Chicago Journal of International Law

Volume 16 | Number 1 Article 4

6-1-2015

The Threshold Requirement in Asymmetric Conflicts: A Game Theory Analysis

Alon Cohen

Raphael Bitton

The Threshold Requirement in Asymmetric Conflicts: A Game Theory Analysis

Alon Cohen and Raphael Bitton·

Abstract

In recent years, ad bellum rules have been interpreted more leniently so as to permit forcible responses to terrorism. Yet the threshold level for an "armed attack" that legitimizes full-scale war in response has remained relatively high. This observation is especially puzzling insofar as customary international law reflects the practice of (strong) states, which, as we show, can benefit from credibly committing to lowering their tolerance towards terror attacks. Why is the threshold requirement relatively tolerant? What would be the critical mass of terror beyond which a full-scale war is legitimate? Under what circumstances are states expected to violate the ad bellum threshold requirement?

This Article seeks compelling answers to these questions using game theory. We argue that a low ad bellum threshold requirement (an "intolerant strategy") is a futile measure to fight terrorism, since normally it cannot underlie a renegotiation-proof equilibrium. Namely, were a strong state to seriously consider waging a full-scale war in response to repeated, low-level terror attacks sponsored by a weak state, both parties would be better off renegotiating their way back to the status quo ante. On the other hand, renegotiation may fail if transaction costs are high enough, which ultimately makes the intolerant strategy sustainable and the equilibrium level of terror lower. The overall conclusion is thus counterintuitive: an effective intolerant strategy in the War on Terror is beneficial for victim states, but implausible unless the barriers towards a nonviolent arrangement are sufficiently high.

· Respectively, alonc@post.tau.ac.il, the Law Faculty, Hebrew University, and rephaelb@gmail.com, Haifa University. We thank Eyal Benvenisti, Sharon Hannes, Shai N. Lavie, Zvika Neeman, Ady Pauzner, Ariel Porat, Eric Posner, David Rosenberg, Ariel Rubinstein, Avraham Tabbach, and the participants of seminars in the law faculties of Tel Aviv University and the Interdisciplinary Center in Herzliya for very helpful comments and insights. All errors are ours.

Table of Contents

I. INTRODUCTION

In the pattern of asymmetric international conflicts, the weak side often uses terror and guerilla warfare against its stronger rival. Surprisingly, the victims of terror are relatively tolerant, despite their military superiority. They tend to obey the ad bellum threshold rule that allows a state to wage a full-scale war only in response to a substantial provocation that constitutes an "armed attack." When the provocation falls short of an armed attack, the victim state normally contains the attack or responds rather moderately. This is surprising considering the repetitive nature of terror attacks, which victim states normally can foresee. It is even more surprising in light of the fact that a credible threat of waging a full-scale war, in itself, can theoretically decrease terror to its minimum. Why do these much stronger states usually comply with the threshold requirement, which seems tolerant toward terror? What is the critical mass of terror, beyond which a full-scale war is legitimate? Under which circumstances are states nevertheless expected to violate the ad bellum threshold rule?

Consider a simple paradigm of an unequal, bilateral, international conflict. A strong state suffers repeated terror attacks, which are directly or indirectly orchestrated by a militarily weaker rival state.[1] Due to its superiority, the strong state can theoretically threaten to wage a full-scale war against the weak state in response to any terror act, regardless of scale or effect it induces.[2] We refer to this as an "intolerant trigger strategy." If the threat is credible, which is feasible in dynamic and repeated conflicts, a rational weak state is expected to submit to

---

1 A good example would be the Turkey-Syria-PKK (Kurdistan Workers' Party) triangular relationship. Being militarily inferior, Syria used the PKK organization to exact a toll on Turkey, with whom it had a long territorial conflict. The applicability of our theory, however, extends beyond this sort of example. The main idea is leverage: the existence of military inequality between the conflicting parties that can effectively come into play. Thus, even if the weak state merely harbors the terror organization without actively directing it, our theory still applies insofar as the harboring state has some ability to control the terrorists. Such was the case, for instance, with Pakistan-Taliban relations prior to 9/11 or Syria-Hezbollah relations prior to the Syrian civil war. Our model also applies to a direct conflict between a strong state and a terror organization, provided that the latter is a state-like entity. In this case, again, the power asymmetry can underwrite a credible threat against the terror organization directly. Such is the case with the ongoing struggle between Israel and Hamas, who controls the Gaza strip (after violently taking it over in the 2007 Battle of Gaza). Our theory, however, is not relevant to all kinds of terrorism. Stateless terrorism that is scattered globally and unanswerable to the authority of any viable entity to whom the credible threat can be made does not suit our paradigm. Ostensible modern examples are Al-Qaeda (in contrast with its pre-9/11 status) and the Islamic State/ISIS. Ironically, when such a terror group becomes more and more powerful, inevitably they gain more assets, interests, and state relations that may yet provide the leverage required for our theory to apply.

2 The notion of *full-scale war* by a state refers to the comprehensive and continuous use of force against a target state or organization, which is far broader than a reprisal that is limited in time or scope, with the goal of continuing such use of force until the desired military or political outcomes are met.

it. Terror would thus be eradicated, or at least minimized, which would clearly benefit the strong state. Despite this apparent benefit, however, strong states that suffer from terror, even those as strong as the U.S., normally do not employ intolerant trigger strategies. Full-scale wars are usually waged only if the preceding armed attack surpasses some substantial critical mass (as was the case following 9/11).[3]

One immediate explanation for this puzzle relies on international law. Strong states, arguably, may prefer to act intolerantly toward terrorists and the states that harbor them, but find themselves obligated to comply with international law against their own interests. In other words, states might actually want to adopt an intolerant approach against terror. And yet, the ad bellum threshold rule that prohibits waging war unless the preceding provocation is substantial precludes them from doing so. Indeed, the traditional international lawyers' wisdom holds that nations have various noninstrumental, voluntary motivations for complying with international law.[4]

Proponents of the new international law scholarship, however, such as Jack Goldsmith and Eric Posner, consider this type of explanation insufficient.[5] Their underlying argument is that states are self-interested, and thus their strategies are primarily the result of rational choices. International law reflects such states' practice, and not the other way around. Hence, in order to explain the relatively tolerant ad bellum threshold rule in unequal conflicts, one must rationalize the underlying states' practice.

This logic guides Goldsmith and Posner in their analysis of customary international law using game theory.[6] Among other things, they also address asymmetric setup of conflict using an "entry deterrence" model similar to ours. The strong state warns the weak one not to engage in some activity, which may harm the former's interests (in our context, this activity would be terror, of course). The nature of the threat entails some sanctioning or even a severe military response against the weak side. If the costs of exercising the threat are sufficiently low, the threat is credible and the weak state will submit to it. The strong state too, for its part, may refrain from the same activity but for different

---

3 *See* discussion and sources cited, *infra* note 16.

4 *See, for example*, Abram Chayes et al., *Managing Compliance: A Comparative Perspective*, *in* Engaging Countries: Strengthening Compliance with International Environmental Accords 39–62 (1998); Thomas M. Franck, Fairness in International Law and Institutions (1995); Harold Hongju Koh, *Why Do Nations Obey International Law?*, 106 Yale L.J. 2599 (1997).

5 *See generally* Jack L. Goldsmith & Eric A. Posner, *A Theory of Customary International Law*, 66 U. CHI L. REV. 1113 (1999); JACK L. GOLDSMITH & ERIC A. POSNER, THE LIMITS OF INTERNATIONAL LAW (2005); Jack Goldsmith & Eric A. Posner, *The New International Law Scholarship*, 34 GA. J. INT'L & COMP. L. 463 (2006).

6 *See generally* Goldsmith & Posner, *A Theory of Customary International Law*, *supra* note 5.

reasons, such as high alternative costs. Eventually both states refrain from that activity, a result that may subsequently evolve into a customary rule in the eyes of the observer.

This model, however, cannot fully explain the paradigmatic unequal conflicts on which we focus. The puzzling situation we review here is one where a weak state *does* engage in a harmful activity (terror) against a strong one, in violation of international law; but the latter, in compliance with international law, neither retaliates with its full power nor threatens to do so.

At first glance, one could argue that our paradigm simply implies the impossibility of making credible threats. That is, the costs of punishing the weak state by waging a full-scale war are not trivial for the strong state, its military superiority notwithstanding. Therefore, the argument goes, the threat to punish simply may not be credible. Because the weak state is aware of that, it is not deterred from engaging in terror against the strong one. This explanation, however, is far from satisfactory.

We show that in an infinitely repeated game setup, where the confrontation takes place continuously, strong states *can* build up credible intolerant threats to which their (rational) weak rivals submit.[7] We show that strong states can indeed credibly threaten to launch a full-scale war even in response to a relatively low level of terror. An antiterror war is expected to reduce the terror level as a result of undermining the enemy's capabilities or motivation. The benefit of carrying out that threat is the terror damage the war manages to prevent over time for the strong state. Repeated interactions increase this benefit and thus may render war cost-effective. One would not spend 100 in war in order to save terror damage of 90 this year. But saving 20 every year in the next 10 years could be reason enough to spend 100 today.[8]

---

7 Industrial organization literature has long established effective deterrence strategies of incumbent firms against their rivals, using "predation prices" and other restrictive measures costly to the incumbent firm itself. These models, however, heavily rely on information asymmetry. *See, for example*, Paul Milgrom & John Roberts, *Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis*, 50 ECONOMETRICA 443 (1982); Paul Milgrom & John Roberts, *Predation, Reputation, and Entry Deterrence*, 27 J. ECON. THEORY 280 (1982). Nevertheless, we argue that if indeed similar equilibrium exists in international conflicts—whereby strong states employ costly strategies to deter weaker states from provoking them—it could be for reasons other than information asymmetry.

8 Our argument, in fact, goes much further than that. We claim that even if the terror level is very low such that the benefits from eradicating it in all future periods fall *below* the current costs of war for the strong state, the threat of war can still be credible. This is because the benefits of war for the strong state should not be measured as the gap between the terror level it suffers before the full-scale war it wages and the level following it. It should be measured as the gap between the terror level in the *absence* of war and following one. Because the terror level plausibly increases when the weak state provokes the strong one and goes unpunished, war actually may prevent more terror than what the strong state currently suffers.

Consequently, a subgame perfect equilibrium, which describes credible strategies of both states in a dynamic and repeated conflict, *can* be based on a strong state's intolerant strategy and ultimately minimize terror.[9] This, of course, reiterates the need to explain why strong states normally do not employ such strategies. We maintain that the reason is the ability of the conflicting (rational) parties to communicate, directly or indirectly. If the attacks sponsored by the weak state should trigger a full-scale war by the strong state according to its subgame perfect equilibrium strategy, the parties would presumably renegotiate their way back to the status quo ante. By definition, neither party is better off deviating alone from its equilibrium strategy; yet both parties are better off if they *jointly* deviate. The strong state avoids severely damaging the weak one by waging a full-scale war; the latter, in return, refrains from escalating terror as a response to its rival containment strategy.

Hence, a subgame perfect equilibrium that is based upon an intolerant trigger strategy cannot survive renegotiation perfection.[10] On the other hand, a subgame perfect equilibrium that survives renegotiation perfection is more plausible. We show that such equilibrium is characterized by relatively high levels of terror in the victim state, which correspond to its tolerant strategy. We argue that this is the main reason for the implausibility of strong states' intolerant trigger strategies. Hence, the ad bellum threshold rule, which prohibits waging a full-scale war in response to mild provocations, may simply reflect that notion.[11]

Next, we turn to discuss the circumstances in which the strong party may breach the ad bellum threshold rule and act intolerantly. Because renegotiation in fact is a barrier to eradicating terror, the strong state may benefit from its failure. Counterintuitively, "transaction costs" may do the trick. Recall that foreseeable renegotiation is what derogates the credibility of the intolerant strategy in the first place. If the parties can predict, however, that such negotiation will fail, intolerance becomes effective. For instance, the leader of a strong state could be anxious to go to war for idiosyncratic political reasons.

---

9 For a more detailed description of the subgame perfect equilibrium notion, see discussion, *infra* note 41.

10 *See* Joseph Farrell & Eric Maskin, *Renegotiation in Repeated Games*, 1 GAMES AND ECON. BEHAVIOR 327 (1989). Renegotiation perfection is a refinement concept, usually used to narrow down the set of subgame perfect equilibria, in order to have a more precise prediction for the outcome of games. Basically, it means that if at some point *both* players agree to change course, the original subgame perfect equilibrium is not renegotiation-proof, and thus can be eliminated as implausible.

