subject to heavy strategic censoring (cf. Werner 2000). This leads to the simple prediction that *violations of neutrality should increase the probability that wars expand to include new belligerents*, a hypothesis thus far missing from most empirical studies of conflict expansion (see, *inter alia*, Altfeld and Bueno de Mesquita 1979, Gartner and Siverson 1996, Siverson and Starr 1990,

Wolford 2014). Further, the states most likely to respond to violations of neutrality by joining wars are relatively weak, because their initial threats to intervene are not so credible as to induce the expansionist *B* into the full compliance equilibrium.

Conclusion

International law can shape state choices, even in wartime, by coordinating expectations on acceptable standards of behavior and the appropriate responses to violations of those standards. In the case of neutrality rights, a body of law rooted in traditional practice and codified by The Hague Convention of 1907 stipulates that belligerents must respect the sovereignty of neutral nations, lest they face intervention in their war by either the neutral states they victimize or third parties acting on their behalf. Nonetheless, the threat of intervention only imperfectly deters violations of neutrality; it seemed to work against Japan in 1904, but it manifestly failed to stop Germany in 1914. Rather than describe these varying responses as a symptom of the weakness or irrelevance of international law, I have argued instead that neutrality rights may work best when they are occasionally violated. The key premise of this argument is that, when it comes to neutrality rights, the ostensible means of punishing violators—i.e., intervening against them in war—may be inherently *desirable* if the violator is truly expansionist.

When intervening against an expansionist belligerent may hasten its defeat and eliminate a possible future threat, the typical assumption that punishment is strictly costly, useful only to the extent that it induces compliance, may not be appropriate. If intervention against expansionist states is attractive, it is nonetheless wasteful against satiable states that, even if they violate neutrality, have less aggressive foreign policy ambitions. Codified neutrality rights, I have

argued, can render credible threats to intervene against states that violate neutrality rights, but there is a sense in which they can be *too* effective; if all states honor neutrality rights, then third parties may find themselves remaining neutral, opting not to intervene against states that turn out, after the fact, to have been worth intervening against. Therefore, neutrality rights can, under some conditions, force states to trade the occasional wasteful intervention—which they might embark on if satiable states opportunistically violate—for regretted neutrality.

However, when the threat of intervention is strong, but not *too* strong, codified neutrality rights can facilitate a separation of types in which only expansionist belligerents violate the law. This can solve a third party's informational problem and help it identify desirable interventions. Thus, in the case of neutrality rights, the apparent failure of deterrence and subsequent punishment may be spurious to a different, though no less important, aspect of international law: helping third parties choose between worthwhile and worthless interventions in ongoing wars. In the model presented here, international law would seem to work best when it deters only some violations of neutrality, such that only expansionist types violate the law; in the event of a violation in this kind of equilibrium, a third party is best able to tailor its intervention decisions to avoid both regretted neutrality and unnecessary interventions.

These insights derive from integrating a common law-and-compliance framework built on coordination in the absence of central authority (Morrow 2002) with an equally common framework for understanding state decisions to join ongoing wars rooted in classical approaches to the balance of power (Altfeld and Bueno de Mesquita 1979, Powell 1999, Wagner 2007). Put differently, answering questions about the effects of the laws of war requires strong theories about both international law and international war.¹⁰ This synthesis shows that, while “punishment”

¹⁰For similar syntheses of theories of law and war, see Huth, Croco and Appel (2011, 2012).

for violations of neutrality may be spurious to a third party's own interest in eliminating future threats, it can be so precisely because international laws of neutrality coordinate of expectations on both the rights and obligations of neutrals and acceptable responses to breaches of those rights. Indeed, the laws of neutrality work best not only when they make credible the threat of war expansion but also when desirable interventions are allowed to occur. Codified neutrality rights, then, can restrain both states that would otherwise violate them opportunistically *and* third parties that might otherwise engage in wasted interventions, even when no states have an inherent interest in upholding neutrality rights for their own sake. In other words, laws of neutrality can shape state behavior even when there is no centralized enforcement and when no one *cares* about the content of the law itself.

