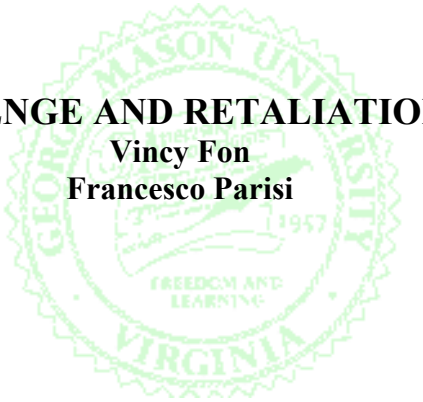
George Mason University

SCHOOL of LAW



REVENGE AND RETALIATION

Vincy Fon
Francesco Parisi

02-31

LAW AND ECONOMICS
WORKING PAPER SERIES

Vincy Fon[1] -- Francesco Parisi[2]

REVENGE AND RETALIATION

Abstract: This paper considers the role of retaliation norms as a way to induce more socially desirable behavior among self-interested parties. The paper first considers the unregulated case in which individuals indulge in mutual aggression, in the absence of other legal or social constraints. Next the relationship between aggressors and their victims is investigated, concentrating on the effect of victim's propensity to retaliate when suffering harm from others. Two retaliatory regimes are examined: proportional retribution and fixed retaliation. Special attention is paid to the impact of these regimes on the parties' interaction. The results suggest that human instincts for revenge may indeed be as important as honesty for the evolution of cooperation. More generally, retaliation norms are an important ingredient for the evolution of desirable social behavior in the absence of other social constraints or legal intervention.

JEL Codes: K10, D70, C7, Z13
Keywords: *Negative Reciprocity, Retaliation, Revenge, Vindictiveness*

Humankind operates within a set of constraints. Some constraints have evolutionary origins, others are the result of human design and still others are the result of sheer accident.[3] Norms of positive and negative reciprocity are important constraints that affect human behavior. Indeed, reciprocity constraints are pervasive among behavioral, social and legal rules. While much attention has been devoted to the economics of reciprocity in cooperation,[4] little consideration has been given to the economics of negative reciprocity and retaliation. The present study fills this gap in the literature and looks into an aspect of retaliation. The model attempts to identify the extent to which retaliation norms can be understood as a mechanism to induce more socially desirable behavior among self-interested parties.

---

[1] George Washington University, Department of Economics.
[2] George Mason University, School of Law. The authors would like to thank Dan Milkove for his comments and Erin Ruane Karsman for her research assistance.
[3] Buchanan, James M. (1978), "Markets, States, and the Extent of Morals," 68 *American Economic Review* 364-368.
[4] See, e.g., Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith (1998) "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry*, 36:3 (July), 335-352; FEHR, E. AND K.M. SCHMIDT (2000), "Theories of Fairness and Reciprocity- Evidence and Economic Applications," in M. Dewatripont, L. Hansen and S. Turnovsky (eds.) *Advances in Economics and Econometrics – 8th World Congress, Econometric Society Monographs*; Fon, Vincy and Parisi, Francesco (2003),"Reciprocity-Induced Cooperation," *Journal of Institutional and Theoretica l Economics* __-__.

This paper is structured as follows. Section 1 starts with the stylized fact that humans have a natural predisposition towards reciprocity and retaliation. Although different theories provide different explanations for the origins of retaliatory norms in human behavior, most evolutionary theories suggest that if reciprocal and retaliatory behaviors pay off, processes of cultural adaptation would generate norms to specify the forms that reciprocity will take. This paper proceeds to investigate whether reciprocal behavior, in particular retaliatory behavior, pays off, and it further identifies the conditions under which retaliatory behavior serves as an instrument for achieving more desirable social outcomes.

In Section 2, we develop an economic model of mutual aggression and unregulated revenge in which parties can draw a benefit from attacking others, and are subject to attack from other parties. We examine the subgame perfect Nash equilibrium outcome in this game, describing the interaction between independent parties in the absence of norms of retaliation and other legal or social constraints. This Section serves as a benchmark case for the subsequent study of retaliation norms.

Sections 3 and 4 consider the impact of two alternative retaliatory regimes. We compare the mutual aggression outcome to the results induced by such regimes. The first regime of retaliation, presented in Section 3, is characterized by *kind-for-kind* retaliation, subject to an *ex post* test of proportionality. In such an environment, victims of a wrong privately carry out in-kind retaliation. The degree of retaliation is chosen by the aggrieved party but is subject to an *ex post* test of proportionality. The aggressor who suffered excessive retaliation is entitled to get even with his retaliator, imposing further harm induced by the excess retaliation. The second regime, presented in Section 4, is characterized by a *measure-for-measure* rule of fixed retaliation. Under such a regime, the original victim is allowed to retaliate by duplicating the harm in the same objective measure and modality as the harm suffered initially. Since retaliation is strictly regulated and the retaliatory action cannot exceed the measure of the original harm, no *ex post* test of proportionality follows under such a regime.