11 We also deal with several critiques of this hypothesis and make a more refined prediction regarding states' practice in that regard, using a more complicated perfection criterion. For example, one may argue that the renegotiation that supposedly restores the status quo ante is not a reliable prediction. What would stop the weak state, the argument goes, from repeating the routine of provocation-then-renegotiation in the future? We show that even this logic cannot help sustain intolerant strategies (and, in turn, the lower levels of terror they generate).

Alternatively, potential mediators in the international community may suffer from a collective action problem. The conflicting parties, in this case, foresee the futility of renegotiation. Hence, a strong state's intolerant strategy can support a renegotiation-proof equilibrium in which the parties will not agree to change course or strategy during the period of the "game." Because an intolerant, renegotiation-proof strategy benefits the strong state, we predict that in these circumstances it may ignore the ad bellum threshold rule and employ an intolerant trigger strategy towards its weaker rivals. As abnormal as this may sound, transaction costs in the international sphere may lead to better results in the War on Terror. An effective international mechanism for negotiating de-escalation arrangements will actually increase the level of state-sponsored terror. We support this insight with some evidence from international conflicts worldwide.[12]

The work proceeds as follows. Section II sketches the practice and law of unequal conflicts, describing our doctrinal premise of a tolerant threshold rule. Section III demonstrates our argument using a simple numeric example. Section IV engages an unequal, bilateral confrontation model, which shows one plausible set of subgame perfect equilibria based on tolerant and intolerant trigger strategies. Section V refines the set of subgame perfect equilibria based on renegotiation perfection, grounding our argument in the origins of the threshold requirement as well as its possible violation. Section VI then concludes by pointing out other areas of international law, such as treaty practice, to which this Article's insights can extend.

II. THE PRACTICE & LAW OF UNEQUAL CONFLICTS

A. Terrorism in Unequal Conflicts

Many international conflicts are unequal in terms of the military capacity of the parties. Any full-scale war in these circumstances, albeit costly also to the strong state, would clearly end up causing enormous damage to the weak one. Hence, the latter often employs an indirect conflict strategy by endorsing

---

12 This Article focuses on patterns of terror sponsoring. However, the analysis it offers can readily serve as the groundwork for understanding many other unequal confrontations in international relations. It can explain, for example, what made the Syrian regime violate the intolerant "red line" threat set by the U.S. concerning the use of chemical weapons. And it can assist in explaining the renegotiation and its ensuing results. For more on first signals of an intolerant trigger strategy, see Mark Landler, *Obama Threatens Force Against Syria*, N.Y. TIMES, (Aug. 20, 2012), http://www.nytimes.com/2012/08/21/world/middleeast/obama-threatens-force-against-syria.html. For an account of the events from the Administration's perspective—from making the threat to it being nonrenegotiation-proof—see *Chemical Weapons Attack in Syria, News and Updates*, WHITE HOUSE, *available at* http://www.whitehouse.gov/issues/foreign-policy/syria (last visited Feb. 25, 2015).

terrorism and guerilla warfare against its stronger rival, hoping to maintain the low intensity of the conflict. In many cases, this indirect strategy is also based on the weaker state's belief that in the long run it will accomplish its territorial or ideological goals.

A clear example of such a strategic use of terror is the situation in Syria. Patrick Seale maintains that Syria's former president, Hafez al-Assad, preferred to manage a conflict over disputed territories with Israel and Turkey via terrorist proxies rather than wage a full-scale war with these much stronger states.[13] Libya adopted a similar strategy against the U.S. during the 1970s and the 1980s.[14] Pakistan also has orchestrated one of the largest wars by proxy in history, using the Taliban and Lashkar-e-Tayyiba against India to release the Indian grip on Kashmir.[15]

Normally, victim states do not retaliate against terror attacks with full-scale military force, but rather with restrained on-the-spot force or mild reprisals.[16] Interestingly, a doctrinal review of ad bellum rules reveals that their interpretation has transformed, especially after the 9/11 attacks.[17] Although there is no international consensus concerning these changes, the conditions under which forcible responses to terrorism are permitted have become more lenient. One crucial rule, however, has remained fairly stable—the ad bellum

---

13 *See generally* PATRICK SEALE, ASAD: THE STRUGGLE FOR THE MIDDLE EAST (1990). Seale reports that Assad jailed PLO leader Arafat when he suspected that Arafat deliberately planned to increase terror to a level high enough to drag Syria into an undesired war with stronger Israel. On the other hand, of course, weak states may make "calculation mistakes." Nasrallah, the leader of Hezbollah, stated following the 2006 war in Lebanon that had he anticipated Israel's severe response, he would not have committed the provocation that ignited it. *See Hezbollah Leader Nasrallah Regrets War*, CBS NEWS, Aug. 28, 2006, http://www.cbc.ca/news/world/story/2006/08/27/nasrallah-abduction.html.

14 *See generally* W. Hays Parks, *Lessons from the 1986 Libya Airstrike*, 36 NEW ENG. L. REV. 755 (2002).

15 *See generally* Bruce Riedel, *Pakistan and Terror: The Eye of the Storm*, 618 THE ANNALS OF THE AM. ACADEMY OF POLI. & SOC. SCI. 31 (2008); Brahma Chellaney, *Fighting Terrorism in Southern Asia: The Lessons of History*, 26 INT'L SEC. 94 (2002).

16 It took quite a few attacks by Libyan proxies in the 1980s before the U.S. responded with the El Dorado Canyon operation. *See* Parks, *supra* note 14, at 757–59. It took years of rocket attacks on Israeli settlements before Israel launched its 1982 campaign in Lebanon and its 2008 campaign in Gaza; in fact, 2,427 rockets were fired on Israel from Gaza in 2007 alone. *Israel Under Fire*, Israel Defense Forces, *available at* http://www.idfblog.com/facts-figures/rocket-attacks-toward-israel/ (last visited Feb. 10, 2015); *see also* Barry A. Feinstein, *The Legality of the Use of Armed Force by Israel in Lebanon—June 1982*, 20 ISR. L. REV. 362 (1985). Similarly, it took Turkey years of Syrian support of PKK's training and attacks before it issued its threat of full-scale war in 1998. *See* Eyal Zisser, *Syria and the United States: Bad Habits Die Hard*, 10(3) MIDDLE E. Q. 29 (2003). The U.S. approached Pakistan with an aggressive threat of war only after 9/11 and after years of the latter's known support of the Taliban and Al-Qaeda. *See* Riedel, *supra* note 15.

17 On the process of softening the ad bellum rules concerning the use of force in the war against terror, see generally Christian J. Tams, *The Use of Force against Terrorists*, 20 EUR. J. INT'L L. 359 (2009).

"threshold requirement." By that we mean the scale and effect of a terror attack that legitimizes a full-scale war by the victim state. Despite some difference of opinion, the consensus is that a full-scale retaliation in response to a relatively mild attack is prohibited. In what follows, we establish that doctrinal premise.[18]

B. Several Ad Bellum Rules Revisited

Recourse to full-scale war is legitimate under several ad bellum conditions, inter alia, where it has a "just cause."[19] Accordingly, Article 51 of the UN Charter conditions the use of force on a preceding "armed attack."[20] As indicated above, several aspects of this rule were reshaped in the course of the War on Terror led by the U.S.

First, in post-9/11 conflicts, it has become virtually a foregone conclusion that acts of terror can constitute an armed attack.[21] These horrific acts have shown that the scale and effect of terror may even exceed the harm of a conventional attack.[22] Accordingly, the U.S. resorting to force against Al-Qaeda and Taliban targets in Afghanistan was widely supported.[23] Furthermore, prominent scholars acknowledge that a non-state actor can be the perpetrator of

---

18 *See also* YORAM DINSTEIN, WAR, AGGRESSION AND SELF-DEFENCE (2011); ANTONIO CASSESE, INTERNATIONAL LAW (2d ed. 2005); Thomas M. Franck, *On Proportionality of Countermeasures in International Law*, 102 AM. J. INT'L L. 715 (2008).

19 In addition, a war must be the last resort, and its intensity must be proportionate. Three other (less important) ad bellum conditions are that war must be declared by an authorized organ, should be fought with right intention, and have a reasonable prospect of success. *See* Thomas Hurka, *Proportionality in the Morality of War*, 33 PHIL. & PUB. AFF. 34, 35–37 (2005). Once a war commences, the rules of legitimate warfare govern the use of force (jus in bello).

20 "Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations." U.N. Charter art. 51.

21 See, for example, Robert D. Sloane, The Cost of Conflation: Preserving the Dualism of Jus Ad Bellum and Jus In Bello in the Contemporary Law of War, 34 YALE J. INT'L L. 47 (2009). Corten is a rare opposing voice in that regard. See OLIVIER CORTEN, THE LAW AGAINST WAR: THE PROHIBITION ON THE USE OF FORCE IN CONTEMPORARY INTERNATIONAL LAW (2010).

22 See Sean D. Murphy, Terrorism and the Concept of "Armed Attack" in Article 51 of the U.N. Charter, 43 HARV. INT'L L.J. 41 (2002).

23 President Bush declared in his speech to a Joint Session of Congress that the events of 9/11 were acts of war. *See* President George W. Bush, Address to a Joint Session of Congress (Sept. 20, 2001). Congress accordingly authorized the President to use force. *See* S.J. Res. 23, 107th Cong. (2001). The U.N. Security Council has issued two resolutions supporting the use of force while indirectly implying that the attacks constitute an armed attack. *See* S.C. Res. 1373, U.N. Doc. S/RES/1373 (Sept. 28, 2001); S.C. Res. 1368, U.N. Doc. S/RES/1368 (Sept. 12, 2001). NATO activated its collective self-defense clause in Article 5, stating clearly that it was responding to an armed attack following 9/11. *See* Lord Robertson, NATO Secretary General, *An Attack on us All: NATO's Response to Terrorism, Remarks* (Oct. 10, 2001), *available at* http://www.nato.int/docu/speech/2001/s011010b.htm.

an armed attack even if such an attack does not reach the scale of the 9/11 events.[24] State practice reaffirms this perception.[25]

Second, imputing liability for an armed attack to the state that harbors the terror organization became easier post-9/11. Originally, the terror organization had to belong to a governmental structure, at least de facto.[26] For example, in its *Tadić* ruling, the International Criminal Tribunal for the former Yugoslavia (ICTY) lowered the requirement of states' responsibility to "overall control" over the terror organization.[27] Again, it seems that the events of 9/11 have reshaped the doctrine on states' responsibility for acts of terrorism (the "Bush Doctrine") to require only the broad basis of "harboring and supporting" the terror group.[28]

---

24 *See, for example*, Theresa Reinold, *State Weakness, Irregular Warfare, and the Right to Self-Defense Post-9/11*, 105 AM. J. INT'L L 244, 260-67 (2011); DINSTEIN, *supra* note 18, at 227–30; CASSESE, *supra* note 18, at 354–55.

25 *See* Cassesse's clear conviction: "Self-defense is the lawful reaction to an 'armed attack', that is to massive armed aggression . . . the aggression need not come from a state; it can also emanate from a terrorist organization or even from insurgents." CASSESE, *supra* note 18, at 354–55. *See also* DINSTEIN, *supra* note 18, at 227; Sloane, *supra* note 21, at 99–100; Tams, *supra* note 17, at 378. For instance, Turkey conducted repeated large-scale military campaigns in Northern Iraq against the PKK (without targeting Iraq at all). *See* Eunyoung Kim & Minwoo Yun, *What Works? Countermeasures to Terrorism: A Case Study of PKK*, 32 INT'L J. COMP. & APPLIED CRIM. JUST. 65, 73 (2008). *See also* Daren Butler & Davis Dolan, *Turkish Military Launch Operation Targeting Kurdish Rebel Hideouts*, REUTERS (Mar. 24, 2015), http://www.reuters.com/article/2015/03/24/us-turkey-kurds-idUSKBN0MK17H20150324. Colombia fought FARC forces in Ecuador's territory. *See* Frank M. Walsh, *Rethinking the Legality of Colombia's Attack on the FARC in Ecuador: A New Paradigm for Balancing Territorial Integrity, Self-Defense and the Duties of Sovereignty*, 21 PACE INT'L L. REV. 137, 137–39 (2009). The U.S. bombed terror camps in Sudan and Afghanistan in 1998 as a response to the attacks on its embassies in Kenya and Tanzania. *See* Tams, *supra* note 17, at 379–80. Israel conducted two full-scale wars in Lebanon while expressly arguing that it was fighting against the hosted terrorist organizations and not against Lebanon. The just cause of its war against Hezbollah in the summer of 2006 was widely recognized, including by the Secretary General of the U.N., the U.S. Senate, the G8 states, and the majority of the members of the Security Council. *See* U.N. SCOR, 61st Sess., 5489th mtg., U.N. Doc. S/PV.5489 (July 14, 2006); G8 St. Petersburg Summit Declaration, Middle East (July 16, 2006); Press Release, Secretary-General, *Secretary-General Says 'Immediate Cessation of Hostilities' Needed in Lebanon, Describes Package Aimed at Lasting Solution, in Security Council Briefing*, U.N. Press Release SC/8781 (July 20, 2006), *available at* http://www.un.org/press/en/2006/sgsm10570.doc.htm.