Appendix

Proof of Proposition 1. Begin with the satiable type \bar{c}_B , rendered indifferent by A 's strategy of mixing joining and neutrality after violation, such that $EU_B(h|\bar{c}_B) = EU_B(v|\bar{c}_B)$. This requires that y satisfy

$$\rho + b = y(\rho\lambda + b - a) + (1 - y)(\rho\lambda + b) - \bar{c}_B,$$

yielding

$$y^* = \frac{\rho(\lambda - 1) - \bar{c}_B}{a}$$

as the probability with which A joins following violation. To ensure that $y^* \in [0, 1]$, it must be the case that

$$\rho(\lambda - 1) - a \leq \bar{c}_B \leq \rho(\lambda - 1).$$

Next, we verify that the expansionist type is happy to violate, or $EU_B(v|\underline{c}_B) \geq EU_B(h|\underline{c}_B)$. Given A 's strategy, this is satisfied when

$$y^*(\rho\lambda + b - a) + (1 - y^*)(\rho\lambda + b) - \underline{c}_B \geq \rho + b.$$

This is sure to be true as long as $\bar{c}_B > \underline{c}_B$, which is true by assumption.

Next, we verify that A 's beliefs are consistent and strategies optimal given B 's strategy. If B

honors, A believes $\phi' = 0$, which implies that $EU_A(n|h) \geq EU_A(j|h)$ can be represented by

$$(\rho + b) u_A(\bar{c}_B) \geq (\rho + b - a) u_A(\bar{c}_B) - c_A,$$

which is true when

$$c_A \geq -a u_A(\bar{c}_B). \quad (1)$$

If B violates, A is rendered indifferent between joining and remaining neutral by the probability with which the satiable type violates (g^*), such that $EU_A(j) = EU_A(n)$. By Bayes' Rule, A 's posterior belief that B is expansionist given violation is

$$\phi' = \frac{\phi}{\phi + (1 - \phi)g} \equiv \chi.$$

Therefore, the indifference condition is

$$\chi(\rho\lambda + b - a) u_A(\underline{c}_B) + (1 - \chi)(\rho\lambda - a) u_A(\bar{c}_B) - d c_A = \chi(\rho\lambda + b) u_A(\underline{c}_B) + (1 - \chi)(\rho\lambda) u_A(\bar{c}_B),$$

which yields the satiable type's probability of violation,

$$g^* = \frac{\phi(d c_A + a u_A(\underline{c}_B))}{-(1 - \phi)(d c_A + a u_A(\bar{c}_B))}.$$

Ensuring that $g^* \in [0, 1]$ requires, first, that

$$-\frac{a u_A(\bar{c}_B)}{d} < c_A < -\frac{a u_A(\underline{c}_B)}{d}$$

which binds over Inequality (1) since $d \leq 1$, and, second, that

$$\phi \leq \frac{d c_A + a u_A(\bar{c}_B)}{a u_A(\bar{c}_B) - a u_A(\underline{c}_B)}. \quad (2)$$

Finally, when $d = 1$, as it is in Figure 2, let $\bar{\phi}$ denote the constraint in Inequality (2), and when $d < 1$, as it is in Figure 3, let $\underline{\phi}$ denote the constraint in Inequality (2). \square

Proof of Proposition 2. Begin with A 's strategy, which is to remain neutral if B honors and intervene if B violates. In the event that B honors, A remains neutral when $EU_A(n) \geq EU_A(j)$, or when

$$\begin{aligned} & \phi(\rho + b) u_A(\underline{c}_B) + (1 - \phi)(\rho + b) u_A(\bar{c}_B) \geq \\ & \phi(\rho + b - a) u_A(\underline{c}_B) + (1 - \phi)(\rho + b - a) u_A(\bar{c}_B) - c_A, \end{aligned}$$

where its posterior belief is simply its prior, $\phi' = \phi$. This inequality can be satisfied, first, when

$c_A \geq -au_A(\underline{c}_B)$ or, second, when $-au_A(\bar{c}_B) < c_A \leq -au_A(\underline{c}_B)$ and

$$\phi \leq \frac{c_A + au_A(\bar{c}_B)}{au_A(\bar{c}_B) - au_A(\underline{c}_B)} \equiv \bar{\phi}.$$

