How Does Reciprocity Work? Evidence from the Laws of War

by James D. Morrow Department of Political Science University of Michigan Ann Arbor, Michigan 48103

email: jdmorrow@umich.edu

Please do not cite or quote without the author's permission.

Abstract

Reciprocity is a central concept in much of international relations, particularly international law. The idea that states should respond to one another in kind is deeply ingrained in how we think about international politics, on topics from the use of force to trade liberalization. Empirical work has shown broad patterns of reciprocity in state interactions (e.g. (Goldstein and Freeman 1990)). Agreements to cooperate are generally believed to be enforced by threats of reciprocal retaliation against defections from the agreement. International institutions, including international law, are believed to aid in reciprocal enforcement in a variety of ways.

Despite this general belief about the logic and role of reciprocity in international politics, there are important gaps in our understanding of it. First, the varieties of reciprocity are not always appreciated. The game of iterated Prisoners' Dilemma game is often used to illustrate the logic of reciprocity; responses are direct, immediate, and proportionate to the defection from the agreement to cooperate. This form of reciprocity can support full cooperation between the parties by deterring violations with reciprocal sanctions. However, this form of reciprocity also requires that the parties know perfectly what actions each other has taken. When the parties have difficulties monitoring the compliance of each other to their agreement, the nature of reciprocity changes. It becomes irregular and disproportionate to perceived violations. Then difficulties in monitoring an agreement—often referred to as "noise"—changes the nature of reciprocity. The presence of noise also means that the level of cooperation that can be sustained decreases below full cooperation. Uncovering evidence of reciprocity then requires us to assess the level of noise present in the situation.

Second, reciprocity requires a shared understanding about how an agreement to cooperate

-1-

will be enforced. This shared understanding allows the actors to anticipate what actions will be viewed as violations and what responses are appropriate when. Because violations and reciprocal responses are the same acts, the actors need to be able to share an understanding allowing them to which acts are violations requiring a response and which are reciprocal responses to which there must not be a response. This shared understanding plays a larger role when noise is prominent in the parties' behavior. The difficulty in discerning what exactly one another has done leads parties to rely on agreed public signals of when behavior has gone "too far" and hence a reciprocal response is warranted. These standards are often referred to as "bright lines" where retaliation only occurs if that line is crossed. Among the roles of international law is the specification of such "bright lines" of conduct, describing the conditions to judge behavior as acceptable or not.

Third, we have little reproducible evidence on how reciprocity works in practice, particularly in light of the two points above. Oye (1986) (Oye 1986) provides case studies of reciprocity in practice on a number of issues through international relations. Goldstein and Freeman show complex patterns of reciprocity in the diplomatic and political actions of the United States, Soviet Union, and China (Goldstein and Freeman 1990). However, neither of these studies address how noise affects reciprocity and how international law can be used to address the difficulties that noise poses for reciprocal enforcement of agreements.

This paper addresses all three points above by examining how reciprocity has worked in practice in the laws of war. This body of international law, also known as international humanitarian law, seeks to regulate how militaries conduct themselves during wartime to limit the destructiveness of war. Compliance with the laws of war is only possible through reciprocity

-2-

and the willingness of warring parties to live within as the parties have no recourse to a stronger sanction. Further, the level of noise varies across the different issues covered by these laws, from protection of civilians to restrictions on the use of chemical and biological weapons. There is also variation across issues in the level of legalization, as I examine areas with treaties in force and an issue, aerial bombardment, where no treaty has ever entered into force despite several efforts to draft a treaty to address that issue. States also vary in their legal obligation to these treaties as not all warring states ratified all treaties in force at the time of the war.

I report the results of statistical analyses conducted on a data set I have collected on compliance with the laws of war during the 20th century. These analyses show that the level of compliance with the norms embodied in the laws of war declines with the level of noise introduced by violations committed by individual combatants. There is evidence for reciprocity in the data. Law has little direct effect on compliance; states legally bound are not more likely to comply with their obligations. However, legal obligation does produce restraint in states, particularly the stronger side, that otherwise would be more likely to commit violations. Additionally, the "bright lines" of conduct produced by the treaties does matter as clear, reciprocal responses are more likely to clear legal violations than those who illegality is in question.

I begin with a brief discussion of the logic of reciprocity drawn from game-theoretic models of reciprocity. This discussion focuses on the consequences of noise for reciprocity in practice. I then describe the data collection used to test those arguments. I present the results of a series of tests of hypotheses concerning reciprocity and noise and conclude with a discussion of the implications of those results.

-3-

Strategic Theories of Reciprocity

Iterated Prisoners' Dilemma has often be used as the starting point for thinking about reciprocal relationships. For the purposes of this discussion, consider a version of Prisoners' Dilemma where the two players choose a degree of cooperation rather than the usual dichotomous choice between Cooperate and Defect. Specifically, each player chooses a number between 0 and 1 inclusive, say *x*, which denotes the player's level of cooperation in that round of the game. The player pays a cost of *-cx* for this effort which produces a benefit of (c+1)x for the other player. If Player 1 chooses *x* and Player 2 *y*, Player 1 receives (c+1)y - cx and Player 2 (c+1)x - cy. If both cooperate fully (x = y = I), both receive a payoff of 1 for the round; if neither cooperates at all (x = y = 0), both receive 0; if one cooperates fully while the other does not cooperate at all, the former receives *-c* and the latter c+1. This game has the familiar strategic logic of Prisoners' Dilemma. Each player does better in this round by not cooperating at all regardless of what the other does, but both are better off if they both fully cooperate. The question is how can they enforce an agreement to play Cooperate in the face of their individual incentives not to do so in the short run.