26 See Derek Jinks, State Responsibility for the Acts of Private Armed Groups, 4 CHI. J. INT'L L. 83, 88–90 (2003).

27 "In order to attribute the acts of a military or paramilitary group to a State, it must be proved that the State wields overall control over the group." Prosecutor v. Tadić, Case No IT-94-1-A, Appeals Judgment, ¶ 131 (Int'l Crim. Trib. for the Former Yugoslavia July 15, 1999), http://www.icty.org/x/cases/tadic/acjug/en/tad-aj990715e.pdf. *See also* Tom Ruys, *Crossing the Thin Blue Line: An Inquiry into Israel's Recourse to Self-Defense Against Hezbollah*, 43 STAN. J. INT'L L. 265, 277 (2007).

28 *See generally* Murphy, *supra* note 22; Steven R. Ratner, *Jus Ad Bellum and Jus In Bello after September 11*, 96 AM. J. INT'L L. 905 (2002).

Third, in calculating the scale of aggression that amounts to an "armed attack," a question arises whether a cumulative approach is possible. This question is relevant in the context of terror, which is usually conducted in "small doses," each of which may fall short of the required critical mass yet taken together may amount to an "armed attack." (The quantitative threshold itself is reviewed in the next subsection.) For instance, Israel relied on such an argument when it attacked Gaza in 2008 (and again in 2014) and Lebanon in 2006 in response to years of sporadic Hamas and Hezbollah rocket attacks on its southern and northern cities.[29] So too did the U.S. in the *Oil Platforms* case.[30] Accordingly, legal discourse shows growing acceptance of the victim state's accounting for the accumulated effect of all single provocative attacks.[31]

Fourth, the 9/11 attacks amplified the question of whether a victim state could preempt the armed attack in self-defense, and if so, to what extent. While anticipatory self-defense is no doubt recognized for the sake of preventing an imminent armed attack, the legitimacy of preempting a more remote attack is unclear. Yet it seems that several scholars acknowledge a state's right to preempt in self-defense future foreseeable attacks that need not be imminent.[32] Other scholars also accept the notion that preventing future attacks may be a legitimate target when the victim state faces repeated provocations.[33]

---

29 On Israel's reliance on the cumulative approach, see David Kretzmer, *The Inherent Right to Self-Defense and Proportionality in Jus Ad Bellum*, 24 EUR. J. INT'L L. 235, 243–44 (2013).

30 In the *Oil Platform* case, the ICJ used language that seemed to permit accumulating several incidents into an "armed attack" so as to legitimize the use of force in self-defense. *See* Oil Platforms (Iran v. U.S.), 2003 I.C.J. 161 (Nov. 6). A similar suggestive tone is used in Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v. Uganda), 2005 I.C.J. 1 (Dec. 19) (using the term "armed aggression").

31 *See, for example*, Roberto Ago, Addendum to the Eighth Report on State Responsibility, U.N. Doc. A/CN.4/318/ADD.5–7 (1980), *reprinted in* [1980] 32 Y.B. INT'L L. COMM'N; Kretzmer, *supra* note 29 (maintaining that there is a growing awareness that transnational terrorist attacks present states with a serious problem and increase the need to acknowledge the cumulative approach); Sloane, *supra* note 21; Tams, *supra* note 17.

32 *See, for example*, Thomas Hurka, *supra* note 19; David Mellow, *Counterfactuals and the Proportionality Criterion*, 20 ETHICS & INT'L AFF. 439 (2006); Eric A. Posner & Alan O. Sykes, *Optimal War and Jus Ad Bellum*, 93 GEO. L.J. 993 (2005).

33 *See, for example*, Oscar Schachter, *The Right of States to Use Armed Force*, 82 MICH. L. REV. 1620 (1984); MICHAEL N. SCHMITT, ESSAYS ON LAW AND WAR AT THE FAULT LINES (2012). In our view, when it comes to conflicts that involve terrorism, the gap between *prevention* and *anticipation* is smaller than contemporary literature suggests. The proponents of anticipation argue that it is sufficient if the victim state knows for sure that the decision to execute such attacks has been taken and that the capacities to execute the attack are within reach. *See* Schmitt, *supra*, 80–84. However, as Schmitt rightfully points out, this is typically the case with terror. Terrorist organizations are in a continuous state of transforming intentions to attack into actions. In this respect, when the victim state acts to *prevent* future terror attacks, in practice, it is acting to intercept a highly probable attack, for almost every preventive strike against a terrorist organization is a strike in anticipation.

C. The Threshold Requirement

In contrast to increasing flexibility in interpreting the above conditions governing resort to force against terrorism, the threshold requirement remains fairly robust. That is, a full-scale war is not legitimate unless the scale and effect of the triggering terror attacks is substantial. They must exceed some critical mass for it to constitute an "armed attack" under Article 51 of the UN Charter. Despite a variety of opinions in this respect, all seem to agree that if the preceding provocation is minor, it cannot be considered an armed attack and thus cannot justify a full-scale retaliation.[34]

In the Nicaragua case, the International Court of Justice (ICJ) required that, for an attack to constitute an armed attack, it must be of a "significant scale" or employ "grave forms of the use of force."[35] The ICJ upheld that approach after 9/11 in the case of *DRC v. Uganda*, considering responses of states to attacks by irregular forces to be potentially legitimate only if directed against "large scale attacks."[36]

The threshold requirement remains firm also beyond the ICJ. The Eritrea Ethiopia Claims Commission contended that "[l]ocalized border encounters between small infantry units, even those involving the loss of life, do not constitute an armed attack for the purpose of the Charter."[37] Further, scholars such as Cassesse view attacks as falling within Article 51 only if they reflect a "massive armed aggression . . . that imperils its life or government . . . the attack must be of such magnitude that one cannot repel it otherwise."[38] Tams concurs: "[T]he distinction [between armed attack and lesser provocations] as such has been defended rather vigorously."[39]

The analysis so far has reviewed the threshold requirement as part of the "just cause" test. That is, we quantified the scale and effect of the provocation

---

34 Dinstein justifies a more restrained retaliation by the attacked state. *See* discussion, *infra* note 40.

35 "As regards certain particular aspects of the principle in question, it will be necessary to distinguish the most grave forms of the use of force (those constituting an armed attack) from other less grave forms." Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. U.S.), 1986 I.C.J. 14, ¶ 191 (June 27). *See also* W. Michael Reisman & Andrea Armstrong, *The Past and Future of the Claim of Preemptive Self-Defense*, 100 AM. J. INT'L L. 525, 534 (2006).

36 *See DRC v. Uganda*, *supra* note 30, at ¶¶ 146–47. *See also Oil Platforms*, *supra* note 30, at ¶ 51, (reaffirming the distinction between "most grave" and "less grave forms" of the use of force).

37 Eri. Eth. Claims Comm'n, Partial Award—Jus Ad Bellum, Ethiopia's Claims 1–8, 4 (The Federal Democratic Republic of Ethiopia and The State of Eritrea), The Hague (Dec. 19, 2005).

38 CASSESE, *supra* note 18, at 354–55.

39 Tams, *supra* note 17, at 387. The threshold requirement is not to be confused with other aspects of self-defense that have been readjusted, such as the accumulation of past events or the inclusion of future possible attacks. Even if one accepts that sporadic past attacks as well as possible future ones add up when the legitimacy of the response is considered, the threshold itself remains high.

that constitutes an "armed attack" under the terms of Article 51. Some scholars, however, consider this threshold requirement as an ad bellum *proportionality* rule. Here too, a full-scale war cannot be deemed ad-bellum proportionate unless the preceding provocation also exceeds the threshold level.[40]

In summary, as opposed to the adaptation of other ad bellum requirements, the threshold requirement seems immutable—remaining stable regardless of whether it originates in the just-cause examination ("armed attack") or in the ad bellum proportionality rule. Given that it is very much a self-imposed norm by stronger states to limit their own defense policy, this requirement is far from self-evident as rational state action. Because customary international law reflects states' practice, we turn to rationalizing this relatively tolerant approach using game theory.

III. A NUMERIC EXAMPLE

The purpose of this section is to capture most of our analysis using a simplified numerical example. Assume a conflict between a strong state and a weak one, where the Subgame Perfect Equilibrium ("SPE") is based on three pillars.[41]

---

40 The ad bellum proportionality doctrine is notoriously vague; its meaning varies among international law scholars, only some of whom view it as a threshold test. Franck argues that determining whether a state has the right to use military force requires considering if such response is proportionate in the context of a specific provocation. *See* Franck, *supra* note 18, at 715, 718, 721. Similarly, Dinstein characterizes it as a reasonableness test of responding to force by counterforce. Minor or localized attacks can justify on-the-spot responses; only more serious attacks with massive effects can justify resorting to war. *See* DINSTEIN, *supra* note 18, at 261–64. Kretzmer believes that in conflicts involving terror groups and a liable harboring state, the victim state may respond with force to prevent future attacks and deter its rivals. Accordingly, proportionality weighs the scale of the preceding attacks, including their accumulated impact combined with the scale of future attacks if not prevented, against the intensity of the response. *See* Kretzmer, *supra* note 29, at 273–76. To some extent our analysis may also fit the "tit for tat" proportionality test, in the sense that we are exploring the scale of the provocation that legitimizes a full-scale war in response. *See id.* at 279–82 (noting reactions to the conflict between Israel and Hezbollah in the summer of 2006, undertaken in response to the latter's provocation).

41 Subgame Perfect Equilibrium (SPE) is a kind of perfection of Nash Equilibrium. Nash Equilibrium is a set of strategies, one for each player, from which unilateral deviation makes neither better off. In other words, if Player 2 adheres to her equilibrium strategy, Player 1's best response is to play according to his own equilibrium strategy as well. This basic concept of game theory has several drawbacks, one of which is the multiplicity of solutions. If there is more than one equilibrium, it is hard to predict which one will prevail. Consequently, several refining concepts have been developed over the years, leaving out equilibria that are for some reason unlikely. Perhaps the most known screening criterion is subgame perfection, which rules out unreliable equilibria. Specifically, an equilibrium may consist of conditional strategies; for example, "if Player 1 provokes you, wage a full scale war against it." If Player 1 does not provoke Player 2 in equilibrium, this scenario should never materialize, thus Player 2 technically never deviates from its prescribed equilibrium strategy. However, subgame perfection asks more than that. It is

First, there is a status quo of low terror attacks induced by the weak state and suffered by the strong one. In each period of this low-intensity conflict, terror is causing the strong state damage of 20. This status quo also reflects some kind of a "red line." If the terror inflicted by the weak state does not exceed this level, the strong state contains it and maintains the low-intensity conflict. Within such a low-intensity conflict, the parties infinitely witness a cycle of low-scale conflict that is not expected to escalate into a full-scale war, provided that the red line is not crossed.

Second, the strong state threatens to wage a full-scale war should the weak one provoke it by increasing the terror level above 20. If indeed such war takes place, for whatever reason, the strong state becomes less tolerant in the future. The new redline is lowered, even to, say, damage of 0 in all subsequent periods. The post-war lower tolerance reflects the "returns" demanded by the strong state on its costly war "investment." A full-scale war reduces the future level of terror by either undermining the weak state's capabilities or by increasing deterrence.

Third, if the weak side provokes the strong state, but the latter does not make good on its threat to wage a full-scale war, deterrence is eroded, and the weak state increases its attacks. The strong state becomes more tolerant, and the new red line in all subsequent periods is higher (instead of 20, it moves to, say, 30).