Next, A intervenes if B violates. Letting ϕ'' denote its posterior (out of equilibrium) belief that B is expansionist, it intervenes when $EU_A(j) \geq EU_A(n)$, or

$$\begin{aligned} \phi''(\rho + b - a)u_A(\underline{c}_B) + (1 - \phi'')(\rho + b - a)u_A(\bar{c}_B) - dc_A > \\ \phi''(\rho + b)u_A(\underline{c}_B) + (1 - \phi'')(\rho + b)u_A(\bar{c}_B). \end{aligned}$$

This inequality is satisfied for any out of equilibrium beliefs ϕ'' when $c_A < -au_A(\bar{c}_B)$, and for

$$\phi'' > \frac{dc_A + au_A(\bar{c}_B)}{au_A(\bar{c}_B) - au_A(\underline{c}_B)}$$

when

$$-\frac{au_A(\bar{c}_B)}{d} < c_A < -\frac{au_A(\underline{c}_B)}{d}.$$

Now, consider the satiable type \bar{c}_B , who honors when $EU_B(h|\bar{c}_B) \geq EU_B(v|\bar{c}_B)$, or $\rho + b \geq \rho\lambda + b - a - \bar{c}_B$. This inequality is satisfied when $\bar{c}_B \geq \rho(\lambda - 1) - a$. The expansionist type honors when $EU_B(h|\underline{c}_B) \geq EU_B(v|\underline{c}_B)$, or $\rho + b \geq \rho\lambda + b - a - \underline{c}_B$. This inequality is satisfied when $\underline{c}_B \geq \rho(\lambda - 1) - a$. Since $\bar{c}_A > \underline{c}_A$, the binding constraint is $\underline{c}_B \geq \rho(\lambda - 1) - a$. \square

Proof of Proposition 3. Begin with B 's incentive-compatibility constraints. The satiable type honors, such that $EU_B(h|\bar{c}_B) \geq EU_B(h|\underline{c}_B)$, or $\rho + b \geq \rho\lambda + b - a - \bar{c}_B$, and the expansionist type violates, such that $EU_B(v|\underline{c}_B) \geq EU_B(v|\bar{c}_B)$, or $\rho\lambda + b - a - \underline{c}_B \geq \rho + b$. Both constraints are satisfied when

$$\underline{c}_B \leq \rho(\lambda - 1) - a \leq \bar{c}_B.$$

Since player-types of B perfectly separate, A 's posterior beliefs are $\phi' = 1$ if B violates and $\phi' = 0$ if B honors. As such, it intervenes following a violation when $EU_A(j) > EU_A(n)$, or when

$$(\rho\lambda + b - a)u_A(\underline{c}_B) - dc_A > (\rho\lambda + b)u_A(\underline{c}_B),$$

which is satisfied when

$$c_A < -\frac{au_A(\underline{c}_B)}{d}.$$

A remains neutral if B honors when $EU_A(n) \geq EU_A(j)$, or when

$$(\rho + b)u_A(\bar{c}_B) \geq (\rho + b - a)u_A(\bar{c}_B) - c_A,$$

which is satisfied when $c_A \geq -au_A(\bar{c}_B)$. Combining these constraints,

$$-au_A(\bar{c}_B) \leq c_A < -\frac{au_A(\underline{c}_B)}{d}$$

ensues that A 's strategy is a best response and consistent with its beliefs. \square

Proposition 4 (Partial Compliance). *When $\underline{c}_B > \rho(\lambda - 1) - a$ and $\phi > \max\{\underline{\phi}, \bar{\phi}\}$, the following strategies and beliefs constitute a Perfect Bayesian Equilibrium. If satiable, B honors. If expansionist, B honors with probability h^* and violates with probability $1 - h^*$. If B honors, A believes $\phi' = \sigma$ and intervenes with probability r^* , remaining neutral with probability $1 - r^*$. If B violates, A believes $\phi' = 1$ and intervenes.*