If the players play the game repeatedly, they may be able to enforce an agreement to cooperate fully through reciprocity. The simplest idea of reciprocity is the well-known strategy of Tit-for-Tat (cf. (Axelrod 1984)); each player cooperates fully in the first round (x = y = 1) and continues to do so until either of them does not do so (x < 1 or y < 1). Then both match the lowest level of cooperation in the previous round for one round and then resume their agreement to play Cooperate after that round. Assuming that they discount future payoffs by a common

-4-

discount factor of * < 1 for each round, each will prefer honoring their agreement to play Cooperate in a round to playing Defect when

$$1 + \delta(1) > c + 1 - cx + \delta(x)$$

$$\delta > c \tag{1}$$

In words, the discounted loss from the reciprocal punishment (1 - x where x is the level of cooperation in the round of the violation), must exceed the short-run benefits of breaking the agreement, c(1-x). If this is not the case, the players may still be able to enforce an agreement to cooperate with a longer period of reciprocal punishment.¹ Reciprocity here is direct, immediate and proportional; the response is certain, comes quickly, and matches the violation. Robert Keohane ((Keohane 1986)) describes this form of reciprocity as specific.

This simple form of reciprocity requires that the players share an understanding that they will enforce their agreement through reciprocity. That agreement is necessary for them to interpret one another's actions correctly; reciprocal punishments are the same actions as violations of the agreement. It is only the shared understanding that acts of less than full cooperation are sometimes violations and other times compliance with the agreement. Further, the length of such punishments must be agreed to as well to avoid mistakes of interpretation of each other's moves.

The world is rarely as clean as this simple model. The players may not know exactly what one another has done. Specific reciprocity raises a dilemma in such a situation because I may be mistaken when I perceive that you have taken actions that are violations of our agreement. If I retaliate against you when you have not broken our agreement, you may see my action as a defection from the agreement, raising the possibility of a retaliatory spiral. Because

-5-

reciprocal retaliation imposes costs on both of us through a loss in the benefits of cooperation, I might not want to respond every time I believe that you have broken the agreement. This unwillingness to respond to possible violations could undermine the deterrent effect of reciprocal sanctions and actors' ability to enforce agreements to cooperate.

We can represent this inability to know exactly whether one another has complied with the agreement by adding "noise" to the model. Imagine that some random error is added to each player's level of cooperation after they have chosen their moves but before payoffs are revealed. Now each player knows when he or she has not received the full benefits of cooperation but not whether that loss is the consequence of the other player's action; the "noise" of the random error has made it impossible for the players to know exactly what one another has done. The players may still be able to use reciprocity to support some level of cooperation provided that they share a public signal of the current state of affairs. We can then condition our behavior on that public signal. George Downs and David Rocke ((Downs and Rocke 1990), 176-92; (Downs and Rocke 1995), 87-127; the original source is (Green and Porter 1984)) describe reciprocal strategies based on triggers for problems of cooperation under noise. The players set a trigger level and then ignore small perceived violations which do not breach that threshold but respond with complete noncooperation if the level of cooperation falls below that level. The precise level of the trigger weighs the two types of error possible. Lowering the trigger level reduces the deterrent effect of breaching that threat, and both sides will lower their cooperative behavior out of self-interest. Raising the trigger level increases the risk of the random error pushing the level of cooperation below that level, leading to a period of noncooperation in response. Reciprocity under noise becomes indirect and disproportionate; the players do not respond to any violations

-6-

of full cooperation, only large deviations from cooperation; when they respond, their response is much larger than the offense that triggered it. Reciprocity begins to resemble Keohane's (Keohane 1986) diffuse reciprocity.

Noise reduces the level of cooperation that can be maintained through reciprocity ((Bednar 2005), (Kreps 1990), 526-31). Because noise raises the possibility of inadvertent retaliation, agreements to cooperate will break down at some times. Further, the "wiggle room" between full cooperation and the trigger leads to less than full cooperation. Both of these issues increase as the level of noise increases. The greater the scope for inadvertent defections and misunderstood actions, the lower the level of cooperation we should expect.

Institutions can play a role in supporting reciprocity as a means to enforce cooperation. As mentioned earlier, all of these reciprocal strategies require that the players understand that they are playing them so they can interpret one another's actions properly. The trigger strategies for reciprocity under noise require "bright lines" to be drawn between acceptable behavior and violations that require a response. One role of international law as an international institution is the creation of such "bright lines." Formal treaties play an important role because they set a single standard of conduct which states accept publicly when they ratify the treaty in question. Ratification then signals acceptance of the standard and its enforcement. Law also sets principles that parties could use to reason through cases in dispute.

Reciprocity then is more complex in practice that the simple model of Tit-for-Tat. This paper examines reciprocal patterns of behavior in compliance with the laws of war during the 20th century. Treaty law here establishes rules of conduct in war, but the application of those rules is left to the parties themselves. The question then is how does reciprocity work in the law

-7-

of war?

The Laws of War as a Test of Reciprocity

The laws of war, or more properly international humanitarian law, seeks to regulate conduct by warring forces to limit the destructiveness of war. The treaties cover a wide range of possible conduct during wartime from protection of civilians to conduct on the high seas. Detter (Detter Delupis 2000) provides an introduction to this body of law including issues of when war is legal and the scope of the applicability of the treaties.

These laws provide a test of arguments about how reciprocity can be used to enforce cooperation. As the parties are already at war when these laws are binding, they lack any larger sanction to use for enforcement. Reciprocity during wartime means that violations of the treaties are met in kind; such reciprocal enforcement is reverting to the absence of restraint on a particular area of the law of war. The parties then could observe some of the laws of war even while placing others in abeyance as a form of retaliation. The laws of war that regulate conduct during wartime can test theories of reciprocal enforcement of cooperation, while other portions of the law of war that address conduct outside of war, such as neutrality law, are not as useful for testing reciprocity because a wider variety of acts could constitute reciprocal responses to violations. For instance, the entry of the United States into the First World War was driven largely by unrestricted submarine warfare by Germany that the United States viewed as a violation of the neutral rights of U.S. shipping. Similarly, examining parties actively at war avoids the difficulties of capability to retaliate posed by law of war that holds after one state has

-8-

been defeated, such as law of occupation. Those cases make a poor test for reciprocity because the citizens of the defeated state may not possess the capability to retaliate should the occupying power violate the relevant treaty standard.