To prove this is indeed an SPE, one needs to show that under no circumstances is a party better off deviating from its prescribed equilibrium strategy. The weak state is worse off (by assumption) fighting a full-scale war against the strong state; therefore, given the threat it faces, it should never cross the red line, whatever it is. The strong state can save 20 in each period, presumably over an infinite period, if it wages war without prior provocation. Assuming that the present value multiplier is 10, the overall benefits of war amount to 10*20=200.[42] Let the costs of war be 300. Deviation by the strong state, then, is not cost-effective. It is irrational to invest 300 in order to save 200. Therefore, the weak state will maintain the terror level at 20, and the strong state will maintain the low intensity conflict. This is the equilibrium path.

---

not enough not to deviate from a strategy never triggered; Player 2 should adhere to its strategy assuming it *will* be triggered. If Player 2 is better off not waging war when provoked, then an equilibrium based on the assumption of nonprovocation is unreliable, because the other party is aware of that fact. Thus, SPE retains only reliable equilibria, in the sense that all players shall act as prescribed by their strategies even if some element of their strategies is not expected to be tested.

42 War prevents damage of 20 in each period, infinitely. However, future benefits are discounted periodically by a discount rate of, say, 1/9. When the damage from terror is summed across all periods infinitely, it converges to the product of 20 and the "present value multiplier," which, with some manipulation of the discount rate, can be expressed as (1+1/9)/(1/9)=10.

In order to substantiate the SPE, one needs to verify whether the strong state will live up to its threat when provoked by the weak one (although such provocation is not expected to take place in the equilibrium path). Suppose, therefore, that some provocation of the strong state above the red line is imminent. If the strong state wages a full-scale war (as prescribed by its equilibrium strategy) its overall costs are 300: these are the costs of war "today" plus the costs of terror in each future period (which is presumably 0).[43] If, on the other hand, the strong state decides not to go to war, it shall suffer terror of 30 in each future period, on top of the costs of the current provocation. With the same present value multiplier (10), total costs in damage amount to more than 300. Therefore, the threat to wage war when provoked is indeed credible.

However, renegotiation between the parties following provocation can change the cost-effectiveness of war for the strong state. The parties can agree to "skip war" and return to the original status quo ante, where the terror level was 20 (this is "weak renegotiation perfection" of SPE, following Farrell and Maskin[44]). Clearly, the weak state would agree: avoiding war with a much stronger rival is axiomatically better for it. As to the strong state, if it decides to avoid war in light of this "new agreement," its periodic costs decrease to 20, which amounts to 200 in present value terms (and its total costs are a bit higher, accounting for the costs of the current provocation). Hence war, which costs 300, becomes cost-*ineffective*. A rational strong state would thus probably also agree to resorting to the status quo ante and merely containing the provocation.[45]

Such conditions lead to both sides deviating from their equilibrium strategies: anticipating successful renegotiation, the weak state provokes the strong one and crosses the red line of 20; the strong state restrains itself and does not retaliate with full-scale war as "promised"; the weak state does not take advantage of its rival's restrained reaction (it agrees not to punish the strong state for violating its commitment to go to war and avoids further escalation of terror). All three pillars of the equilibrium, respectively, are violated. Therefore, the original equilibrium—in which the terror level is 20—is not renegotiation-proof. The underlying equilibrium path, whereby an intolerant strategy keeps

---

43 It should be noted that even if a full-scale war against the weak State reduces the costs of terror in the future to 0, it does not refer to the influence of war over the conduct of unmodeled exogenous actors, such as other states or organizations.

44 Farrell & Maskin, *supra* note 10, at 330–31.

45 Arguably, the weak state has an ongoing incentive to deviate from the status quo, with the anticipation of another successful renegotiation for similar reasons. Because the strong state knows that, one can argue that it might exercise its threat after all—an outcome the weak state seeks to avoid. In the proposed argument, we deal with this complexity by using a more refined perfection criterion than the one presented in this section.

terror at a (low) level of 20, is implausible. The threshold ad bellum rule that forbids waging a full-scale war in response to low-scale terror reflects this dynamic.

Suppose, alternately, that the terror level in each period is 100 instead of 20. The SPE is similarly based on the same three pillars: so long as the weak state does not cross the red line, the strong one keeps the conflict in its low-intensity form. However, if that line is crossed, the strong state threatens to wage a full-scale war. If a full-scale war erupts, for any reason, the subsequent new red line is lowered to 70. If the red line is crossed, however, and the strong state refrains from exercising its threat, deterrence is eroded and the new red line increases to 120. Showing why this is also an SPE tracks our first example.

In these circumstances, if the weak state provokes the strong one, returning to the status quo ante would *not* make the latter better off. Recall that war costs 300, and following it, terror decreases to 70 in each period. Thus, the overall cost of sticking to the equilibrium strategy is 1000 (periodical terror level of 70 times the present value multiplier of 10 is 700, whereas the cost of war is 300). On the other hand, deviating from the equilibrium strategy according to the concept of "weak renegotiation" restores the original terror level in each period, 100, which also comes to 1000 over time—and even more, if the costs of provocation are accounted for. The strong state has no incentive to deviate from its subgame perfect equilibrium strategy. It has threatened to go to war if a terror level of 100 is crossed and is expected to do so. Anticipating the failure of renegotiation, the weak state will accordingly not provoke the strong one by inflicting terror damage above the red line. So the original equilibrium—where the terror level was 100 in the status quo—is indeed weakly renegotiation-proof. The underlying equilibrium path, whereby a tolerant strategy sustains terror at 100, is plausible. The relatively tolerant threshold ad bellum rule in the context of such unequal conflicts reflects this equilibrium.

Given these two numeric variations, we speculate that this is why the terror level strong states experience in reality reflects the higher figure (100) rather than the lower (20). It also explains why low levels of terror cannot amount to an armed attack and justify a full-scale war by the victim state. Given the possible benefits of an intolerant trigger strategy, however, strong states can hypothetically benefit from credibly disabling the weak, terror-sponsoring state's ability to negotiate. High transaction costs can cause negotiation between the parties to fail, and indirectly perform as an exogenous commitment mechanism. If renegotiation is frustrated, the SPE—which is based on an intolerant trigger strategy—becomes renegotiation-proof. Thus, in terms of our numeric example, the strong state will suffer terror in the amount of 20, instead of 100, in each

period. Clearly this makes it better off and may induce it to violate the ad bellum threshold rule.[46]

IV. MODEL OF UNEQUAL BILATERAL CONFRONTATION

We now turn to presenting our argument in more detail. We start by discussing an unequal conflict in a one-shot game. In this conflict, the parties face one single interaction: an unequal bilateral confrontation where the weak side uses terrorism indirectly against its stronger counterpart. We then move to the more realistic scenario of a game with repeated interactions. The purpose is to introduce the concept of credible threats and refine the SPE analysis.

A. One-Shot Confrontation

Assume a conflict between a strong state, S, and a weak state, W, with complete and symmetric information. The conflict can be of low intensity, in which W uses guerilla tactics or terrorism against S either directly or via a proxy. And S, the strong state, confines its response to limited reprisals that do not escalate to full-scale war. Alternatively, the conflict can be of high intensity; namely, a full-scale war between the two states.

Due to its military inferiority, state W always prefers to avoid a full-scale war. Technically speaking, this means that the costs of going to war against S are higher than any alternative strategy that averts such war. In a low-intensity conflict, hostile actions against S serve W's long term interests, so we assume W derives benefits from inflicting damage on S. Therefore, W prefers the *net* terror damage suffered by S (that is, after accounting for the counter-terror measures taken by S) to be as high as possible.[47]

Denote the net terror level that S suffers by t, and assume that W controls t. It can restrain or encourage terror activity against S by either restraining or assisting the terrorist organizations it harbors. Accordingly, W "chooses" the

---

46 We assume that transaction costs are exogenous. Naturally, they are not. The strong state benefits from high transaction costs and may try to commit to keep them high (for instance, by electing hardline leaders). The international community, on the other hand, is composed of rational states whose motivation and ability to intervene can also be affected by the parties. This analysis, however, is suppressed in our work because it is ancillary to our research interests. This Article aims to rationalize the relatively tolerant threshold ad bellum rule, as well as to predict its level and when it will be ignored. We show that only tolerant strategies are renegotiation-proof when transaction costs are low (which explains the tolerant threshold requirement); intolerant ones are renegotiation-proof only when transaction costs are high (which predicts when this rule may be violated). This is true, arguably, irrespective of how the level of transaction costs is determined.

47 Of course, this is merely a reduced form. The weak state may not benefit per se from terrorizing the strong one. It might target the strong state for terror to achieve ideological, religious, or territorial gains.

level of terror damage, t, to be suffered by its rival, although it may not have the ability to fully and absolutely control t. It could be that the terror organization—albeit directed by *W*—is not completely subject to its authority. To reflect that, assume some minimum level of terror damage, $\underline{t}$, which takes place even in the face of maximal effort by *W* and *S* to prevent terror. Clearly, it could be that $\underline{t} = 0$, but this need not be the case. Therefore, $\underline{t}$ captures some "constant" level of terror suffered by *S* for which *W* cannot be held responsible.

During the course of a low-intensity conflict, *S* observes *W*'s actions and thus can predict the imminent net level of terror, t, before it actually materializes. In contrast to its rival, *S* naturally prefers t to be as low as possible. The only way, however, for *S* to prevent the expected net terror damage is by waging a full-scale war against *W* (recall that t already accounts for low-scale counter terror measures undertaken by *S*). In such case, *S* would have to endure the costs of war, which we denote by $\bar{t}$. Note that in this case, the assumption is that the terror damage does not materialize.[48] Naturally, we assume that $\bar{t} > \underline{t}$, which means that the cost of war is higher, probably much higher, than the lowest terror level that *W* can induce.

Thus, the one-shot confrontation is a simple two-stage sequential game. In the first stage, state *W* "sets" the expected level of net terror, t (but it does *not* materialize yet). In the second stage, *S*, which observes *W*'s actions and thus can predict t, has a binary decision. It can either endure the net terror damage, t, thereby maintaining a low intensity conflict with state *W* who benefits from t; or *S* can elect to escalate the conflict into a full-scale war and bear its costs, $\bar{t}$ (with *W* suffering enormous damage).[49]

The one-shot game has a unique subgame perfect equilibrium (SPE): *W* chooses $t = \bar{t}$; *S* maintains a low intensity conflict if $t \leq \bar{t}$, and otherwise switches to full-scale war (note that the SPE strategy by *S* addresses all possible actions by *W*, including the ones that are not expected to take place). On the equilibrium path, thus, there is a low intensity conflict, where the threshold for launching a large-scale retaliation by *S* is rather high: $\bar{t}$.

To see why this is indeed an SPE, consider first the strategy of *W*. As the weaker state, it always prefers to avoid a direct war against *S*. Subject to that constraint, it prefers t to be as high as possible. *W*'s choice of the maximal terror level that still allows avoiding war ($t = \bar{t}$), therefore, maximizes its well-being.

---

48 The concept of war that fully prevents the imminent terror level is technically convenient. It could be, of course, that war does not prevent terror in full. The idea, however, is that the terror damage is absorbed in the costs of war borne by country *S*, yielding an overall damage level of $\bar{t}$.

49 This model is in fact a simple entry-deterrent model, as used by Goldsmith and Posner. *See generally* Goldsmith & Posner, *A Theory of Customary International Law*, *supra* note 5. However, the decision of the weak state is not binary (whether or not to "enter") but continuous.

Hence, it cannot profitably deviate from its equilibrium strategy: lower terror undermines its goals, and higher terror is expected to trigger a full-scale war.

S, on the other hand, cannot profit from war if $t \leq \bar{t}$, because in this case the cost of war outweighs its benefits (namely, the terror damage it eliminates). However, it will exercise its threat if the terror damage is expected to exceed the costs of war; that is, if $t > \bar{t}$. Therefore, the expected critical mass of terror, above which *S* wages a full-scale war against *W*, is in fact its cost of war, $\bar{t}$.

What would happen if *S* tried to impose a lower, less tolerant threshold? Assume that the trigger for a full-scale war is $\underline{t}$. Namely, if *W* does not do everything in its power to reduce terrorism to its minimal controlled level, then *S* will retaliate with a full-scale war. *S* will not deviate from this strategy on the equilibrium path because it reduces its damage from terror to a minimum. Due to *W*'s military inferiority, it also will not deviate, instead setting the terror level at $\underline{t}$. Therefore these strategies represent an equilibrium under which *S* is better off. It is not, however, an SPE and thus is not expected to happen in practice. It is based on a noncredible strategy by *S*, whereby if *W* provokes it and increases terror above $\underline{t}$ (but below $\bar{t}$), *S* wages a cost-inefficient full-scale war. In other words, this equilibrium is based on a dubious assumption, whereby *W* believes that *S* would launch a cost-ineffective war. This presumption renders the whole concept of an intolerant trigger strategy implausible.