Section 5 compares the results of our two regimes of retaliation, examining the differential impact of those alternatives when heterogeneous parties are involved in a conflict. Section 6 summarizes the results of the paper, exploring the extent to which

retaliation norms can be viewed as important ingredients of the evolution of peaceful cooperation.

1. Retaliation and Negative Reciprocity

In spite of great variation of ethical values from one culture to another, norms of reciprocity and retaliation stand as universal principles in virtually every human society, both historical and contemporary. No single principle or judgment is as widely and universally accepted as the reciprocity principle, in both its positive and negative versions. The relative importance of the positive and negative components of the reciprocity principle appears to depend on the state of advancement of society and administration of justice. More notably, reciprocity norms first materialize in their negative form in lesser developed societies, while norms of positive reciprocity dominate in more developed societies. In early codes of the Babylonian and Biblical tradition, the reciprocity principle takes the first form as a principle of negative reciprocity or retaliation.[5] The talionic principle of "an eye for an eye, a tooth for a tooth," is the most notable illustration of the early principles of negative reciprocity. Similar incarnations of principles of retributive justice emerge in virtually every early legal system for the treatment of wrongdoing, both voluntary and involuntary.[6] These rules in turn represent a broader concept of reciprocity, which was subsequently articulated as a positive mandate. The command to "love thy neighbor as thyself" sums up the positive and prescriptive nature of the rule of positive reciprocity**.**

Economists and behavioral scientists have devoted considerable attention to both positive and negative connotations of reciprocity. Already by the early 1870s, Charles Darwin in his *Descent of Man,* wrote:

> …[A]s the reasoning powers and foresight…became improved, each man would soon learn from experience that if he aided his fellow-men, he would commonly receive aid in return. From this low motive he might

[5] Early notions of punitive justice are embedded in the ancient practices of indiscriminate personal revenge. In this sense, Biblical scholars describe practices of retaliation as a form of "revenge traveling towards justice". See Blau, Joel (1916), "Lex Talionis," 26 *Yearbook of the Central Conference of American Rabbis* 1, p. 4.

[6] Parisi, Francesco (2001), "The Genesis of Liability in Ancient Law," 3 *American Law and Economics Review*, 3: 82.

> acquire the habit of aiding his fellows; and the habit of performing benevolent actions certainly strengthens the feelings of sympathy, which gives the first impulse to benevolent actions.[7]

In recent years, several seminal studies have developed theories of reciprocity and retaliation that provide an evolutionary explanation of such behavior in humans and animals. With this expansion has come debate over what motivates and propagates reciprocity and retaliation. Evolutionary biologists and sociobiologists argue that the species that have successfully evolved over time are those that have incorporated some form of positive and negative reciprocity into their preference profiles and behavioral patterns. In particular, evolutionary forces have led individuals to acquire a taste for retaliatory justice, which generates a private benefit by imposing retaliation on a wrongdoer. Conversely, evolutionary psychologists and Darwinian anthropologists suggest that reciprocity requires a more complex explanation than simple genetics provides.[8] This school of thought focuses on conscious (rather than subconscious) motivations for human behavior, suggesting that environmental factors greatly influence how and why humans act as they do.[9]

The common ground of understanding between these schools of thought is that, as a result of evolution, be it genetic or cultural, humans have developed an innate sense of fairness. This sense of fairness is the foundation for both positive and negative reciprocity attitudes. Research has shown that, in conjunction with positive and negative reciprocity attitudes, human actors are particularly skilled at detecting cooperators and cheaters in social interactions. Interestingly, current research shows that people are in many ways better at solving problems that require cheater detection (deciding whether a social contract had been violated), relative to problems that involve detecting cooperators.[10] This suggests that human psychology developed in such a way as to give a greater role to

---

[7] Wright, Robert (1994), *The Moral Animal* 189, Vintage Books, New York, quoting Charles Darwin *The Descent of Man, and Selection in Relation to Sex* vol. 1, pp.163-64.

[8] Evolutionary psychologists and Darwinian anthropologists do not view reciprocal and retaliatory behavior as the result of a gene that subconsciously motivates its host to act in particular ways. Evolutionary psychologists hypothesize that our minds are predisposed to learn behavioral responses that promote cooperative outcomes. While humans are not born with fair, cooperative or reciprocal responses, we learn such responses developmentally from social exposure.

[9] Friedman and Singh (1999) show that positive or negative regard for others is not an innate and unconditional sentiment but rather is contingent on others' behavior. Friedman, Daniel, and Nirvikar Singh (1999) "On the Viability of Vengeance," UC Santa Cruz, Discussion Paper, 11.

negative reciprocity than to positive reciprocity, facilitating the second-party enforcement of such norms.