Reciprocity could also occur indirectly through the actions of soldiers on the battlefield independent of an explicit state decision to violate or retaliate against a perceived violation ((Morrow 2001)). The laws of war create obligations for individual soldiers in addition to states that accept the laws. Individual soldiers can commit acts that can be violations or retaliation against acts of enemy soldiers. Studies of battlefield behavior ((Fritz 1995; Holmes 1986; Linderman 1997)) show that soldiers do respond in kind to atrocities committed against their comrades and believe that the other side is likely to respond to atrocities they and their comrades commit. In some situations, such reciprocity operates at the level of retaliation against specific identifiable enemy units which have been responsible for atrocities ((Linderman 1997), 119-21, 137-39). The laws of war then can be enforced at the individual level in part through reciprocity among soldiers on the battlefield. These individual acts of retaliation can produce reciprocity at a broader level than individual units if a spiral of retaliations leads to a general breakdown of a treaty standard (cf. (Fearon and Laitin 1996) for a model of such retaliatory spirals in ethnic conflict). Standards for treatment of the wounded and prisoners on the battlefield broke during the Second World War on both the Eastern Front in Europe and in the war in the Pacific between the United States and Japan ((Fritz 1995; Linderman 1997)).

The scope for violations by individual soldiers varies with the issues covered by different treaties in the laws of war. For issues like the use of chemical and biological weapons, individual solders lack the ability to commit violations on their own as they can use such

-9-

weapons only if they are given them by their command authority. At the other extreme are issues like treatment of the wounded, prisoners, and civilians on the battlefield where individual soldiers always possess the ability to commit violations on their own. In between these extremes are issues like conduct on the high seas and aerial bombardment where officers in command may be able to commit violations on their own but individual soldiers do not.

This variance in the scope for individual violations provides a way to test the role of noise in reciprocal enforcement. Individual violations are an important source of noise in determining whether the other side is complying with their legal obligations. In the first months of U.S. participation in the Second World War in the Pacific, U.S. soldiers learned through experience that Japanese forces would not only fight to the death rather than surrender but they would also exploit the protections of the laws of war to gain advantages on the battlefield through playing dead, attacking medical corpsmen, feigning surrender and other acts of perfidy ((Linderman 1997), 151-61). Because states rarely publicly declare their intention to violate a law of war, they must infer the conduct of the other side from the behavior they observe on the battlefield and from reports of conduct behind enemy lines. Individual violations complicate these inferences of intent by adding violations that may not be intentional state policy. This is the problem of reciprocal enforcement in the face of noise as discussed previously. Because the scope for individual violations varies with the specific issues of the laws of war, we can test the hypothesis that the level of cooperation should decline across issues as noise becomes more prominent. Put simply, those issues which pose the greatest scope for individual violations should be those with the least compliance.

We can also test for the role of formal treaty law in reciprocal enforcement because there

-10-

is some variation on which warring states have ratified which treaties. Although most states have ratified most of the relevant treaties, there are states that have not done so. Additionally, I include one issue–aerial bombing–where no treaty has ever been signed, much less entered into force through ratification. Draft treaties addressing restrictions on aerial bombing were negotiated during the 1930s, providing us with explicit standards of what conduct would qualify as unacceptable. The variation in treaty ratification across states and issues allows us to test whether legal acceptance of a standard affects behavior, and if so, how.

Reciprocity is not the only possible explanation for why states might comply with the law of war; they might choose to do so regardless of what their enemy does out of a sense of legal obligation. Constructivists (e.g. (Finnemore and Sikkink 1998)) argue that actors internalize norms over time, complying with them because they see the conduct prescribed by those norms as proper and correct. Treaties in this view could still play a key role by precisely describing what conduct was proper and what acts unacceptable as a guide to actors who wished to comply but might be unclear about how they needed to act to "do the right thing." Moral principles clearly undergird the laws of war (e.g. (Price 1997)), and those treaties can be seen as the legal codification of standards of humanity during wartime. What exactly this view predicts about patterns of compliance is unclear; a claim that actors should always comply seems impossibly hard to meet in general and is certainly not true in the case of the laws of war. Cases where one did not retaliate against extensive violations by the other side even though they had the ability to do so would support internalization over reciprocity. The frequency of such cases are a useful piece of evidence in comparing reciprocity to internalization as competing explanations of compliance.

-11-

In summary, patterns of compliance with the laws of war could provide evidence on whether and how reciprocity operates in practice during wartime. First, do states meet violations with violations and compliance with compliance? Second, does compliance decrease as the role of noise through violations by individuals increase? Third, do treaties increase compliance through either legal obligation or by clarifying what responses are appropriate after unacceptable behavior?

The Data

This paper reports the results of analyses of the patterns of reciprocity in compliance with the laws of war. The time period of the study is the 20th Century because the formal body of international law of war begins with the Hague Convention of 1899 and grows with the Geneva Conventions and other treaties regulating conduct between warring parties. The basic unit is the directed warring dyad-issue area. The question is what leads to compliance, with reciprocity as one mechanism leading to compliance.²

First, we take all Correlates of War interstate wars from the Boxer Rebellion to the Gulf War (1991, not the sequel). Each multilateral war is broken into all warring dyads by pairing off each member of each side with every member of the other side. Additional research determined whether military action occurred between the members of each of the possible warring dyads. When neither state in the dyad engaged in military action against the other, the dyad is dropped. For example, World War I expands to a full set of 44 dyads matching each of the 11 states that were members of the Allies with the 4 states of the Central Powers. From this set, dyads such as United States-Bulgaria are dropped because they did not actually fight one another.

The set of dyads is reduced further by consolidating states that fight under unified command into one actor. This process eliminates overcounting of non-independent observations because states that fight under united command have a single leader or leadership group that has the power to order subordinate units to comply. In this sense, states that fight under unified command do not have separate policies. Additional research was conducted to determine when such unified command existed. For example, all dyads in World War I between Portugal and the various Central Powers are absorbed into the corresponding dyads with Great Britain because Portuguese forces fought under British command.³ This consolidation also eliminates some cases where it may be difficult to determine if the two states in question actually fought as one of them fights under the command of another state.