Nevertheless, this notion does not fully explain why strong states do not use intolerant strategies more often in unequal confrontations. Specifically, international conflicts are almost always dynamic and repeated. In this case, the "returns" from war should also take into account future benefits in the form of terror attacks that the war can prevent. In these circumstances, the strong state can more plausibly build a credible threat. If so, our original question regarding the rationalization of the threshold rule remains unanswered.

B. Dynamic Confrontation

An intolerant trigger strategy by the strong state cannot establish an SPE in the single confrontation scenario because it relies on noncredible threats. This section shows that in an infinitely repeated game, almost any set of payoffs can be supported in an SPE.[50] Namely, the strong state can build credible threats that are supported in equilibrium.[51]

---

50 See also Quan Wen, A Folk Theorem for Repeated Sequential Games, 69 REV. ECON. STUD. 493 (2002).

51 Credible threats play a crucial role in asymmetric conflicts that involve terrorism. A clear example is the explicit Turkish threat against Syria in 1998. Turkey demanded that the PKK leader be expelled from his shelter in Syria and that Syria immediately cease its support of PKK. Determining the threat to be credible and understanding the grave implications of its

The threat of a full-scale war is credible if and only if waging it when provoked is cost-effective for the strong state. This happens when the benefits of war for the strong state are higher than its costs. The costs of war in the dynamic formation are the same as in the one-shot scenario, $\bar{t}$. And yet, the benefits differ. The benefits of war in a one-shot game reflect the terror that war prevents in that period; in the repeated game, the benefits of war also account for the terror the full-scale war prevents in all subsequent periods.

For instance, assume two periods of confrontation between S and W, both of whom possess perfect foresight. If S chooses not to go to war and maintains a low intensity confrontation, it suffers t in each period. If it goes to war, it suffers $\bar{t} = 300$ in the first period and $\underline{t} = 30$ in the second one. To enable inter-temporal comparison of values, denote the periodical discount factor of state S by δ. Let $\delta = 0.9$.[52]

If S contains the terror attacks, it suffers $t + 0.9t$.[53] If S wages a full-scale war, it suffers $\bar{t} + 0.9\underline{t}$.[54] Thus, S is indifferent between both options if $t + 0.9t = \bar{t} + 0.9\underline{t}$, which implies that $t \approx 172$. Namely, if the terror level in each period is 172, the costs of war equal its benefits. If the terror level in each period is higher than 172, war will be preferable to the low intensity conflict. Accordingly, if the terror level in each period is lower than 172, the low-intensity conflict alternative is preferable to S.

Figure 1 generalizes the above example to any repeated confrontation game with two features. First, without war, the current level of terror, t, remains in subsequent periods; second, following war, the terror level decreases to the minimum, $\underline{t}$:

FIGURE 1: THE INDIFFERENCE THRESHOLD

consummation, Assad complied with the demands in full. *See* Zisser, *supra* note 16, at 35–37; Kim & Yun, *supra* note 25, at 73. Another example concerns Pakistan's leader, General Musharaf, who explained that Pakistan's cessation of relations with the Taliban was purely the result of an American ultimatum to join its War on Terror or suffer the consequences. *See* Riedel, *supra* note 15, at 37. The self-imposed Libyan and Iranian cessation of their nuclear programs in 2003 is another example of a credible threat of war by a strong state toward a weaker one. *See generally* Oscar Schachter, *The Extraterritorial Use of Force against Terrorist Bases,* 11 HOUS. J. INT'L L. 309 (1989).

52 The "discount factor" equals 1/(1+discount rate). Thus, the discount rate here is 1/9.

53 S actually suffers two periods of terror level t. The terror in the second period is multiplied by δ so that it equals $0.9t$ in present terms. Therefore, overall terror damage is $t + 0.9t = 1.9t$.

54 If S instigates war, it will suffer war's cost in the first period $\bar{t}$. In the second subsequent period, S will suffer the minimal terror level, being reduced to $\underline{t}$ as the benefit of war, calculated in its present value to be $0.9\underline{t}$. The total cost of terror in both periods then is $\bar{t} + 0.9\underline{t}$.

If $t < \tilde{t}$, the costs of war are higher than its benefits. If $t > \tilde{t}$, the costs of war are lower than its benefits. Accordingly, $\tilde{t}$ is the *indifference threshold* where the costs of war equal its benefits.[55] As for W, a full-scale war against S makes it worse off regardless of the alternative. Therefore, we retain our assumption that W aims to maximize the terror level while avoiding a full-scale war with S.[56]

The game remains exactly the same, only that instead of playing it in a single period, it is played infinitely. Namely, in each period, W chooses a level of terror t, and S, who observes that, decides whether to contain it or wage a full-scale war. The SPE we focus on in this dynamic game is similar to the one-shot game.[57] Both parties "agree" on a certain level of terror. The strong side backs this level up with an appropriate trigger strategy. The trigger materializes if W provokes S by deviating and increasing terror beyond the initially "agreed" level.[58] The difference, however, is that in the repeated game scenario, the equilibrium also accounts for future periods.

To establish an SPE, the following conditions must hold:

(1) The costs of war for S should be higher than its benefits from an unprovoked war. Thus, S will not wage war against W when not provoked.

(2) The costs of war for S should be lower than its benefits from a provoked war. Thus, S will wage war against W when provoked.

(3) The costs of war for W are higher than any possible benefit. Thus, W will not provoke S.

Figure 2 presents a *possible* set of SPE equilibria. The horizontal axis describes imminent levels of terror damage S suffers in any given period; the vertical axis describes the periodical terror level in the following future periods, given the parties' strategies.

---

55 Note that $\tilde{t}$ is the current terror level for which S is indifferent between going to war and subsequently reducing the terror level to minimum, and not going to war (thereby maintaining the current terror level). If the horizon of future confrontations is infinite, then:

$$(1)\ \tilde{t} = \delta \underline{t} + (1 - \delta)\bar{t}.$$

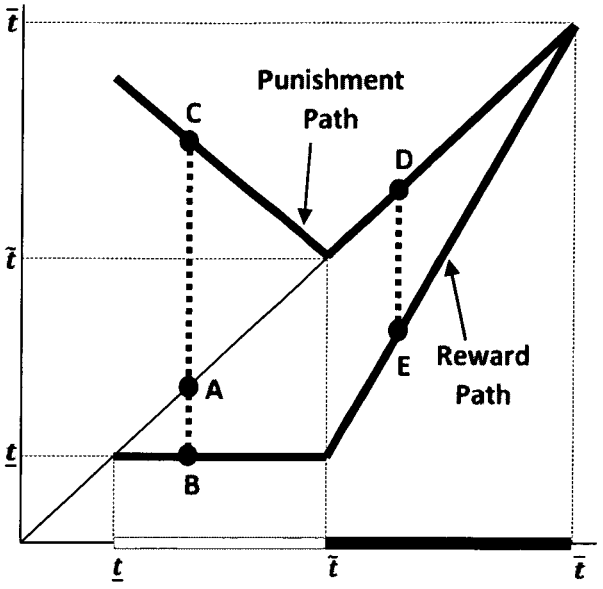
Namely, the indifference threshold is a combination of the minimal terror level, $\underline{t}$, and the costs of war, $\bar{t}$. It is positively correlated with the costs of war, $\bar{t}$, the minimum terror level, $\underline{t}$, and the discount factor, δ.

56 Formally, it means W is better off sponsoring no terror at all than suffering the costs of war, and "benefitting" from the maximal terror level, $\bar{t}$, afterwards in all periods.

57 We argue that an SPE can be formed above the indifference threshold, $\tilde{t}$, and also below it. To that end, both assumptions of the dynamic confrontation in the example above are relaxed: without war, the current terror level may change in the future and, following war, the terror level is not necessarily minimal.

58 Naturally, we are not arguing that these strategies are the only ones that can establish an SPE. We do argue, however, that these strategies deserve special attention, because unequal and indirectly managed conflicts in reality are often characterized by similar strategies: a trigger strategy by the strong side, with constant attacks by the weak side.

FIGURE 2: A POSSIBLE SPE IN AN INFINITELY REPEATED GAME



The terror level on the equilibrium path can be anything between $\underline{t}$ (below which is impossible by assumption) and $\bar{t}$ (above which war is preferable irrespective of any future periods, as shown in the previous section). Thus, assume that the terror level in equilibrium is some t within this range. If the parties remain loyal to their strategies (which, by the definition of an equilibrium, they are expected to do), this status quo will persist indefinitely. Namely, in each period W will cause net terror at the level t, and S will contain it without escalating the conflict any further. This is exactly why the equilibrium path is in fact any point (above $\underline{t}$) on the 45-degree line.

If, however, W deviates and intends to raise the terror level beyond t, then S faces two alternatives: it can keep up its strategy and launch a full-scale war, or it can avoid war contrary to its predefined threat. If it chooses war, the terror level in subsequent periods is reduced as a result due to W's diminished capacity and S's increased deterrent effect. The level of terror in all future periods (depicted by the vertical axis) is lowered to what the "Reward Path" describes: the future post-war level of terror that corresponds to the present, imminent pre-war level of terror. Clearly, the extent to which the post-war terror level falls below the current one depends upon the intensity of the power inequality between the states, among other exogenous elements. Greater military inequality

results in lower levels of post-war terror damage. Naturally, therefore, the reward path in equilibrium can take many forms, of which Figure 2 displays only one.[59]

If, however, contrary to *S*'s threat, war is avoided, then *W* punishes *S* by increasing the terror level in all future periods to that described by the "Punishment Path": future periodical levels of terror from which *S* will suffer in the absence of its full-scale war retaliation for a current deviation by *W*. It reflects the erosion in *S*'s deterrent effect when its threats are shown not to be credible. Clearly, the punishment path can also take many forms, of which Figure 2 displays only one.[60]

For example, Point *A* on the 45-degree line depicts a possible equilibrium path below the indifference threshold, $\tilde{t}$. If the level of terror in a certain period is *A*, the terror level in all subsequent periods is also *A* (on the equilibrium path). This status quo also marks the "red line" which *W* must not cross. Otherwise, *S* threatens to wage a full-scale war against it. If *W* deviates and increases the terror level beyond that described by *A*, *S* can either respond with a full-scale war according to its threat and reduce the future periodical terror level to *B*, or avoid war, prove its threat to be empty, and therefore induce punishment in the form of *W*, increasing the future periodical terror level to *C*.

Note that if war takes place, *S* is "rewarded" by a lower terror level, and that new terror level—which manifests a lower tolerance by *S* for terror—becomes the new status quo.[61] The lower post-war terror levels are the "returns" *S* receives for its "investment" in a full-scale war.[62] Therefore, the post-war

---

59 In fact, the reward path describes the *maximal reward path* for which unprovoked war is not cost-effective. That is, if the post-war terror level is lower than the reward path presented in Figure 2, *S* is better off waging a full-scale war even if not provoked, in violation of the first condition of the SPE.

60 Given the reward path, the punishment path displays the *minimal punishment path* for which provoked war is cost-effective. That is, if the "post-bluff" terror level is lower than the punishment path presented in Figure 2, then *S* is better off not carrying out its threat when provoked, in violation of the second condition of the SPE. Each possible reward path (out of infinite possibilities) has infinite corresponding punishment paths that together form an SPE. The punishment path in Figure 2 is the lowest among all possible punishment paths in all SPEs of the sort we describe (the lower envelope). The reason, of course, is that the reward path in Figure 2 is the maximal one—it exhibits the lowest possible post-war terror level in an SPE.

61 It should be noted that war could be launched whether or not following provocation by *W*. However, a strategy that reflects SPE, by definition, requires that an unprovoked war always be cost-inefficient.

62 Following the war in Afghanistan, the anti-American Al-Qaeda terror level dropped dramatically, as was the case following the 2006 Israeli war in Lebanon. *See* Amos Harel, *Five years after war, Israel-Lebanon border is quieter than ever*, HAARETZ (July 12, 2011), *available at* http://www.haaretz.com/blogs/2.244/five-years-after-war-israel-lebanon-border-is-quieter-than-ever-1.372727; Jean-Loup Samaan, *From War to Deterrence? Israel-Hezbollah Conflict Since 2006*, U.S. ARMY WAR COLLEGE STRATEGIC STUDIES INSTITUTE (2014), *available at* http://www.strategicstudiesinstitute.army.mil/

(periodical) terror level sets in fact a new red line, $\underline{t}$, as marked by point B on the "reward path." The above trigger strategy applies to the new red line exactly as to the old one.