Research in behavioral and experimental economics recognizes the stylized fact that humans are predisposed towards negative reciprocity.[11] Be it genetic, cultural, or institutional, retaliatory attitudes often characterize human action. Actual circumstances and experiences may lead to retaliatory behavior by many people. Experimental and behavioral evidence show that people exhibit a strong tendency towards reciprocity. This suggests that there are punitive and retaliatory motives that lead humans to retaliate even when it is privately suboptimal to carry out punishment. Humans demonstrate a willingness to punish defectors, even when punishment is personally costly and there are no plausible future benefits from so behaving.[12] The presence of a taste for negative reciprocity has been confirmed by experimental evidence showing that, although payoff consequences are important, other motives that are not captured by the objective payoffs of the game constitute the driving force of retaliatory behavior. These retaliatory attitudes are triggered when humans interact with other humans, but are not present when the game is played against impersonal entities. For example, people react differently when playing against a computer (to which they cannot attribute defection intentions) as opposed to playing against other humans. Conversely, harm suffered from the action of other humans justifies a retaliatory response, even in one-shot interactions where reputational incentives are not at work. No such retaliation is generally observed when playing against a mindless computer. Likewise, if the harm was occasioned by subjects who had no alternative behavioral choice, blame is absent and the victims' natural instinct for retaliation is less likely to be present.[13]

---

[10] Cosmides, Leda and John Tooby, "Evolutionary Psychology: A Primer."

[11] Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith (1998), "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry*, 36:3 (July), 335-352.

[12] Gintis, H. (2000), *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Behavior*, Princeton, N.J.: Princeton University Press, p. 262.

[13] Intention-based theories predict an absence of punishment in games in which no intention can be expressed. The motive to punish unfair intentions plays an important role in negative reciprocity. Fehr and Schmidt (2000) provide evidence suggesting that many subjects who reduce the payoff of other players lack the desire to change the equitability of the payoff allocation. Instead, many subjects seem driven by the desire to punish. The desire to hurt other players is consistent with intention-based models of reciprocity. Fehr, Ernst, and Klaus M. Schmidt (2000), "Theories of Fairness and Reciprocity- Evidence and Economic Applications," prepared for the invited lecture session on Behavioral Economics at the 8th World Congress of the Econometric Society in Seattle, (December), 31.

Current theories and experimental evidence thus concur in suggesting that an instinct for negative reciprocity is deeply rooted in human nature.[14] As Friedman and Singh put it, "vengeance or a taste for negative reciprocity is an important part of the human emotional repertoire."[15] These findings raise the fundamental question as to whether revengeful and retaliatory behavior pays off. This study centers its focus on this question and identifies the environments under which retaliatory behavior may be explained as instrumental to achieve more desirable outcomes. In the following sections, we develop an economic model of unregulated revenge under which parties benefit from unilaterally attacking others and where individuals benefit from retaliation when suffering wrongdoing from others. Subsequent sections consider alternative regimes of retaliation and their impact on outcomes in case of conflict.

2. Violence in the State of Nature

When conflicts arise in the absence of a commonly recognized rule of conduct, interaction between individuals is governed by the most elementary law of nature: what one party can do to another, the other can do as well. Subject to their relative strength, parties engage in a relationship of mutual aggression. Relationships based on force likely permeated interaction between individuals and groups for a great part of human history.

---

[14] Friedman and Singh (1999) point out viability and credibility problems with vengeance and negative reciprocity. These problems find different answers in the literature. Friedman, Daniel, and Nirvikar Singh (1999) "On the Viability of Vengeance," UC Santa Cruz, Discussion Paper, 12. Bowles and Gintis (2001) consider the genetic evolution of vengeance in the context of a voluntary contribution game. They contemplate a direct tie between two discrete traits, a preference for punishing shirkers and a preference for helping a team of cooperators. Bowles, Samuel and Herbert Gintis (2001), "The Evolution of Strong Reciprocity," (December 2001), available at http://www.umass.edu/preferen/gintis/evolsr.pdf. Friedman and Singh (1999) propose a coevolution model to demonstrate the vengeance motive, where coevolution refers to the interaction of individual level ("gene") selection and group level ("meme") selection. Bowles and Gintis (2001) further note that all known groups of humans maintain social norms, or memes, that prescribe appropriate behavior towards fellow group members and typically prescribe different appropriate behavior towards individuals outside the group. The success of the meme, as with any other adaptive unit, is measured by its ability to displace alternatives, or its fitness. A meme prescribing a particular pattern of vengeful behavior is more fit than existing alternatives when it brings higher average fitness to group members.

[15] Friedman, Daniel, and Nirvikar Singh (1999) "On the Viability of Vengeance," UC Santa Cruz, Discussion Paper, 19. The authors model the important role of vengeance in sustaining cooperative behavior but highlight an intrinsic free-rider problem: "the fitness benefits of vengeance are dispersed through the entire group but the fitness costs are borne personally."

We consider a stylized setting where aggressors obtain a unilateral benefit by attacking other parties, and where victims are allowed to indulge in like behavior. After considering the subgame perfect Nash equilibrium, we derive the social optimum in such a mutual aggression. These results serve as a benchmark for evaluating subsequent regimes of regulated retaliation.

2.1 Asymmetric Parties

Consider two parties with different relative strengths. Both parties are potential aggressors or victims of aggression. Regardless of their strength, aggressors can draw a unilateral benefit when attacking another party. One party's attack imposes a cost on the other party. Individuals differ in their subjective propensity to attack and retaliate. The subjective net benefits from aggression are assumed to differ between parties according to their strength, while costs imposed on the victims are assumed equal.