The period of fighting for a particular warring dyads may differ from the general dates of the war, and so each is dated from the beginning of military action until fighting ends by agreement.⁴ For example, the Netherlands and Germany fight one another from May 10, 1940 to May 14, 1940 in the Second World War. States that reentered World War II are dropped (e.g. Vichy France); consolidation under unified command makes this distinction less important. Each warring dyad then leads to two directed dyads. For instance, Germany and France fought against one another in the First World War, giving rise to the directed dyads of Germany Y France and France Y Germany.

For each warring directed dyad, we then code behavior of the first member toward the second member on nine different issue-areas in the law of war. The issue-areas are as follows:

- Aerial bombardment

-13-

- Armistice/Ceasefire
- Chemical and biological weapons
- Treatment of civilians
- Protection of cultural property
- Conduct on the high seas
- Prisoners of war
- Declaration of war
- Treatment of wounded

Each of these issue-areas is defined by the set of treaties, including draft treaties, in the issuearea. We used the text of treaties found at the website of the International Committee of the Red Cross.⁵ The treaties found there were grouped into the nine issue-areas above. Issue-areas such as neutrality law were dropped because they do not address the conduct of one warring party toward another. Genocide was also dropped as acts that could be considered genocide during wartime were subsumed under the treatment of civilians. This selection of issue-areas encompasses both areas with well-developed treaty law as well as those which lack any formal treaty law, such as aerial bombardment. This design allows a test whether the existence of a formal treaty aids compliance with the norms of proper conduct in an issue-area. In the analysis that follows, we include 8 of the 9 issue-areas, dropping declaration of war because there is no possibility of reciprocity on that issue. If one side begins the war with a clear attack, whether declared or undeclared, the other side is coded as fully complying with its obligation to declare war, making reciprocity impossible.⁶

The relevant treaties in each issue-area were read to identify major and minor violations.

These coding rules structured the collection of information on violations and compliance for each directed warring dyad. Historical works and contemporary journalistic sources were searched for examples of violations and general judgments on degree of compliance by each warring party toward the other member of a directed dyad. Once files of the information had been assembled, we sought to code the following dimensions of compliance:

- Magnitude: how bad were the violations?

A four-point scale from none (coded 1) to many major violations such that compliance does not matter (4).

- Frequency: how frequent were violations?

A four-point scale from none (1) to massive violations to the point where the standard is ignored (4).

- Centralization: what was the role of central military and political authorities concerning violations?

A five-point scale from no violations (1) to central authorities punishing individual violators (2) to positive identification of state intent to violate (5).

- Clarity: did the actions clearly violate the treaties?

A four-point scale from no violations (1) to definite legal violation (4).

When a state commits violations, we also attempt to determine the date of first violation. This breakdown of compliance into four dimensions is designed to make the coding more reliable than a single scale of compliance. Each dimension can generally be coded when we have available evidence on the acts.

This design does not allow us to test reciprocity of actions directly because the codings are judgments about compliance across the entire period of fighting. Ideally, one would like to have a complete list of all instances of violations where one could trace the patterns of reciprocal responses directly. Although such sequences might be found in some cases, it is impossible to find them for even a notable set of cases. Understandably, those who commit atrocities often attempt to conceal their own participation in them, meaning there is never any record of many violations. Charges that the other side has committed atrocities are also common; sorting out what actually happened can be difficult. We prefer secondary historical sources, particularly academic works written decades after events, because the authors of such sources have often done the difficult work of separating truth from unfounded accusation. They also often make judgments about the nature and extent of violations from an examination of many incidents.

Quality and coverage of the data are key questions facing any analysis on this topic for two reasons. Violations are not possible for some issue-areas in some wars. For example, conduct on the high seas was not an issue in the Hungarian-Allies war of 1919 because all fighting took place on land. Missing data is an immense problem for any comprehensive study of compliance with the laws of war. Atrocities are often not reported. Some of these wars are obscure, meaning little or no information is available. Because the amount of information available to use in coding varies greatly from observation to observation, each directed-dyad-

-16-

issue area is also coded the quality of the data used in the coding. A fifth variable was then collected for each observation that was coded:

Quality of the data: do we have confidence in the coding because it is based on substantial and reliable information?
A four-point scale from sketchy evidence (1) to excellent documentation providing strong confidence in coding (4).

The score for the quality of the data is used in some analyses to place greater weight on the cases where we have confidence in the coding.

Given the amount of directed-dyad issue-areas where we had no evidence on which to base a judgment, standardized codings were employed for the following issue-areas: treatment of civilians, cultural protection, conduct on the high seas, prisoners of war, and treatment of the wounded. These standardized codings reflect my view that even the best disciplined armies commit some violations.⁷ These standardized codings are superceded if any information is available. In the analysis, a warring-dyad-issue-area will be dropped from the analysis if both sides have a standardized coding on that issue. Otherwise, the results could be biased in favor of reciprocity because a large number of cases would have identical standardized codes, inflating the appearance of reciprocity. These standardized codings allow me to use warring dyads where I have information only about what one member of the dyad did. The standardized codings are given data quality 0, which matters when I weight analyses by the quality of the data.

A standardized coding is also used for the issue of chemical and biological weapons

when we had no reports of use by a side. Accusations of the use of such weapons have been extensively investigated by others (e.g. (Harris and Paxman 2002; Stockholm International Peace Research 1971)), allowing us to conclude that a side did not use such weapons if the sources that focus on chemical and biological warfare do not mention that it used such weapons in that war. For the issue of chemical and biological warfare, the data set coded a warring party as having no violations (magnitude = 1 and frequency = 1) if there are no reports in these sources. Data quality is rated as a 2 for such cases.