On the other hand, if W deviates from the equilibrium path (by escalating the terror level above the red line) and S fails to act on its threat, deterrence is eroded. S is "punished" for its empty threats. Hence the higher "post-bluff" terror levels S suffers as depicted by the "punishment path." Accordingly, the new red line is marked by point C on the "punishment path," exhibiting S's higher tolerance. The above trigger strategy applies to the new red line exactly as to the old one.[63]

The vertical difference between points C and B represents the future periodical benefits of a *provoked* war. It is the level of terror damage that war can prevent in each future period, after prior provocation by W. The vertical difference between points A and B represents the future periodical benefits of *unprovoked* war. It is the level of terror damage that war can prevent in each future period, without prior provocation by W. Because this is an unprovoked war, its benefits do not include the saved punishment; not going to war when unprovoked is not expected to entail punishment (recall that the punishment reflects the erosion of deterrence, which is relevant only if a threat proves empty).

In the above example, points A, B, and C form an SPE equilibrium because the following conditions apply. First, the benefits of an unprovoked war are not higher than the cost of war (the first SPE criterion).[64] Hence, S is never

---

pdffiles/pub1198.pdf. Indeed, at least in the short run, this is also the result of destroying the terror infrastructure. However, terror organizations gradually restore their capabilities, and thus long run quiet is gained mainly by the deterrence effect of the full-scale war (as mentioned, this model does not take account of the influence of war over external parties. However, if these expected influences, whether harmful or beneficial, are foreseeable, they can be incorporated into the calculation of the expected cost of war).

63 India's slow and mild response to Pakistani-sponsored terrorism, including surrender to hijackers of an Air-India jet, arguably eroded deterrence and contributed to the increase in anti-Indian terror. *See* Chellaney, *supra* note 15, at 97–99. The Israeli case also demonstrates this "drizzling effect." Sporadic rocket attacks on both its northern (by Hezbollah) and southern borders (by Hamas and Islamic Jihad) gradually increased in frequency and volume until a critical mass was reached and a large military campaign launched to restore deterrence.

64 Formally, points A, B, and C comprise an SPE equilibrium since the following conditions apply, as reflected by this inequality:

$$(2)\ A + \frac{\delta}{1-\delta}(C - B) \geq \bar{t} \geq A + \frac{\delta}{1-\delta}(A - B).$$

The inequality on the right, $\bar{t} \geq A + \frac{\delta}{1-\delta}(A - B)$, corresponds to the first condition of the equilibrium: the left-hand term is the cost of war; the right-hand term is the benefit from an *unprovoked* war. The latter has two components: the present imminent level of terror that war absorbs, A, and the level of terror that war prevents in each future period, $A - B$, multiplied by a

tempted to go to war when not provoked. Second, the benefits from a provoked war are at least as high as the cost of war (the second SPE criterion).[65] Hence, S is never tempted to forego its threats. (Generally, the reward and punishment paths are constructed such that the first two conditions of the SPE hold for *any* level of terror.[66]) The third condition of the SPE, inducing state W not to provoke S, is implied. Given that its stronger rival punishes it for deviation, W not only suffers extreme costs of war but is deterred into "settling" for a lower terror level in future periods.

A second example can be found, given a different benchmark terror level, as reflected by point D in Figure 2. This time, the elected terror level is above the indifference threshold, $\tilde{t}$. If W deviates from the equilibrium path implied by D, and S responds with a full-scale war, then the post-war terror level is reduced to E on the reward path. If S fails to respond with a full-scale war to provocation by W, the terror level is at least as high as before, at point D, because the punishment path coincides with the equilibrium path in this area. The current level of terror is high enough to make provoked war cost-effective. In this case, thus, no actual punishment of S by W is needed in order to induce it to wage a full-scale war when provoked.[67]

In conclusion, as implied by Figure 2, any terror level can be supported in an SPE. Specifically, an intolerant trigger strategy, which is based on a very low threshold, can be supported in an SPE. Every point on the 45-degree line is a combination of rewards and punishments, such that the three SPE conditions attain. Low levels of terror can be supported in an SPE because there can always

---

present value multiplier, $\frac{\delta}{1-\delta}$ (recall that δ is the periodical discount factor). Therefore, an unprovoked war is not cost-effective.

65 Recall equation (2), $A+\frac{\delta}{1-\delta}(C-B) \geq \bar{t} \geq A+\frac{\delta}{1-\delta}(A-B)$. The inequality on the left, $A+\frac{\delta}{1-\delta}(C-B) \geq \bar{t}$, corresponds to the second condition of the equilibrium: again, the right-hand term is the cost of war; the left-hand term is the benefit from a *provoked* war. The latter has two components: the present imminent level of terror that war absorbs, A, and the level of terror that war prevents in each future period, $C-B$, multiplied by the present value multiplier. Therefore, a provoked war is cost-effective.

66 Note that the vertical gap between the punishment and reward paths increases as the level of terror on the equilibrium path decreases. The intuition is simple. The benefits of a provoked war are the terror damage it manages to prevent, imminently and in the future. These benefits must be at least $\bar{t}$ for the threat of war upon provocation to be credible (condition (2) of the SPE). Hence, when the imminent benefits of war (as captured by the terror level on the equilibrium path) decrease, the future benefits (as captured by the gap between the punishment and the reward paths) must increase, so that the benefits of war remain at least $\bar{t}$.

67 Points D and E form an equilibrium because all three criteria of forming an SPE are met. Accordingly, the following conditions apply (in equality):

(3) $D+\frac{\delta}{1-\delta}(D-E) \geq \bar{t} \geq D+\frac{\delta}{1-\delta}(D-E)$.

be a high enough punishment (if war is avoided) to render a provoked war cost-effective. So, if indeed such a strategy can practically eradicate terror, why do strong countries so uncommonly use it? Why do strong states rarely apply intolerant and apparently effective strategies in a manner that would reshape the ad bellum threshold rule of customary international law?

V. RENEGOTIATION-PROOF EQUILIBRIA

A central problem with the set of SPEs discussed above is indeterminacy. Any level of terror can be supported in equilibrium, including very low ones. The ad bellum threshold rule, however, implies that states practice relatively tolerant approaches towards terror. The above analysis does not make the high-terror equilibria (based on tolerant trigger strategies) more likely than the low-terror ones (based on intolerant trigger strategies). It seems that, at least in customary international law, states choose tolerant strategies for the most part.

The final phase of our argument offers a rational basis for this refined choice of strong states and the consequent tolerant ad bellum threshold rule. In fact, the proposed argument even assists in predicting the less common occasions in which states are likely to violate the ad bellum threshold rule and apply an intolerant strategy. In explaining the likelihood of tolerant strategies, we use renegotiation-proof refinement criteria, following Farrell and Maskin. This analysis of foreseeable renegotiation filters out intolerant trigger strategies as implausible.[68]

A. Weak Renegotiation-Proofness

An equilibrium strategy is subgame perfect only if its underlying threat is credible: if called upon to carry out a threat, neither party wishes to *unilaterally* back out of the strategy prescribed to it. Thus, when provoked, S is better off waging a full-scale war against W, although it is a costly action. The fact that such punishment harms both parties, however, may encourage *bilateral* deviation from the SPE strategies if it makes both parties better off. This refinement is also known as Pareto perfection.[69]

Renegotiation perfection has two refinement levels. We start with the one that is referred to as "weak" by Farrell and Maskin.[70] According to the predefined strategy, if both states "cooperate" in agreeing (tacitly or otherwise) on some terror level, and yet W deviates, S should wage a full-scale war.

---

68 *See* Farrell & Maskin, *supra* note 10.

69 *See* Eric Van Damme, Stability and Perfection of Nash Equilibria (1991); Drew Fudenberg & Jean Tirole, Game Theory (1991).

70 Farrell & Maskin, *supra* note 10.

However, because war harms both parties, if they can successfully communicate, either may propose to the other to return to the status quo ante, ignoring the one-time provocation. Such a proposal can be Pareto superior to adhering to the SPE strategies. If both parties are expected to mutually deviate from their prescribed strategy as a result of foreseeable renegotiation, the underlying original SPE is *not* weakly renegotiation-proof (and thus implausible). Accordingly, all the players' strategies not resistant to such renegotiation should be filtered out in advance from the set of SPEs. This refinement would foreground the implausibility of intolerant strategies that accounts for the relatively tolerant nature of the ad bellum threshold requirement.

Consider the SPE that is based on an intolerant trigger strategy, as displayed in Figure 2. It originally has three possible paths of terror: the equilibrium path terror level, A; the corresponding reward path terror level, B; and the relevant punishment path terror level, C. As we showed in the previous section, if W deviates from the equilibrium path by planning to increase terror above A, S is better off pursuing its prescribed strategy, given that W does so as well. However, both parties are better off avoiding the war (and its subsequent post-war terror path) and returning to the original equilibrium path. That is, S would agree to look the other way "just this once" and avoid going to full-scale war against W despite its provocation (such a war would have brought terror level to B on the reward path). W, in return, would agree not to "punish" S, namely, not to take advantage of the latter's containment reaction to the provocation, and also return to the original equilibrium path. (Note that such punishment would have increased terror level to C, on the punishment path.).

This renegotiation leads to Pareto improvement. W is clearly better off: it manages both to avoid a costly war as well as the lower level of terror followed by it (recall that higher terror level is in W's best interest). As for S, its original incentive to wage full-scale war is conditional on the benefits of the provoked war. These benefits largely depend on the future periodical gap between the punishing and rewarding levels of terror [$(C - B)$ in Figure 2]. Clearly therefore, if S is *not* punished for not waging a war when provoked, a full-scale war becomes cost-ineffective. Thus, S would also be better off not carrying out its threat and returning to the status quo ante. Therefore, this SPE is implausible since it is not a weakly Renegotiation-Proof Equilibrium ("RPE").

Generally speaking, such renegotiation would be successful if and only if the minimal punishment path is higher than the equilibrium path. In Figure 2, the minimal punishment path is indeed higher than the equilibrium path below the indifference threshold, $\tilde{t}$, and it is *not* higher than the equilibrium path above

the indifference threshold. *Consequently, any SPE in the interval* $[\underline{t}, \tilde{t})$ *is not weakly RPE and any SPE in the interval* $[\tilde{t}, \bar{t}]$ *is weakly RPE.*[71]

The intuition is simple. When the level of terror damage is relatively low, there is not much to gain by war. Even the "best" war that is expected to eliminate terror may not be a sufficient incentive for S to carry out its threat when needed. In this case, only if S is punished for false threats will war become cost-effective and thus ex ante credible. That is, only if war prevents more severe terror attacks than currently are suffered will it become cost-effective. Any actual punishment, however, leaves some wedge for this simple, and very specific, so-called "weak renegotiation." This renegotiation can only succeed if both parties have some benefit to "sell" to the rival party. These benefits will render the initial strategy inferior compared to the successful conclusion of the negotiation. S, the stronger state, can offer W to skip the destructive war. W offers in return to refrain from punishing S if the latter does not carry out its threat. As a result, renegotiation will be successfully concluded only when a punishment is relevant. In Figure 2, this applies to all points below the threshold level $\tilde{t}$. Insofar as the original SPE must be supported by punishing S, then both sides have something to gain by resorting to the original status quo, instead of going through a costly war.

On the other hand, when the equilibrium path overlaps the minimal punishment path, the terror level is relatively high. Hence, the benefit of a full-scale war for S exceeds the alternative suggestion, namely, returning to the status quo and containing the provocation. In this case, the offer to restore the status quo ante is simply not enough to discourage S from carrying out its threat to wage a full-scale war. The underlying SPE is therefore sustainable.