The aggression strategy adopted by each party is $s_i \in [0,1]$. Each party's payoff function depends on the level of aggression exercised against others and the level of harm suffered due to others' aggression. We assume that the marginal benefit of aggression is constant, while there is increasing marginal cost from the harm suffered due to others' aggression.

Parties have different predispositions towards aggression, represented by different net marginal benefits of harm and retaliation for the two parties. For any given level of aggression, the losses imposed on the respective opponents are assumed equal. Thus, the payoffs for the parties are given by:

$$P_1(s_1, s_2) = -as_2^2 + bs_1,$$

$$P_2(s_1, s_2) = -as_1^2 + cs_2.$$

We assume that $0 < b < a$ and $0 < c < a$ to ensure that the highest levels of mutual aggression ($s_1 = s_2 = 1$) generate negative payoffs for both parties. Further, without loss of generality, consider the case in which party 1 enjoys a higher net marginal benefit from harming the other party: $c < b$. This implies that party 1 has a greater predisposition towards aggression than party 2. This may be because of a stronger subjective preference for aggression and retaliation or because of greater strength and

lower costs to aggression. We thus generally refer to party 1 as the stronger or more aggressive party and party 2 as the weaker or less aggressive party. Combining the above assumptions, the parameters a, b, c satisfy the requirement $0 < c < b < a$.

2.2 The Subgame Perfect Nash Equilibrium

Aggressors can benefit from unilaterally attacking others, and potential victims can undertake similar strategies against their aggressors. Parties choose the level of aggression according to their subjective strength and predisposition. In stage 1, one party (party 1) chooses his strategy $s_1 \in [0, 1]$. In stage 2, the other party (party 2) chooses his strategy $s_2 \in [0, 1]$. Information is complete. That is, the structure of the game and the rationality of the parties is common knowledge.

Now consider the subgame perfect Nash equilibrium of the game by backward induction. In stage 2, as $\frac{\partial P_2}{\partial s_2} = c > 0$, party 2 chooses $s_2 = 1$ for any given s_1. In stage 1, given that party 2 will choose $s_2 = 1$, party 1 chooses $s_1 = 1$ since $\frac{\partial P_1}{\partial s_1} = b > 0$. Thus, the subgame perfect Nash equilibrium outcome is $(s_1^N, s_2^N) = (1, 1)$ with full mutual harm for both parties. The subgame perfect Nash equilibrium payoff for parties 1 and 2 are $b - a$ and $c - a$ respectively. These payoffs are both negative.

2.3 The Social Optimum

In order to evaluate the outcomes of mutual aggression in the state of nature and of the alternative regimes of retaliation considered in the following sections, we identify the socially optimal levels of mutual aggression as those that maximize aggregate payoffs for the parties. The social outcome which fulfills the Kaldor-Hicks criterion is given by:

$$\arg\max_{s_1, s_2} \quad (-as_2^2 + bs_1) + (-as_1^2 + cs_2).$$

Thus, the efficient outcome consists of strategies $(s_1^S, s_2^S) = (\frac{b}{2a}, \frac{c}{2a})$. Note that this outcome is characterized by a positive level of mutual aggression –a level, however, that falls short of the subgame perfect Nash equilibrium level that would dominate in the state of nature.[16]

The payoffs for parties 1 and 2 under the socially efficient outcome are $\frac{2b^2 - c^2}{4a}$ and $\frac{2c^2 - b^2}{4a}$ respectively, while the total social payoff equals $\frac{b^2 + c^2}{4a}$. Note that although the payoff in a social optimum for the stronger party, party 1, is always positive, the payoff for the weaker party, party 2, may not be. For the weaker party's payoff to be positive, its level of aggressiveness must be fairly high compared to the more aggressive counterpart.[17] However, the parties' total payoff in a social optimum will always be positive.

Since the parties cannot effectively bind their strategies to one another, each party faces a dominant strategy of aggression. As a result, greater overall violence than is socially optimal will obtain in the subgame perfect Nash equilibrium. This highlights the problems associated with unregulated revenge, as anthropologists and historians have often suggested. We finally note that in the mutual aggression case examined above, the subgame perfect Nash equilibrium is not order-dependent. The same outcome, characterized by maximum harm and maximum retaliation, $(s_1^N, s_2^N) = (1, 1)$, would obtain regardless of the order of moves and in both simultaneous and sequential games. Likewise, the socially optimal outcome is independent of the order of moves and equal in both simultaneous and sequential games: $(s_1^S, s_2^S) = (\frac{b}{2a}, \frac{c}{2a})$.

---

[16] Since $0 < c < b < a$, $s_1^S = \frac{b}{2a} < \frac{b}{a} < 1 = s_1^N$ and $s_2^S = \frac{c}{2a} < \frac{c}{a} < 1 = s_2^N$.

[17] That is, $\frac{2c^2 - b^2}{4a}$ is positive if $2c^2 > b^2$. Since $c < b$ implies $c^2 < b^2$, the requirement that $2c^2 > b^2$ must hold for the payoff of the weaker party to be positive means that c cannot be too much smaller than b.