Finally, I would like to make clear what the data set is not. As mentioned above, it is not a comprehensive listing of all violations for a given warring directed dyad-issue-area. Such a comprehensive listing would be wonderful for testing the dynamics of reciprocity, but it is impossible to collect for even a small set of the cases. Second, the data is not based on a precise legal analysis of whether particular acts constitute violations of the treaty in question. The legal status of some acts are contested, particularly when questions of military necessity and proportionality arise. Instead, the codings capture whether the broad pattern of acts by a warring party are consistent with the standards of the relevant treaty. When such acts are not agreed to be clear legal violations, the score for the legal clarity of the violations reflects that uncertainty. At the level of aggregation of the data, precise legal analysis of all acts is not necessary to make broad distinctions between behavior that is compliance and that which is not. These limitations are important to remind us what we can learn from this data and what we cannot. The data can help us see broad patterns in how reciprocity has worked in practice, even if it cannot illuminate the full details of such reciprocity.

Data Analysis

The variables in the data set are defined for the ease and clarity of collection and coding from the historical sources used. We seek to analyze when compliance occurs and whether and when reciprocity influence compliance. To do so, we need a measure of compliance drawn from the variables collected. A simple measure of compliance multiplies the scores for magnitude of violations and frequency of violations. This score then ranges from 1, no violations, to 16, frequent and multiple major violations. However, these scores should not be treated as a ratiolevel variable because equal intervals in the score do not necessarily indicate equal changes in the level of compliance. It is more appropriate to treat compliance as the product of magnitude and frequency of violations as an ordinal variable. We can say clearly that some cases show lower levels of compliance than others without asserting that equal differences between scores represent equal differences in compliance. The ordinal scale of compliance has five levels as follows:

- < Full compliance, where no violations are reported (magnitude = 1 and frequency =1),
- < High compliance, where only minor violations are reported even if they are frequenrt (magnitude = 2)
- < Partial compliance, where infrequent major violations are reported (magnitude = 3 or 4 and frequency = 2),
- < Low compliance, where major violations occur frequently but the standard is not ignored (magnitude > 2 and frequency > 2, but at least one of them < 4)
- < Noncompliance, where violations are both major and frequent (magnitude = 4 and

frequency = 4)

This ordinal scale of compliance should be robust in its separation and order of categories even if we cannot say that the differences in compliance between the levels are comparable.

I begin with some contingency tables to show the broad patterns in the data before turning to multivariate analysis of compliance. Table 1 reports the pattern between the compliance of both sides for each warring dyad-issue-area. To place each warring dyad in this table only once, I place the compliance of the less compliant side on the rows of the table and the compliance of the more compliant side on the columns. This procedure eliminates doublecounting of cases. Each warring dyad produces two observations in the data set for a given issue, and including both observations in Table 1 would inflate the appearance of reciprocity in the table.

Table 1 about here

Table 1 shows a reciprocal generally with some exceptions. The compliance of the two sides are positively related, and the two sides match compliance in almost exactly one-half of the cases (267 out of 533). Looking at the detail of the table, when one side is noncompliant, the other side almost always commits some major violations (partial compliance at best, 52 out of 54 cases of noncompliance by the less compliant side, the bottom row of the table). When the more compliant side commits no major violations (high compliance at worst), the other side commits frequent major violations only 9% of the time (26 out of 297 cases in the first two columns). Cases where the sides differ greatly in their compliance are rare, suggesting that reciprocity is generally present.

The role of noise in reciprocity can be seen by the variation in compliance across issue-

areas. Those issues where individual soldiers have greater scope to commit violations on their own should have lower levels of compliance because the noise makes maintaining compliance through reciprocity more difficult. Table 2 presents means, standard deviations, and medians for compliance divided by issue. Compliance is on a scale of 1 to 5 as described above with 1 meaning full compliance and 5 noncompliance. Higher scores in Table 2 then mean less compliance. To be careful, the ordinal nature of the compliance measure could make the means reported deceptive, which is why I report medians as well. The pattern of compliance across issues is that chemical and biological warfare is the issue with the greatest compliance and treatment of civilians the issue with the least compliance. There are then two groups in-between with aerial bombardment, armistice, and conduct on the high seas have higher levels of compliance on average than protection of cultural property, prisoners of war, and treatment of the wounded. For now, I note this pattern and hold off on a discussion of its significance until I show that it holds even when we control for reciprocity and other effects in a multivariate model.

Table 2 about here

Multivariate analysis allows us to analyze the effects of many variables on compliance at once to judge their effects of each controlling for others. Such analysis of reciprocity faces the issue of simultaneity bias. If reciprocity is present, then each side's compliance depends upon the other side's compliance. To test for reciprocity between sides A and B, we includes B's compliance toward A on the issue in question in the equation for A's compliance toward B. Because B's compliance towards A is believed to depend on A's compliance toward B in part, we are effectively including A's compliance toward B as part of the right-hand side of the equation predicting itself. This simultaneity bias will inflate the size of the coefficient for the effect of B's

-21-

compliance towards A on A's compliance toward B. Instrumental variables provide a way to address the simultaneity bias at a cost. We create an instrument for the other side's compliance by regressing it against all the variables in the regression we would like to estimate and a set of variables not included in that regression but correlated with the other side's compliance. By construction, the instrument is uncorrelated with the other independent variables and the bias is eliminated from the estimate. The cost is that the instrument by definition is not as accurate a measure of the other side's compliance as the original variable. Effectively, the instrument includes error not present in the original variable, and so will not fit as well as that variable. The issue of lack of fit is aggravated here by the ordinal scale of the compliance measure as instrumental variable analysis, such as two-stage least squares, treats the included endogenous variable as a ratio-level variable.

I deal with this pair of methodological issues by running two flawed analyses. I first run an ordered probit using the compliance score of the other side directly. The coefficient for the other side's compliance, how we assess reciprocity in these models, should be biased upward. The second analysis creates an instrumental variable for the other side's compliance, but treats compliance as a ratio-level variable. This analysis addresses the simultaneity bias in our estimate of the effect of reciprocity at the cost of efficiency in our estimates of that effect. The hope is that the pair of analysis will provide similar results, and we can conclude that the patterns we find are not the product of this pair of methodological concerns.