This simple refinement tool provides our main explanation for the tolerant nature of the threshold requirement. Although we have shown that intolerant trigger strategies are credible and can theoretically eradicate terror, they are not commonly used in reality because such strategies are not robust to the weakly renegotiation-proof criterion. Hence, the customary ad bellum threshold rule, which reflects this practice, is not an intolerant one.[72]

B. Strong Renegotiation-Proofness

The concept of weakly RPE suffers, in our context, from three downsides. First, while it manages to narrow the set of terror levels that are sustainable in

---

71 *See* proof in the Appendix, *infra.*

72 This argument is restricted only to the SPE based on the specific trigger strategies described above. It does not purport to eliminate, based on the weak RPE refinement criterion, *any* equilibrium whose path displays relatively low levels of terror.

equilibrium, the remaining scope is still rather large. Indeed, this perfection criterion explains why a low threshold rule is not sustainable; however, it cannot give a more precise prediction as to the actual threshold rule.[73]

Second, the weak renegotiation refinement is restricted to a very specific kind of renegotiation, in which only one proposal is on the table: returning to the status quo ante. This limitation, of course, is avoidable. If the parties indeed can communicate following a provocation, they are not necessarily bounded by the status quo ante (despite its status as a focal point in international relations).

Third, one may argue that ruling out an SPE by weak renegotiation perfection is reasonable only if the provocation is isolated or mild or inadvertent, or if the new arrangement it offers can somehow be backed up by the international community. Otherwise, despite the promise to return to the status quo, S may wage a full-scale war after all. The reason, arguably, is that nothing will stop W from repeating the "provocation and renegotiation" routine in the future as well. So, in fact, when repeated provocations are expected, the option of restoring the status quo ante is not realistic. In such a case, S may prefer a full-scale war after all.[74]

To address these concerns, we employ the stricter perfection criterion offered by Farrell and Maskin—strong renegotiation-proofness. The idea is that equilibrium should not be robust only to an offer that transfers the parties, internally, from one path to another. It should be robust to an offer that transfers the parties to *any* other weakly RPE. The intuition is simple: if both parties are better off deviating concurrently to a different equilibrium, at any point during the course of a given equilibrium, then that equilibrium is Pareto inferior and thus implausible. Formally, an SPE is considered a strongly RPE if it is weakly RPE and there exists no other weakly RPE that Pareto dominates it. We submit that *the lower bound of the weakly RPE set is the only strongly RPE.*

---

73 Further, recall that Figure 2 displays only one SPE, but there are infinitely more of these. For instance, the reward path can be higher than the one in Figure 2 (but still below the equilibrium path) if the military inequality between the states is lower. That is, the results of a full-scale war deter the weak state to a lesser degree, and therefore the post-war terror level is higher than the one described in Figure 2. In such a case, the minimal punishment path must also be higher than the one in Figure 2.

74 A similar pattern emerges from the Israeli-Hamas conflict. The involvement of the U.S. and effective Egyptian mediation, as well as the heavy costs of war to both parties, enabled several rounds of "renegotiation" following Hamas deviations from the fragile status quo. In the 7 days of the conflict in 2012, which involved targeted Israeli air strikes and massive Hamas shelling of more than half of Israel's population, effective mediation prevented it deteriorating into a full-scale war and a massive Israeli incursion. Israel accepted the renegotiated terms, however, not based on the assessment that Hamas *wanted* to keep its word but rather based on the assessment that Hamas *needed* to comply based on its own interests. *See* Aaron David Miller's analysis of expected Hamas compliance with the renegotiated terms, *infra* note 86.

To see why, take the interval $(\tilde{t}, \bar{t}]$ in Figure 2. All the SPEs that correspond to the terror levels within this interval are weakly RPE, as established in the previous section. If the terror level in equilibrium is anywhere above the lower bound, $\tilde{t}$, however, a Pareto superior deal can still be achieved after provocation by W. In return for "looking the other way" when S is provoked, W commits to a *lower* terror level in a different weakly RPE than the original one. Both states are better off this way, implying that the original weakly RPE is not strongly RPE.

The reason is simple. W is always better off avoiding a full-scale war. For marginal provocations, S is only slightly better off following its equilibrium strategy and waging a full-scale war. If offered a new arrangement whereby avoiding war strictly decreases the terror level, it becomes better off not waging war. Therefore, both states will agree. Clearly, however, such renegotiation is not possible if the original equilibrium path is $\tilde{t}$. S cannot be persuaded to forego the war for the promise to switch to a lower level terror, because the latter does not underlie a weakly RPE. Therefore, only $\tilde{t}$ is a strongly RPE, which reflects the best prediction for the ad bellum threshold requirement.[75]

The example of the recent U.S.-Syria conflict is striking in this regard. President Obama had set a clear red line for the Syrian regime—that is, the use of chemical weapons. Syrian president Assad violently breached that line. The U.S.'s threat to go to war was credible, and the strike was imminent. Last minute discussions brokered by Russia yielded a deal which involved disarming Syria of its chemical capabilities in exchange for not suffering war against a militarily superior state. In other words, the U.S. agreed not to attack and Syria has agreed to switch to a different equilibrium, whereby its mass destruction capabilities are neutralized.[76]

We have shown that the strongly RPE addresses the first two downsides of the weak renegotiation-proofness refinement criterion. That is, the prediction as to the actual threshold rule became clearer, indicating a specific critical mass of terror, $\tilde{t}$, which justifies resort to a full-scale war by the victim state. Further, the nature of the negotiation is no longer restricted to returning to the status quo. The third critique is different, however. It does not question the equilibria that

---

75 Interestingly, $\tilde{t}$ is the lower bound for all possible strongly RPEs, because it is based on the globally minimal punishment path (the lower envelope for all punishment paths). Therefore, even if the parties are extremely differentiated by their military capabilities, terror cannot be totally eradicated (or contained to its minimum level, $\underline{t}$). The terror level is bounded by the indifference threshold, $\tilde{t}$. Therefore, it reflects the most plausible threshold rule.

76 *See, for example, Syria Hails U.S.-Russia Deal on Chemical Weapons*, BBC NEWS (Sept. 15, 2013), *available at* http://www.bbc.com/news/world-middle-east-24100296; U.S. Dept. of State, Framework for Elimination of Syrian Chemical Weapons (Sept. 14, 2013), *available at* http://www.state.gov/r/pa/prs/ps/2013/09/214247.htm.

survive weak renegotiation refinement. On the contrary, it questions the weak renegotiation refinement criterion itself. Specifically, it does not consider the status quo ante as a viable alternative to war because the weak state might simply repeat the same pattern in the future. Therefore, it could be that waging war is preferable for S after all.

Using the example in Figure 2 again, that critique relates to any equilibrium path, t, within the set $[\underline{t}, \tilde{t})$.[77] When a provocation occurs and renegotiation begins, S knows that W shall not truly return to the status quo ante. It realizes that W may repeat the pattern of provocation and renegotiation in the future. This means, effectively, that what is really "on the table" from the strong state's standpoint is either to adhere to the SPE strategy, go to war, and suffer the post-war (lower) terror levels, or, alternately, to suffer from an effectively *higher* terror level than before. In other words, S contemplates whether to stick to the current SPE or move to a different one.

However, W has a primary incentive to avoid war. Thus, it is not expected to effectively increase the future terror level to the corresponding punishment path (or above it). This is because if it did so, war would become cost-effective for S. In fact, W is expected to increase the terror level to be marginally *below* the corresponding punishment path, keeping S marginally better off not waging a full-scale war.[78] In other words, the provocation-renegotiation pattern exhausts the maximal tolerance level of S, but never goes further than that. Consequently, even if the offer to return to the status quo is unrealistic, S is still unlikely to go to war because W would most likely impose terror just lower than the level that should trigger war in response. In these circumstances, even if S does not believe that the return to the status quo is feasible, going to war is not cost-effective for it. Thus, any terror level below $\tilde{t}$ remains implausible, this critique notwithstanding. This rationale, however, expires when terror level reaches $\tilde{t}$. Because $\tilde{t}$ underlies a strongly RPE, the weak state cannot repeat the strategy of "provocation and renegotiation" as before. If it does provoke S by intending to increase terror above $\tilde{t}$, it will suffer the consequences of a full-scale war; no renegotiation can help to avoid it.

We conclude, accordingly, that in any unequal conflict, the most plausible SPE is the strongly RPE. Therefore, the expected level of terror in equilibrium

---

77 In the range $[\tilde{t}, \bar{t}]$, renegotiation is not relevant ab initio. This critique relates to weakly renegotiation refinement, which as discussed, applies to cases where punishment of S by W is foreseeable. Namely, only relevant to the range $[\underline{t}, \tilde{t})$.

78 The idea is simple. For any $t \in [\underline{t}, \tilde{t})$ the benefits of war are $t + \frac{\delta}{1+\delta}\big(t_2(t) - t_1(t)\big) = \bar{t}$ where $t_2(t)$ and $t_1(t)$ are the punishment and reward level of terror that corresponds to t, respectively. Hence keeping the punishment just below $t_2(t)$ renders war cost-inefficient for S.

would be the lowest among all possible levels that are immune to weak renegotiation.

Still, one may argue that we have obtained a subjective ad bellum threshold rule. After all, $\tilde{t}$ depends on the discount factor of the strong state. Different states may have different discount factors, implying different states may be eligible to wage war under different levels of provocations. But why should such a conclusion be surprising? For example, a victim state with a developed civil defense infrastructure will suffer fewer casualties and less destruction compared to a less protected state facing the very same attack. Theoretically, this may lead to the conclusion that two states that are similarly attacked may experience different levels of damage, and exercise different levels of retaliation. Consequently, the ad bellum review of their response may be different.[79]

In fact, this differential outcome aligns with basic intuitions about the right of self-defense. Different levels of vulnerability or offensive capacity of the victim lead to different conclusions in determining whether the victim's response is indeed a legitimate instance of self-defense.[80] The requirements of imminence, last resort, and proportionality are all dependent upon the varying victim's vulnerability or offensive capacity.[81]

C. Violating the Threshold Requirement

The previous discussion yields two main conclusions. First, any SPE whose equilibrium path reflects a rather low terror level is not weakly RPE. We maintain that this is why the threshold rule in customary laws of armed conflicts is not intolerant. Second, insofar as there are multiple weakly RPEs, we argue that $\tilde{t}$, which constitutes the lower bounds of all strongly RPEs, is the most plausible candidate to quantitatively describe the ad bellum threshold rule.

Still, an SPE based on an intolerant trigger strategy, which can decrease terror below $\tilde{t}$, is better for the strong state. The only thing precluding that better outcome in our model is the possibility to renegotiate (and, presumably, not the threshold requirement doctrine). Nevertheless, successful consummation of renegotiation also depends on the level of transaction (or negotiation) costs.

---

79 In the military clash between Israel and Hamas in Gaza during the summer of 2014, Hamas launched more than 100 rockets and mortars a day for 50 days on Israeli civilians. Yet the anti-rockets system Iron Dome exhibited Israel's technological superiority, leaving almost zero casualties among Israeli civilians as a result of these rockets. Therefore, many classified Israel's strikes on Gaza as excessive using that argument exactly.

80 *See* Sloane, *supra* note 21, at 53.

81 In domestic law, it may seem reasonable that a physically challenged victim facing a young attacker may shoot him in self-defense, yet a highly capable athlete under the very same circumstances responding in the same way may not be permitted to claim self-defense.

Like any other negotiation, its efficient conclusion can be prevented if transaction costs are sufficiently high. Surprisingly, for the strong state, this is in fact good news. If transaction costs are indeed high enough such that successful renegotiation is impossible, then the intolerant trigger strategy and its underlying SPE is viable (that is, renegotiation-proof). In other words, if renegotiation is expected to fail, the strong state in unequal conflicts is expected to violate the ad bellum threshold rule by employing an intolerant trigger strategy (and carrying it out when needed).

Transaction costs with respect to renegotiation between conflicting states are not rare. Leaders may have a hidden agenda to go to war. They also may not necessarily have their country's best interests in mind (agency costs). Despite the fact that a full-scale war is best avoided when the status quo ante is Pareto superior to it, leaders from both sides may nonetheless exploit the situation for political benefit. In which case the original threat to carry out such wars is renegotiation-proof, and thus its underlying SPE is viable.[82]

This conclusion is far from trivial. It indicates, counterintuitively, that "bad" strong-state leaders who are determined to go to war may be the best candidates to decrease state-sponsored terror and potentially avoid this war. The fact that they ignore the rational limitations on war signals to the leaders of the weaker state that they might stand behind their threat to go to war even if this threat reflects an intolerant strategy—that is, even if it marks a very low level of terror as a red line. Effectively eliminating the option of renegotiation is actually a good thing for the strong side.