2.4 From Mutual Aggression to Revenge and Retaliation

All human societies practiced retaliation at one stage or another. Practices of retaliation evolved over time. In the early phase of discretionary retaliation, there were no formal or legal controls on the victim's behavior. Early customs of retaliation granted victims some degree of discretion over the severity of punishment imposed on wrongdoers. The early conceptions of retaliatory justice, however, often imposed qualitative limits on punishment. Retaliation contained the idea of punishment in-kind – captured by the etymology of the word *talio* (retaliation), which comes from the word *talis* (equal in kind) – without imposing any limit on the measure of punishment. In other words, early norms of retaliatory justice embedded the notion of "kind-for-kind" punishment without imposing the additional constraint of "measure-for-measure".[18]

Although no rational departure from the peaceful equilibrium would be expected under a kind-for-kind regime, given that a disturbance of the peaceful equilibrium could prove very costly, an involuntary shock could trigger a medium-term feud with considerable dissipation of wealth.[19] These problems were subsequently mitigated by the emergence of norms of proportional retribution, which led to an *ex post* scrutiny of the private retaliation carried out by the original victim. At this stage, retaliation was still privately carried out and thus was influenced by the victim's subjective instinct for revenge, but in case of blatantly disproportionate retaliation, the unequal harm done to the parties could be brought into balance by imposing in kind punishment on the overreacting party.[20] We refer to this regime as *kind-for-kind* retaliation.

---

[18] See Blau, Joel (1916), "Lex Talionis," 26 *Yearbook of the Central Conference of American Rabbis* 1, p. 7 and Parisi, Francesco (2001), "The Genesis of Liability in Ancient Law," 3 *American Law and Economics Review.* Under a kind-for-kind regime, wrongdoing was vindicated by the victim (or in the case of homicide, by the closest kin within the group). These practices of retaliation initially allowed private revenge with talionic multipliers greater than one and gave origin to possible spirals of violence.

[19] Parisi, Francesco (2001), "The Genesis of Liability in Ancient Law," 3 *American Law and Economics Review* notes that, in the event of an involuntary disturbance of the peaceful equilibrium (e.g., involuntary harm, mistakenly attributed wrongdoing, etc.), a costly game of mutual aggression would follow.

[20] In the absence of a commonly accepted rule, the measure of revenge was left to the discretion of the victim or his clan. As historians and anthropologists tell us, in kind punishment often came in multiples of the harm originally suffered by the victim. Parisi (2001) considers the dynamics of retaliation under this early regime of retaliatory punishment. Parisi, Francesco (2001), "The Genesis of Liability in Ancient Law," 3 *American Law and Economics Review.*

The subsequent evolution of norms of retaliation led to the articulation of fixed retaliatory penalties, imposing a *measure-for-measure* constraint on the parties' retaliatory strategies. The Biblical *lex talionis* and the comparable provisions found in ancient codifications introduced an *ex ante* constraint on private retaliation by the victim, with an upper limit of 1:1 to the measure of legitimate retaliation.

Historically, the administration of retaliatory justice proceeded from privately carried out revenge to forms of supervised retaliation. In this context, it should be noted that the two regimes of retaliation impose different monitoring requirements for their implementation. The kind-for-kind retaliation regime allows parties to carry out retaliation without any adjudicature and only requires a system for the *ex post* correction of excessive retaliation. As the development of the law progressed, the restrictions regulating vengeance were extended. The injured party, who was originally allowed to carry out the execution himself (subject to a constraint of proportionality), later was only allowed to do so under the supervision of authority, and eventually was only permitted to attend the execution. The second regime of measure-for-measure retaliation utilized third party supervision of the talionic punishment to prevent excesses. Under this phase of supervised retaliation, the *talion* was carried out by the victim (or his family) in the presence of witnesses and under the direct supervision of an official executioner.[21]

It is interesting to note that the historical illustrations of retaliation considered in this paper were instrumental to both promoting and constraining practices of retaliation.[22]

In some situations the human instinct for revenge provides the natural impetus for carrying out retaliation even when it is privately costly and may appear *ex post* irrational. In those situations, the natural instinct for revenge may extend beyond proportional retaliation and the measure-for-measure limit to retaliation serves to constrain such human impetus.[23] Yet in other situations, retaliation norms emerge to encourage

[21] The absence of a police system across different clans may thus explain why the kind-for-kind practices of retaliation historically preceded the institution of measure-for-measure retaliation.

[22] Unlike prior practices carried out in the absence of customary or codified rules, the *lex talionis* created an express and well-defined punitive rule. The victim (or his family) was entitled – and, at the same time, obligated – to perpetrate literal talionis in the measure indicated by the law: "[t]hou shalt give life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound for wound, stripe for stripe" (Exodus 21:23-25).