Before presenting and discussing those two analyses, I turn to a brief discussion of the other variables included in these analyses. I include the POLITY democracy and autocracy scores for the side in question to assess the effect of the domestic political system of the country

-22-

in question on its compliance. Each of these is measured on a scale from 0 to 10. To measure the relative power of the two sides, I include the power ratio of the two sides corrected for distance to the battlefield and aggregated across actors with a unified command. I use the Correlates of War composite capabilities measure and calculate the proportion of the capabilities of the two sides that the side in question possesses. This measure ranges from 0 to 1, with larger figures indicating that the side in question is more powerful on the battlefield relative to its opponent. I assess whether the side in question is legally bound to uphold the current standard of conduct in that issue-area by whether both sides have ratified the most recent treaty in the issuearea when the war began. I interact joint ratification with the side's democracy score and the power ratio to see if legal obligation is conditioned by the domestic system of a state or the power relationship. I then include dummy variables for the issue-areas to test for differences in compliance across them; treatment of the wounded is the excluded category that serves as the baseline for comparison. I also include variables for whether the side in question initiated the war as judged by the Correlates of war project and the battle deaths per 1000 prewar population suffered by the side during the war. The latter variable provides a way to assess whether more intense wars lead to more violations and less compliance.

Table 3 reports the results of an ordered probit predicting compliance by one side from the compliance of the other side on that issue-area and the issue-area in question. Recall that the measure of compliance of the side in question, the dependent variable here, increases as the side in question complies less with the standard. Positive coefficients then signify lower compliance, and negative ones greater compliance. The scale is reversed so that the lowest category is full compliance, the middle is partial compliance, and the highest category is noncompliance.

-23-

Table 3 about here

Table 3 again shows the broad pattern of reciprocity in the data. Noncompliance tends to be met with noncompliance. Again, the coefficient here is likely to be biased upwards by simultaneity bias, so we should be cautious about concluding that we have definitive evidence in favor of reciprocity. I defer the discussion of the other results in Table 3 until after I show that they are robust when we use an instrumental variable analysis to address the issue of simultaneity bias.

Table 4 presents the results of an instrumental variable analysis to correct for the simultaneity bias. The pattern of results is the same as the ordered probit analysis in Table 3, providing some assurance that the evidence of reciprocity in the data set is not just the result of simultaneity bias. The coefficients cannot be directly compared across the models in these two tables because the ordered probit analysis treats compliance as an ordinal variable while the instrumental variable regression treats compliance as a continuous, ratio-level variable. Judging from the statistical significance of the coefficients, which reflect the precision of our estimates in the two methods, the effects are more difficult to assess precisely in the instrumental variable analysis. We should expect this because the instrument for the compliance of the other side is not as accurate a measure of its compliance as the original variable. Still, the pattern of results is robust across the two analysis, and I now turn to discuss the common pattern of results.

Table 4 about here

Both analyses show moderate reciprocal effects. Shifting the other side's compliance from full (scored 1) to noncompliance (scored 5) raises the violations of the other side roughly about one level, as from partial compliance to low compliance. The magnitude of the effect is

easiest to see from the instrumental variables estimation in Table 4; raising the noncompliance score by 4 raises noncompliance by .936 (= .234 * 4), just short of one level. The effect is harder to see in the ordinal probit analysis because it does not treat the five levels of compliance as equally distant. Here a shift of 4 in the compliance of the other side shifts the latent variable in the ordinal probit by 1.79 units. This change is large enough to shift the observed noncompliance up at least one level and sometimes two or three levels depending on the value of the latent variable based on the other exogenous variables.⁸ Reciprocity is present in the laws of war, but it is not the perfect matching of tit-for-tat reciprocity. Instead, reciprocity is part but not all of a warring party's decision whether to comply with a treaty standard.

The pattern across issue-areas reflects the different role of individual versus state violations. The omitted category is treatment of the wounded, so the coefficients on the dummy variables show whether noncompliance is more or less likely in the other issue-areas relative to treatment of wounded. The grouping of issues into a hierarchy of the likelihood of compliance is the same as we found looking at the average level of compliance by issue. Compliance is most likely on the issue of chemical and biological warfare. Treatment of civilians is the most likely to have noncompliance, with protection of cultural property, prisoners of war, and treatment of the wounded having the next worst records of compliance. The group of aerial bombing, armistice, and conduct on the high seas fall in between, having a worst record of compliance than CBW and better than the latter group of issues. This pattern holds then even when we control for reciprocity, indeed even when we bias the effect of reciprocity upward in our analysis in the ordered probit analysis.

This hierarchy of compliance across issues in the laws of war matches the scope of

-25-

individual violations across issues. Individual soldiers lack the ability to use chemical weapons on their own initiative; they can only use such weapons if their military and political leaders choose to use such weapons and distribute them to the troops. At the other end of the spectrum of compliance, every individual soldier possess the ability to commit atrocities against civilians simply by using his personal weapon to loot or murder. Further, civilians often lack the ability to retaliate against soldiers who commit such atrocities, removing an immediate risk of retaliation as a deterrent to such acts. Among the issues that fall between these extremes in their typical level of compliance, the group with higher levels of compliance-aerial bombing, armistice, and conduct on the high seas-are those where the decisions of central political and military authorities dominate over the scope for violations by individuals. Violations on these issues can be triggered by officers below the top ranks; ship captains can refuse to take in the sailors from enemy ships they sink, pilots can choose to bomb targets that are off-limits. However, most violations on these issues are the result of state policy to violate the relevant standard, such as unrestricted submarine warfare or indiscriminate bombing of cities. The other issues-protection of cultural property, treatment of POWs and enemy wounded-provide individual soldiers with the opportunity to use protected cultural locations for a military advantage, to kill enemy soldiers trying to surrender, and to refuse to treat enemy wounded. Unlike treatment of civilians, however, the threat of reciprocity from enemy soldiers on these issues is larger and more immediate.