Proof of Proposition 4. Begin to the expansionist type \underline{c}_B , rendered indifferent by A 's strategy of mixing joining and neutrality following honoring, such that $EU_B(v|\underline{c}_B) = EU_B(h|\underline{c}_B)$. This requires that r satisfy

$$\rho\lambda + b - a - \underline{c}_B = r(\rho + b - a) + (1 - r)(\rho + b)$$

yielding

$$r^* = \frac{\rho(\lambda - 1) + a + \underline{c}_B}{a}$$

as the probability with which A intervenes if B honors. To ensure that $r^* \in [0, 1]$, it must be the case that

$$\rho(\lambda - 1) - a \leq \underline{c}_B \leq \rho(\lambda - 1).$$

Next, we verify that the satiable type is happy to honor, or $EU_B(h|\bar{c}_B) \geq EU_B(v|\bar{c}_B)$. Given A 's strategy, this constraint is satisfied when

$$r^*(\rho + b - a) + (1 - r^*)(\rho + b) \geq \rho\lambda + b - a - \bar{c}_B,$$

or when

$$c_A \geq -\frac{au_A(\underline{c}_B)}{d}.$$

Next, we verify that A 's beliefs are consistent and strategies optimal given B 's strategy. if B violates, A believes $\phi' = 1$, which implies that $EU_A(j|v) > EU_A(n|v)$ can be represented by

$$(\rho\lambda + b - a)u_A(\underline{c}_B) - dc_A > (\rho\lambda + b)u_A(\underline{c}_B),$$

which is true when

$$c_A < -\frac{au_A(\underline{c}_B)}{d}. \quad (3)$$

If B honors, A is rendered indifferent over joining and remaining neutral by the probability with

which the expansionist type honors (h^*), such that $EU_A(j|h) = EU_A(n|h)$. By Bayes' Rule, A 's posterior belief that B is expansionist given that it honored is

$$\phi' = \frac{\phi h}{\phi h + (1 - \phi)} \equiv \sigma.$$

Therefore, the indifference condition is

$$\sigma(\rho + b - a)u_A(\underline{c}_B) + (1 - \sigma)(\rho + b - a)u_A(\bar{c}_B) - c_A = \sigma(\rho + b)u_A(\underline{c}_B) + (1 - \sigma)(\rho + b)u_A(\bar{c}_B),$$

which yields the expansionist type's probability of honoring,

$$h^* = \frac{-(1 - \phi)(c_A + au_A(\bar{c}_B))}{\phi(c_A + au_A(\underline{c}_B))}.$$

Ensuring that $h^* \in [0, 1]$ requires, first, that

$$\phi \geq \frac{c_A + au_A(\bar{c}_B)}{au_A(\bar{c}_B) - au_A(\underline{c}_B)} \equiv \bar{\phi}$$

and, second,

$$-au_A(\bar{c}_B) \leq c_A \leq -au_A(\underline{c}_B),$$

which binds over Inequality (3) since $d \leq 1$. □

References

- Altfeld, Michael D. and Bruce Bueno de Mesquita. 1979. "Choosing Sides in Wars." *International Studies Quarterly* 23(1):87–112.
- Arena, Philip. n.d. "Crisis Bargaining, Domestic Opposition, and Tragic Wars." *Journal of Theoretical Politics* forthcoming.
- Bull, Hedley. 1977. *The Anarchical Society: A Study of Order in World Politics*. New York: Columbia University Press.
- 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):669–689.
- Chapman, Terrence L. 2011. *Securing Approval: Domestic Politics and Multilateral Authorization for War*. Chicago, IL: University of Chicago Press.
- Chapman, Terrence L. and Stephen Chaudoin. 2013. "Ratification Patterns and the International Criminal Court." *International Studies Quarterly* 57(2):400–409.