There are many examples of world leaders making this sort of seemingly biased decision. A clear one is the decision of the Argentinean Junta to invade the Falkland Islands in 1981. It is widely argued that this act was primarily motivated by the need to divert domestic unrest to an international crisis.[83] President Clinton was also accused of using the attack on Sudan and Afghanistan in August 1998 as a tool for distracting public attention from the effects of the Lewinski scandal. Similar accusations were made against President Bush, charging that the campaign against Iraq was not meant to address fears of weapons of mass destruction but rather predominantly to gain control of oil reserves or impose a new regional order.[84] Their truth or falsity notwithstanding,

---

82 On the interaction between domestic and international considerations of leaders, see Robert D. Putnam, *Diplomacy and Domestic Politics: The Logic of Two-Level Games*, 42 INT'L ORG. 427 (1988).

83 See, for example, Amy Oakes, Diversionary War and Argentina's Invasion of the Falkland Islands, 15 SEC. STUD. 431 (2006).

84 See David Hastings Dunn, Myths, Motivations and 'Misunderestimations': The Bush Administration and Iraq, 79 INT'L AFF. 279 (2003); James P. Pfiffner, Did President Bush Mislead the Country in His Arguments for War with Iraq, 34 PRESIDENTIAL STUD. Q. 25 (2004).

such accusations depict a situation in which renegotiation is expected to be fruitless and war illegitimate.

Indeed, the international community can sometimes force the leaders of the conflicting states to halt military actions and talk their way back to the status quo ante. This assumes, of course, that it has the power to influence reluctant leaders and is motivated to do so. The former relates, to some extent, to the coercive power of international law and international institutions in general. (The latter relates to the self-interests of these other states within the international community, which we do not address in this work.)

The India-Pakistan standoff in 2001 demonstrates the impact of an effective international community. Following an attack on the Indian parliament in December 2001, India deployed a large military force on its border with Pakistan accusing the latter of hosting and supporting the perpetrators' terror organization. The military tension—on the edge of a full-scale war—was eased within weeks of effective U.S. and E.U. diplomatic efforts. Clearly, the international community was motivated to take action and avoid a potentially catastrophic war between two nuclear superpowers.[85]

In November 2012, an active international community helped to successfully finalize negotiations between Hamas in Gaza and Israel, preventing hostilities from developing into a full-scale war and a massive Israeli incursion into Gaza.[86] During the 1990s, a similarly active international community prevented escalation into full-scale wars of the outbreaks of violence between Israel and Hezbollah, resulting in a repeated pattern of understandings.[87]

---

85 *See Pakistan to Withdraw Front-line Troops*, BBC NEWS (Oct. 17, 2002), *available at* http://news.bbc.co.uk/2/hi/world/south_asia/2335599.stm.

86 Reasons that the renewed cease-fire is expected to hold include "the fact that Netanyahu's credibility is on the line, that Hillary Clinton is involved, that the U.S. is involved," according to Aaron David Miller, a former U.S. Middle East negotiator and now vice president at the Woodrow Wilson International Center for Scholars in Washington. Jonathan Ferziger, Mariam Fam & Saud Abu Ramadan, *Gaza Cease-fire Success Hinges on Israel-Hamas Efforts*, BLOOMBERG (Nov. 22, 2012), *available at* http://www.bloomberg.com/news/2012-11-21/israel-hamas-agree-to-truce-starting-today-after-egypt-talks.html. "Hamas doesn't want a military incursion and doesn't want to jeopardize prospects of opening the border." *Id.* Unfortunately, in the summer of 2014, another collision occurred in the endless circle of violence. Israel, however, did not wage a full-scale war against Hamas and, after 50 days of violence, the parties returned exactly to the status quo ante, with the help of Egypt as mediator. Jodi Rudoren, *Cease-Fire Extended, but Not on Hamas's Terms*, N.Y. TIMES (Aug. 26, 2014), *available at* http://www.nytimes.com/2014/08/27/world/middleeast/israel-gaza-strip-conflict.html?_r=0.

87 See Reuven Erlich, Agreements, Arrangements and Understandings Concerning Lebanon to which Israel was Involved During the Past 30 years- Background, Data, Lessons and Conclusions, INTELLIGENCE AND TERRORISM INFORMATION CENTER AT THE CENTER FOR SPECIAL STUDIES, available at http://www.terrorism-info.org.il/data/pdf/PDF_06_235_2.pdf (last visited Feb. 25, 2015). A very similar pattern of collision and negotiation back to the status quo ante between Israel and the Hezbollah took place again in January 2015.

But this is not always the case. Even if the benefits of de-escalation are clear to third parties, each individual third-party state would rather not incur the costs of mediating—and certainly not actively buffering—the conflicting states (a classic collective action problem). These too are types of transaction costs, which may ultimately frustrate the negotiation between the conflicting parties. These are the circumstances in which the threshold rule is expected to be violated.[88] Once again, this conclusion is not trivial. It indicates that an international community determined to avoid war might unintentionally increase the level of terror. It creates an atmosphere that undermines the credibility of intolerant threats. These threats are not credible because both parties know in advance that they are not renegotiation-proof. Both parties know that the strong state will not execute its threat due to an expected renegotiation. An indifferent or an incapable international community undermines the chances of such negotiation being successfully concluded and, as a result, increases the credibility of the intolerant strategy of the stronger state.

The low incentive for the international community to interfere helped make Turkey's threat against Syria negotiation-proof, thus lowering the (Syria originated) terror level by the PKK virtually to zero. The U.S., Israel, and the Sunni Muslim neighbors of Syria would not have objected to Syria being taught an expensive lesson.[89] An anemic and unmotivated international community is also visible in the U.S. threats and invasions of both Panama and Grenada. These two cases demonstrate a combination of agency costs (it was argued that U.S. presidents had a hidden agenda for the invasions) and an international community that had neither the power nor the incentive to stop the world's strongest superpower.[90]

Diplomacy regarding North Korea similarly demonstrates the combination of an extreme agency problem alongside an ineffective international community. As expected, the provocations seem to work well for the highly manipulative regime that repeatedly violates international laws by threatening to attack its southern neighbor, among others.[91]

---

88 A collective action problem is not the only obstacle in this respect. For instance, if a strong third party is similarly situated in an unequal conflict of its own, it might be reluctant to force the parties back to the status quo. Because successful renegotiation prevents the strong party from sustaining low levels of terror, the third party, acknowledging that, might act strategically to set a precedent for its own benefit.

89 *See generally* Zisser, *supra* note 16 (providing a general analysis of the Syrian position against U.S.-Turkey-Israel).

90 *See generally* Ved P. Nanda, *The Validity of United States Intervention in Panama Under International Law*, 84 AM. J. INT'L L. 494 (1990); Schachter, *supra* note 51.

91 *See* Max Fisher, *Why North Korea loves to threaten World War III (but probably won't follow through)*, WASH. POST (Mar. 12, 2013), *available at* http://www.washingtonpost.com/blogs/worldviews/

Finally, a strong state should probably try to actively increase its transaction costs and even frustrate any possible international mediation in regard to a terror-sponsoring state. Clearly, it may succeed. Transaction costs are neither fixed nor exogenous. Endogenizing transaction costs and the decision-making of third parties, however, extends beyond the scope of our work. We have confined ourselves to discussing the consequences of exogenously given transaction costs, because our research interest is mainly to explain the tolerance of the threshold requirement as well as the conditions of its possible violation. While both turn on the level of transaction costs that renegotiation entails, they do not depend upon how their magnitude is determined.

VI. CONCLUSION

The contribution of this work is threefold. First, it shows that an intolerant trigger strategy can be credible and that credibility is thus not the reason for the relative tolerance of the existing ad bellum threshold rule. Second, it posits that a plausible reason for such rule is renegotiation perfection. Strong states normally refrain from adopting intolerant trigger strategies because such strategies are not renegotiation-proof, and thus have not evolved into a customary rule. Third, it describes the circumstances under which strong states shall deviate from the tolerant approach and violate the threshold rule. Specifically, if the transaction costs are sufficiently high, renegotiation is expected to fail, and the strong state is better off breaching the threshold rule.

These predictions can be tested empirically in future works. Clearly, the heart of our model is the ability to communicate and successfully negotiate, directly or indirectly, especially when the confrontation is on the verge of escalation. If renegotiation is expected to succeed, then strong states are expected to comply with the threshold rule. Subsequently, however, they must be more tolerant and thus suffer more terror. On the other hand, if renegotiation is expected to fail, then strong states are expected to act less tolerantly, in violation of the threshold rule, but to benefit from a lower level of terror. This insight provides a testable implication: the effect of transaction costs on the terror level suffered by strong states. Other things being equal, as transaction costs increase, the terror level from which a given state suffers should be lower.

Lastly, the theme of this work can also be extended to the law of treaties, challenging the conventional wisdom of compliance in that regard.[92] In any

---

wp/2013/03/12/why-north-korea-loves-to-threaten-world-war-iii-but-probably-wont-follow-through.

92 For a theory of international agreements, see generally Goldsmith and Posner, THE LIMITS OF INTERNATIONAL LAW, *supra* note 5.

bilateral treaty, the weak party that can be effectively sanctioned by the stronger one seemingly cannot violate the agreement at all. However, as this model predicts, costly sanctioning the weak side in response to mild violations is not renegotiation-proof. Arguably, therefore, the strong state's power becomes idle in such occasions. Stated differently, minor violations of the status quo are expected to occur insofar as the transaction costs of renegotiation are relatively low.

This work shakes some seemingly safe presumptions. Transaction costs reveal a positive effect. Obeying international law seems credible in the eyes of those terror-sponsoring states who, by definition, violate it. An effective international community that takes action to prevent escalation of conflicts actually increases the level of terror. A hidden agenda of leaders to resort to war at any cost may actually lead to a lower level of terror and even to avoidance of war. And maybe, this is only an expression of what a deeper intuition tells us: that being strong may be a source of weakness, and weakness may sometimes be a source of strength.

VII. APPENDIX

Assume a sequential confrontation game in which state W chooses the level of terror damage to be suffered by state S, $t \geq \underline{t}$, and S, observing W's actions, chooses whether to contain the imminent terror damage or to wage full-scale war, $w = \{0,1\}$, respectively. State W always seeks to avoid a full-scale war, but at the same time, to keep t as high as possible. State S, on the other hand, prefers t to be as low as possible. The cost of war for S is denoted by $\bar{t}$, where $\bar{t} > \underline{t}$.

The only SPE is: $\{\bar{t};\ w = 0 \text{ if } t \leq \bar{t} \text{ and } w = 1 \text{ otherwise}\}$. Infinitely repeating the single-shot game, however, suggests that any $t_0 \in [\underline{t}, \bar{t}]$ can be supported in SPE:

1. W plays t_0; S plays $w = 0$ if $t \leq t_0$ and $w = 1$ otherwise,
2. If $w = 1$ replace t_0 with $t_1(t_0) \leq t_0$ and repeat stage 1 ("punishing W"),
3. If $t > t_0$ and $w = 0$ replace t_0 with $t_2(t_0)$ and repeat stage 1 ("punishing S"),

The conditions for these strategies to form an SPE are:

$t_0 + \frac{\delta}{1-\delta}(t_2 - t_1) \geq \bar{t} \geq t_0 + \frac{\delta}{1-\delta}(t_0 - t_1)\ s.t\ t \geq \underline{t}$, where δ is S's discount factor.

The middle term is the costs of war, which must be sufficiently high to prevent unprovoked war, namely, when $t \leq t_0$, but sufficiently high to encourage war upon provocation, namely, when $t > t_0$. The terms on the right and on the left denote the present value of the benefits of unprovoked and provoked war, respectively. For simplicity, assume that both terms hold in equality, subject to $\min t_1 = \underline{t}$ (thereby we focus on the SPE presented in Figure 2).

Note that by comparing the left side to the right side, $t_2 \geq t_0$. Further, the left side implies that $t_2 = t_0 \Leftrightarrow t_0 \geq \tilde{t} \equiv (1-\delta)\bar{t} + \delta\underline{t}$ (since $\min t_1 = \underline{t}$). Therefore, any SPE in which $t_0 < \tilde{t}$ entails that $t_2 > t_0$.

Any SPE wherein $t_2 > t_0$ is not weakly RPE: S will not punish W for deviating and W will not punish S for not punishing it. Both parties prefer the equilibrium path over the continuation equilibrium that requires them to engage in a full-scale war. Hence, in any weakly RPE, it must be that $t_2 = t_0$, which implies that the set $[\tilde{t}, \bar{t}]$ support the weakly RPE. In which case, only $\tilde{t}$ is a strongly RPE; for any other weakly RPE, S would be better off not punishing W for deviating, and, in exchange, W will decrease t to some lower weakly RPE.