Legal obligation has little effect on its own, but important indirect effects in advancing compliance. The coefficients for joint ratification, which obligates both parties to observe the standard, are positive, implying that states with legal obligations commit *more* violations that

-26-

those without. However, neither coefficient is statistically significant, meaning that we should place little confidence in this result. Instead, joint ratification ameliorates the effect of other variables that lead states to fail to comply with the relevant standard. According to the results in both analyses, democracies are *less* likely to comply with the relevant standard than other types of political systems. Legal obligations through joint ratification effectively eliminates this effect. Democracies then are more likely to commit violations when they have not accepted a legal obligation to restrain their acts during wartime. When democracies have a legal obligation, they are not more or less likely than other types of states to observe that standard. Legal obligation also reduces the effect of power on compliance. More powerful states are less likely to comply with the relevant treaty standard. Part of this effect is opportunity to commit violations; the stronger side is more likely to be winning on the battlefield and so more likely to be able to insulate those at risk from enemy acts against them. For instance, the United States has carried out aerial bombing campaigns against the enemy homeland in some form in all of its wars of the 20th century, but its own homeland has been safe from such campaigns in all those wars, in part because American power guaranteed that those wars were fought on other continents. Legal obligation reduces the tendency of stronger states to commit violations by close to two-thirds. In both of these cases, legal obligation works indirectly by restraining parties that would otherwise violate the standard more.

To conclude this section, I note in passing that initiators are less likely to comply, and that more costly wars have more violations and less compliance.

Discussion

-27-

I conclude by pulling together the results of the statistical analysis with the theoretical discussion of how reciprocity works in practice at the beginning of the paper. In theory, reciprocity can be direct and immediate, as in tit-for-tat, because both sides know what one another has done. In the laws of war, reciprocity is hard to employ because warring parties rarely declare their intent to commit violations, forcing them to make judgments about the intent and policy of the other side from what happens on the battlefield. Battlefields are chaotic, and soldiers commit atrocities in even the best disciplined armies. The noise produced by individual violations make inferences about when and how to retaliate difficult in practice. What we can expect and do find is that average levels of compliance vary with the scope of noise induced by individual violations on the battlefield. As a consequence, reciprocity is not fully effective in inducing compliance because it must be irregular in the face of noise concerning the policy of the other side.

International law still has a critical role to play in supporting reciprocity as a tool of enforcement. Bright lines of acceptable and unacceptable conduct aid the parties in anticipating when and how one another will act in response to what happens on the battlefield whether violations are deliberate state policy or not. Law and its obligations are not a substitute for reciprocal enforcement; they can aid it. This view limits what restrictions on violence during wartime can be realized as not all desirable restrictions can be enforced either through reciprocity in the face of noise.

The results cast doubt on arguments about how states will internalize norms of conduct over time. Almost all cases of noncompliance are met with some major violations by the side suffering from the former side's unwillingness to observe the relevant standard in any time or

-28-

way. Stronger parties and democracies-the actors where internalization of norms should be strongest-commit more violations when they are not legally obligated to observe such standards. If compliance occurred simply because actors came to accept norms as the only proper way to act, we would not see these patterns in the data. This is not to say that norms cannot grow in strength over time; indeed, this view puts increasing legalization as a key step in compliance by reinforcing reciprocity. Instead, it suggests that good will and shared views of proper behavior will not lead states to act nice apart from other consequences.

The argument concerning the central role of noise does suggest a different tie to internalization of norms. Military training which incorporates the law of war is one mechanism by which norms of proper conduct during wartime could be internalized.⁹ Soldiers then comply with the laws of war in this view because they have been trained to do so. In the view presented here, military training is a tool to reduce noise through the combination of instruction in the rights and responsibilities that the laws of war create and impose for individual soldiers with enforcement of those laws against individuals who violate them on their own initiative. Reducing noise reinforces the efficacy of reciprocity as a way to enforce standards, and instruction in the laws of war for soldiers supported by a system of military justice that seeks to identify and punish individual violators would reduce the noise that arises from individual violations.

Endnotes

1. I do not present the mathematical details of such longer punishments in the interest of conserving space.

2. Full details of the construction of the data set including coding rules will be available when the data set is published in the near future.

3. Although it might be more accurate to say the Portuguese ran under British command...

4. This is typically the surrender of one party to the other, but I also end fighting when one side is no longer capable of resisting the other even if there is no formal surrender document.

5. http://www.icrc.org/ihl.nsf/WebFULL?OpenView

6. Reciprocity is possible on declaration of war for cases where the war arises out of border clashes where it is not possible to determine which side launched the first clear attack of the war. By the coding rules, both sides will have similar scores for the magnitude–how large their military actions were–and frequency–how many such attacks were made–during the escalatory period when the war started.

	7. Standardized	l codings :	for fol	lowing issues	as indicated	l in table below:
--	-----------------	-------------	---------	---------------	--------------	-------------------

Issue Area	Frequency	Magnitude	Centralization	Clarity	Data
					Quality
Civilians	2	3	2	4	0
POWs	2	3	2	4	0
Wounded	2	3	2	4	0
High Seas	1	1	1	1	0
Cultural	2	2	2	3	0

8. This estimate comes from examining the cutoff points between the different categories of noncompliance reported at the bottom of Table 3. If the latent variable falls below the first cutoff point, 1.32, the case is predicted to be full compliance; the latent variable between 1.32 and 1.57, high compliance; 1.57 and 2.91, partial compliance, 2.91 and 4.07, low compliance, and above 4.07 noncompliance. A shift of 1.79 in the latent variable raises scores of the latent variable from -0.47 to 4.07 at least one level of violations.

9. For example, see Jeffrey Legro's (1995) arguments concerning the role of military culture–which he defines as beliefs about proper employment of forces–in producing cooperation on the laws of war.

References

Axelrod, Robert M. 1984. The evolution of cooperation. New York: Basic Books.

Bednar, Jenna. 2005. Is Full Compliance Possible? Conditions for Shirking with Imperfect Monitoring and Continuous Action Spaces. Ann Arbor, MI.