- Chayes, Abram and Antonia Handler Chayes. 1993. "On compliance." *International Organization* 47(2):175–205.
- Che, Yeon-Koo and Ian Gale. 2000. "Difference-Form Contests and the Robustness of All-Pay Auctions." *Games and Economic Behavior* 30(1):22–43.
- Connaughton, Richard. 2004. *Rising Sun and Tumbling Bear: Russia's War with Japan*. London: Cassell Military Paperbacks.
- Conrad, Courtenay R. and Emily Hencken Ritter. 2013. "Treaties, Tenure, and Torture: The Conflicting Domestic Effects of International Law." *Journal of Politics* 75(2):397–409.
- Dai, Xinyuan. 2005. "Why Comply? The Domestic Constituency Mechanism." *International Organization* 59(2):363–398.
- 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.
- Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3):379–414.
- Gartner, Scott Sigmund and Randolph M. Siverson. 1996. "War Expansion and War Outcome." *Journal of Conflict Resolution* 40(1):4–15.
- Hague Convention (V) Respecting the Rights and Duties of Neutral Powers and Persons in Case of War on Land*. 1907. Available at: http://avalon.law.yale.edu/20th_century/hague05.asp [accessed 28 January 2014].
- Hague Convention (XIII) Concerning the Rights and Duties of Neutral Powers in Naval War*. 1907. Available at: http://avalon.law.yale.edu/20th_century/hague13.asp [accessed 28 January 2014].
- Hastings, Max. 2013. *Catastrophe 1914: Europe Goes to War*. New York: Alfred A. Knopf.
- Hershey, Amos S. 1904. "Some Questions of International Law Arising from the Russo-Japanese War." *The Green Bag* 16:306–311.
- Hurd, Ian. 2005. "The Strategic Use of Liberal Internationalism: Libya and the UN Sanctions, 1992-2003." *International Organization* 59(3):495–526.
- Huth, Paul K., Sarah E. Croco and Benjamin J. Appel. 2011. "Does International Law Promote the Peaceful Settlement of International Disputes? Evidence from the Study of Territorial Conflicts since 1945." *American Political Science Review* 105(2):415–436.
- Huth, Paul K., Sarah E. Croco and Benjamin J. Appel. 2012. "Law and the Use of Force in World Politics: The Varied Effects of Law on the Exercise of Military Power in Territorial Disputes." *International Studies Quarterly* 56(1):17–31.

- Jia, Hao, Stergios Skaperdas and Samarth Vaidya. 2013. "Contest functions: Theoretical foundations and issues in estimation." *International Journal of Industrial Organization* 31(3):211–222.
- Kim, Moonhawk and Scott Wolford. 2014. "Choosing Anarchy: Institutional Alternatives and the Global Order." *International Theory* 6(1).
- Morrow, James D. 2001. "The Institutional Features of the Prisoners of War Treaties." *International Organization* 55(4):971–991.
- Morrow, James D. 2002. "The Laws of War, Common Conjectures, and Legal Systems in International Politics." *Journal of Legal Studies* 31:S41–S60.
- Morrow, James D. 2007. "When Do States Follow the Laws of War?" *American Political Science Review* 101(3):559–572.
- Posner, Eric A. 2003. "A Theory of the Laws of War." *University of Chicago Law Review* 70:297–317.
- Powell, Robert. 1999. *In the Shadow of Power*. Princeton: Princeton University Press.
- Reinhardt, Eric. 2001. "Adjudication without Enforcement in GATT Disputes." *Journal of Conflict Resolution* 45(2):174–195.
- Simmons, Beth A. and Allison Danner. 2010. "Credible Commitments and the International Criminal Court." *International Organization* 64(2):225–256.
- Siverson, Randolph M. and Harvey Starr. 1990. "Opportunity, Willingness, and the Diffusion of War." *American Political Science Review* 84(1):47–67.
- Valentino, Benjamin, Paul K. Huth and Sarah E. Croco. 2006. "Covenants Without the Sword: International Law and the Protection of Civilians in Times of War." *World Politics* 58(3):339–377.
- Voeten, Erik. 2005. "The Political Origins of the UN Security Council's Ability to Legitimize the Use of Force." *International Organization* 59:527–557.
- Wagner, R. Harrison. 2007. *War and the State: The Theory of International Politics*. Ann Arbor: The University of Michigan Press.
- Wallace, Geoffrey P.R. 2012. "Welcome Guests, or Inescapable Victims? The Causes of Prisoner Abuse in War." *Journal of Conflict Resolution* 56(6):955–981.
- Waltz, Kenneth. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.
- Werner, Suzanne. 2000. "Deterring Intervention: The Stakes of War and Third-Party Involvement." *American Journal of Political Science* 44(4):720–732.
- Wolford, Scott. 2014. "Power, Preferences, and Balancing: The Durability of Coalitions and the Expansion of Conflict." *International Studies Quarterly* 58(1).