[23] A possible explanation for the higher-than-proportional instinct for revenge is given by the presence of enforcement errors, requiring higher multipliers to maintain effective deterrence.

retaliatory practices. This second function of norms of retaliation can be understood considering that retaliation often imposes a private cost on the retaliator, as well as on the wrongdoer.[24] Revenge duplicates the loss, rather than effectuating a compensatory transfer, and as such, is incapable of undoing the original harm. Given the sunk nature of the original loss, carrying out costly retaliation is often irrational *ex post*. Further, retaliation creates a public benefit to society at large in terms of maintained deterrence for future wrongdoing, while imposing a private cost on those who enforce talionic penalties. Similarly, when members of the victim's family need to carry out retaliation in the absence of a central enforcement authority, a second-order collective action problem may arise, with free-riding in retaliation. These factors may generate a suboptimal participation incentive to carry out the retaliatory action. In this sense, social and religious norms and other retaliatory institutions can be viewed as instruments to maintain an optimal level of private enforcement of retaliatory punishment.[25]

The following two sections examine the effect of these retaliation norms in promoting a more desirable social coexistence. Section 3 analyzes norms of kind-for-kind retribution and Section 4 considers rules of measure-for-measure retaliation, studying the effects of such constraints on the parties' interaction.

3. "Kind-for-Kind" Revenge with *Ex Post* Test of Proportionality

The first regime of regulated retaliation is characterized by retaliation subject to an *ex post* test of proportionality. Victims of a wrong carry out in-kind retaliation. The

[24] The avengers of blood risked their own lives for the vindication of the group as a whole. Sulzberger, Mayer (1915), *The Ancient Hebrew Law of Homicide.* Philadelphia: Julius H. Greenstone.

[25] The vindication of an innocent victim is generally carried out by the victim's clan. As pointed out by the rabbinic interpretations, under the older tradition described in Genesis, the blood avenger is not a definite person, any member of the tribe could carry out the retaliation. See, Sulzberger, Mayer (1915), *The Ancient Hebrew Law of Homicide.* Philadelphia: Julius H. Greenstone, at p. 1 and 116. The action, however, was ordinarily orchestrated by the chief of the clan, who, acting as a residual claimant, had an interest in minimizing the external losses to his group. Under the later rule of Deuteronomy, we find a more detailed specification of the procedure for the talionic punishment. The diffuse punitive entitlement of the earlier customs described in Genesis rendered the administration of justice rather uncertain and unrestrained. The change brought about by the rules of Deuteronomy reduced the risk of coordination errors, avoiding the possibility of multiple reprisals for the same wrong as well as the likelihood of leaving some wrongs unpunished. Under the regime of Deuteronomy, the institution of *Go'el*, the nearest of blood, evolves. In the absence of a central law-enforcement system, the closest family member of the victim had the right –

degree of retaliation is chosen by the aggrieved party but is subsequently subject to a test of proportionality. In case of excessive retaliation, the party suffering excessive retaliation can seek relief and impose further harm on the overreacting retaliator. Norms of proportional retribution render excessive retaliation a wrong in and of itself and justify action by the victim of retaliation to reestablish the balance of reciprocal harm with its retaliator.

In order to consider the outcome of retaliatory interaction in the presence of a proportionality test, we modify the previous game by introducing a test of proportionality after the first exchange of harm and retaliation by the parties. Given the possibility of excessive retaliation by the second mover, we extend the simple game to a three-stage game to allow for the getting-even round. In particular, in stage 1 we assume that the first party chooses its initial level of aggression. We refer to this as the *aggression stage*, where the first party harms the other party. In stage 2, the *retaliation stage*, the second party retaliates. The extent of this retaliation may trespass into the region of excessive retaliation. In stage 3, the behavior of the second party is evaluated in light of a test of proportionality. Punitive actions greater than the socially accepted level of retaliation would be considered excessive.

Although the extent of acceptable retaliation changes over time, the following model considers the case in which imposing punishment equal in severity to the original harm constitutes the limit of acceptable retaliation. If the retaliation carried out by the victim or his clan exceeds such a limit, the first party may get even with its retaliator under the criterion of proportional retribution. We refer to this as the *getting even stage*.

Note that in the following models, tests of proportionality and rules of fixed retaliation focus on objective harm caused by the parties, rather than subjective loss suffered by the parties. Thus, for example, the loss of an arm justifies an equal mutilation of the wrongdoer's limb, regardless of the parties' subjective valuation of their respective bodily integrity. This objective application of retaliatory norms is consistent with

---

and more importantly, the duty – to carry out retaliation. Failure to carry out such a gloomy task was considered disgraceful. This further assured the consistent punishment of wrongdoers.

historical examples of retaliation, which often called for a mechanical duplication of harm regardless of the subjective circumstances of the case.[26]

Similar to the case of mutual aggression in the state of nature, we assume that an aggressor can draw a unilateral benefit when attacking another party. Victims benefit from engaging in vindictive behavior when suffering an unjustified attack from another. This benefit may result from an evolved taste for revenge, or from compliance with existing social norms, requiring victims to retaliate in order to preserve their honor and reputation. The harmful activity imposed by one party imposes a cost to the other party subject to attack. Individuals differ in their subjective propensity to attack and retaliate. Each party's payoff function depends on the level of aggression exercised against others and the level of harm suffered from others. We assume that there are bounds to the levels of aggression and retaliation: $s_i \in [0,1]$. Without loss of generality, we consider the case in which party 1 enjoys a higher net marginal benefit from harming the other party: $c < b$. This means that party 1 has either a greater predisposition towards aggression or a greater instinct for revenge.