Detter, Ingrid. 2000. The law of war. 2nd ed. New York: Cambridge University Press.

- Downs, George W., and David M. Rocke. 1990. *Tacit bargaining, arms races, and arms control*. Ann Arbor: University of Michigan Press.
- Downs, George W., and David M. Rocke. 1995. *Optimal imperfection? : domestic uncertainty and institutions in international relations*. Princeton, N.J.: Princeton University Press.
- Fearon, James D., and David D. Laitin. 1996. Explaining interethnic cooperation. *American Political Science Review* 90 (4):715-735.
- Finnemore, Martha, and Kathryn Sikkink. 1998. International norm dynamics and political change. *International Organization* 52 (4):887-917.

Fritz, Stephen G. 1995. Frontsoldaten : the German soldier in World War II. Lexington, Ky:

University Press of Kentucky.

- Goldstein, Joshua S., and John R. Freeman. 1990. *Three-way street : strategic reciprocity in world politics*. Chicago: University of Chicago Press.
- Green, Edward J., and Robert H. Porter. 1984. Noncooperative Collusion under Imperfect Price Information. *Econometrica* 52 (1):87-100.
- Harris, Robert, and Jeremy Paxman. 2002. A Higher Form of Killing. New York: Random House.
- Holmes, Richard. 1986. *Acts of war : the behavior of men in battle*. 1st American ed. New York: Free Press.
- Keohane, Robert O. 1986. Reciprocity in International-Relations. *International Organization* 40 (1):1-27.
- Kreps, David M. 1990. *A course in microeconomic theory*. Princeton, N.J.: Princeton University Press.
- Legro, Jeffrey. 1995. Cooperation under fire : Anglo-German restraint during World War II, Cornell studies in security affairs. Ithaca: Cornell University Press.

Linderman, Gerald F. 1997. *The world within war : America's combat experience in World War II*. New York: Free Press.

Morrow, James D. 2001. The institutional features of the prisoners of war treaties. *International Organization* 55 (4):971-991

Oye, Kenneth A. 1986. Cooperation under anarchy. Princeton, N.J.: Princeton University Press.

Price, Richard M. 1997. The chemical weapons taboo. Ithaca: Cornell University Press.

Stockholm International Peace Research, Institute. 1971. *The problem of chemical and biological warfare; a study of the historical, technical, military, legal and political aspects of CBW, and possible disarmament measures*. New York : Humanities Press.

Crosstabulation of Compliance for Both Sides

		Full	High	Partial	Low	Noncompliance
	Full	135	-	-	-	-
Compliance by	High	15	8	-	-	-
Less Compliant	Partial	90	23	78	-	-
Side	Low	14	10	66	40	-
	Noncompliance	2	0	22	24	6

Compliance by More Compliant Side

Note: cells above the diagonal left empty by design of table.

Summary Statistics of Compliance by Issue

Issue	Mean Compliance	Standard Deviation	Median Compliance
Aerial Bombing	2.17	1.18	2
Armistice	2.15	1.14	2
Chemical and Biological Weapons	1.39	0.98	1
Treatment of Civilians	3.46	0.96	4
Protection of Cultural Property	2.56	0.90	2
Conduct on the High Seas	2.28	1.35	3
Prisoners of War	3.08	1.21	3
Treatment of Wounded	2.91	1.05	3

High scores indicate lower levels of compliance

Variable	Estimated Coefficient	Standard Error	Significance Probability
Victim's Level of Noncompliance	.448	.034	<.001
Both Sides Ratified Treaty?	.279	.176	n.s.
Democracy Score of Violator	.065	.026	.013
Democracy Score of Violator times Joint Ratification	064	.019	.001
Autocracy Score of Violator	.019	.028	n.s.
Power Ratio	.884	.180	<.001
Power Ratio times Joint Ratification	569	.263	.031
Aerial Bombing	579	.146	<.001
Armistice	527	.175	.003
Chemical and Biological Warfare	-1.213	.153	<.001
Treatment of Civilians	.400	.130	.002
Protection of Cultural Property	164	.165	n.s.
Conduct on the High Seas	418	.154	.007
Prisoners of War	.132	.130	n.s.
Violator Initiated War?	.314	.076	<.001
Violator Battle Deaths per 1000 Population	.023	.0027	<.001
N = 1066 Log-Likeliho	pod = -1133.7	$P^2 = 664.4 \text{ w}/16$	d.f.

Ordered Probit Analysis Predicting Noncompliance with the Laws of War

Significance Probability of Model < .0001 Pseudo R² = .227

The omitted category for the dummy variables of the issue-areas is treatment of the wounded.

Constant is set to 0 to identify cutoff parameters. Estimated cutpoints between categories are 1.32, 1.57, 2.91, 4.07.

Variable	Estimated Coefficient	Standard Error	Significance Probability
Victim's Level of Noncompliance	.234	.104	.025
Both Sides Ratified Treaty?	.169	.148	n.s.
Democracy Score of Violator	.047	.021	.030
Democracy Score of Violator times Joint Ratification	051	.016	.001
Autocracy Score of Violator	.008	.023	n.s.
Power Ratio	.668	.161	<.001
Power Ratio times Joint Ratification	478	.217	.028
Aerial Bombing	685	.153	<.001
Armistice	630	.170	<.001
Chemical and Biological Warfare	-1.140	.191	<.001
Treatment of Civilians	.453	.129	<.001
Protection of Cultural Property	216	.152	n.s.
Conduct on the High Seas	530	.149	<.001
Prisoners of War	.113	.117	n.s.
Violator Initiated War?	.225	.064	.001
Violator Battle Deaths per 1000 Population	.020	.0026	<.001
Constant $F(16, 1049) = 45, 71$	1.539	ability of Model < 00	2

Instrumental Variable Regression Predicting Noncompliance with the Laws of War

N = 1066 F(16,1049) = 45.71 Significance Probability of Model < .0001 $R^2 = .458$

The omitted category for the dummy variables of the issue-areas is treatment of the wounded.