To better understand the features of this regime of retaliation with an *ex post* test of proportionality, we consider two scenarios. In the first scenario, the more aggressive party (party 1) moves first, undertaking an aggression on the other, less aggressive party (party 2). We refer to this scenario as the *MLM* game to signify the sequence of moves: *more* aggressive first, *less* aggressive second, and *more* aggressive third. In the second scenario, the order of moves is reversed. We shall refer to this alternative scenario as the *LML* game to signify the reversed sequence of moves: *less* aggressive first, *more* aggressive second, and *less* aggressive third.

3.1 The MLM Case

In the *MLM* game the more aggressive party (party 1) moves first, harming the other, less aggressive party (party 2). The less aggressive party retaliates in the second

[26] Anthropologists and legal historians note that talionic rules always impose harm in the same objective gravity (eye for an eye) regardless of the subjective loss suffered by the victim. See, e.g., Parisi, Francesco (2001), "The Genesis of Liability in Ancient Law," 3 *American Law and Economics Review*, 3: 82.

stage, and the more aggressive first party imposes an eventual getting-even punishment in the third stage.

Recall that under this regime of retaliation, parties choices in stages 1 and 2 are unconstrained, and are subject to a subsequent test of proportionality such that, if the measure of retaliation in stage 2 exceeded the measure of harm occasioned by the original aggression in stage 1, a getting-even stage 3 provides the occasion to reestablish the balance of harm suffered by the parties. Excessive retaliation is possible because retaliation is carried out in the absence of a monitoring authority, and the second party's instinctive preference for revenge may occasionally lead to excessive retaliatory action. Parties know that third stage behavior will be evaluated in light of a test of proportionality and subject to possible correction.

Backward induction will help us identify the parties' strategies and the equilibrium outcome of this game. In stage 3, given the retaliatory action of the less aggressive party in stage 2 (s_2), and recalling the original level of aggression in stage 1 (s_1), the allowed "getting-even" reaction of party 1 is the following: [27]

$$\tilde{s}_1^R = \begin{cases} s_2 - s_1 & if \ s_2 > s_1 \\ 0 & if \ s_2 \le s_1 \end{cases}$$

In stage 2, given the choice of party 1 in stage 1 (s_1) and knowing the feasible reaction of party 1 in stage 3 ($\tilde{s}_1^R$), party 2 confronts the following problem:

$$\underset{s_2}{Max} \quad P_2 = -a(s_1 + \tilde{s}_1^R)^2 + cs_2 .$$

This is equivalent to:

$$\underset{s_2}{Max} \quad P_2 = \begin{cases} -as_2^2 + cs_2 & if \ \ s_2 > s_1 \\ -as_1^2 + cs_2 & if \ \ s_2 \le s_1 \end{cases} .$$

The first branch represents *excessive retaliation*. Here, the maximization problem captures the effect of the reciprocity constraint in the retaliation game, showing the expected cost of the getting-even stage as a result of excessive retaliation in stage 2. The second branch shows *fair retaliation*, in which the retaliatory action does not exceed the original harm. In this latter case, since retaliation s_2 does not exceed the boundary s_1, the

retaliator faces no expected cost from stage 3 in terms of punishment for disproportionate reaction. Thus, if carried out within the confines of proportionality, retaliation in stage 2 can create benefits without imposing any cost on the retaliator (party 2).

If $s_2 > s_1$, $\arg\max_{s_2} P_2 = \frac{c}{2a}$. If $s_2 \le s_1$, $\arg\max_{s_2} P_2 = s_1$. Hence the reaction function of party 2 is:

$$s_2^R = \begin{cases} \frac{c}{2a} & if \ \frac{c}{2a} > s_1 \\ s_1 & if \ \frac{c}{2a} \le s_1 \end{cases}.$$

In stage 1, knowing the reaction of party 2 (s_2^R) and the possibility to get even in stage 3, party 1 confronts the problem:[28]

$$\underset{s_1, \tilde{s}_1}{Max} \quad P_1 = -a s_2^{R2} + b(s_1 + \tilde{s}_1)$$

$$= \begin{cases} -a(\frac{c}{2a})^2 + b(s_1 + \tilde{s}_1) & s.t. \ s_1 + \tilde{s}_1 = \frac{c}{2a} \quad if \ \frac{c}{2a} > s_1 \\ -a s_1^2 + b s_1 & s.t. \ \tilde{s}_1 = 0 \quad if \ \frac{c}{2a} \le s_1 \end{cases}$$

In the first branch, representing the choice of excessive retaliation ($\frac{c}{2a} > s_1$), party 1 chooses a small harmful activity in stage 1, expects a disproportionate retaliation from party 2 in the following stage, and is allowed to get even in stage 3 to reestablish the overall balance between respective harms suffered by the parties. In this case, there is an infinite number of solutions to s_1 and $\tilde{s}_1$, subject to the conditions $s_1 + \tilde{s}_1 = \frac{c}{2a}$ and $s_1 < \frac{c}{2a}$. Whatever the choices of s_1 and $\tilde{s}_1$, the payoff for party 1 is $\frac{c(2b-c)}{4a}$.

---

[27] The notation ~ is used to represent the choice variable of a party in stage 3, the subscript *i* refers to party *i*, and the superscript *R* denotes decisions made under the retaliation game.

[28] The problem confronting party 1 can be described more compactly by writing the two sub-problems with more than one constraint and without an "if" proposition. However, our formulation renders the backward induction more transparent.

In the second fair retaliation branch ($\frac{c}{2a} \le s_1$), party 2 matches the harmful activity imposed by party 1. In this case, $\arg\max_{s_1} -as_1^2 + bs_1 = \frac{b}{2a}$, and the best payoff for party 1 is $\frac{b^2}{4a}$. Since $\frac{b^2}{4a} > \frac{c(2b-c)}{4a}$ is equivalent to $(b-c)^2 > 0$ and the latter inequality always holds, the payoff from fair retaliation always exceeds the payoff from engaging in excessive retaliation. This leads party 1 to opt for $s_1^R = \frac{b}{2a}$. Consequently, this regime of retaliation induces equilibrium strategies $(s_1^R, s_2^R, \tilde{s}_1^R) = (\frac{b}{2a}, \frac{b}{2a}, 0)$. The payoffs for parties 1 and 2 are $\frac{b^2}{4a}$ and $\frac{b(2c-b)}{4a}$, respectively.

3.2 The LML Case

In this scenario the less aggressive party (party 2) moves first, harming the other, more aggressive party (party 1). The more aggressive party retaliates next, and the less aggressive party moves again in the third stage to impose the getting-even punishment.

In stage 3, the getting-even stage, party 2 chooses its strategy given the previous choice of the more aggressive party in stage 2 (s_1), and recalling its own choice in stage 1 (s_2), according to the following reaction function:

$$\tilde{s}_2^R = \begin{cases} s_1 - s_2 & \text{if } s_1 > s_2 \\ 0 & \text{if } s_1 \le s_2 \end{cases}$$

In the retaliation stage 2, given the choice of party 2 in stage 1 (s_2) and knowing the expected reaction of party 2 in stage 3 ($\tilde{s}_2^R$), party 1 confronts the problem:

$$\underset{s_1}{Max} \quad P_1 = -a(s_2 + \tilde{s}_2^R)^2 + bs_1 .$$

This is equivalent to:

$$\underset{s_1}{Max} \quad P_1 = \begin{cases} -as_1^2 + bs_1 & \text{if } s_1 > s_2 \\ -as_2^2 + bs_1 & \text{if } s_1 \le s_2 \end{cases} .$$

Recall that the first branch represents excessive retaliation and the second branch represents fair retaliation. In this latter case, the retaliator faces no expected cost from the getting-even stage.

If $s_1 > s_2$, $\arg\max_{s_1} P_1 = \frac{b}{2a}$. If $s_1 \le s_2$, $\arg\max_{s_1} P_1 = s_2$. Hence the reaction function of party 1 is:

$$s_1^R = \begin{cases} \frac{b}{2a} & if \ \frac{b}{2a} > s_2 \\ s_2 & if \ \frac{b}{2a} \le s_2 \end{cases}.$$

This allows us to consider party 2's strategy in stage 1, the aggression stage. Knowing the expected retaliation of party 1 (s_1^R) and the possibility to get even in stage 3, party 2 is confronted with the problem:

$$\underset{s_2, \tilde{s}_2}{Max} \quad P_2 = -a s_1^{R2} + c(s_2 + \tilde{s}_2)$$

$$= \begin{cases} -a(\frac{b}{2a})^2 + c(s_2 + \tilde{s}_2) & s.t. \ s_2 + \tilde{s}_2 = \frac{b}{2a} \quad if \ \frac{b}{2a} > s_2 \\ -a s_2^2 + c s_2 & s.t. \ \tilde{s}_2 = 0 \quad if \ \frac{b}{2a} \le s_2 \end{cases}$$

In the first branch, characterized by excessive retaliation ($\frac{b}{2a} > s_2$), a small harmful activity by party 2 in stage 1 would trigger excessive retaliation from party 1 in the following stage. The less aggressive original wrongdoer would be allowed to get even in stage 3, and this would lead to a level of mutual harm that exceeds his private optimum. As in the previous case where the order of moves was reversed, an infinite number of possible solutions s_2 and $\tilde{s}_2$ satisfy $s_2 + \tilde{s}_2 = \frac{b}{2a}$ and $s_2 < \frac{b}{2a}$.

A particularly interesting possibility is $s_2 = \frac{c}{2a}$ and $\tilde{s}_2 = \frac{b-c}{2a}$, with payoff equal to $\frac{b(2c-b)}{4a}$. This case is interesting because $s_2 = \frac{c}{2a}$ constitutes party 2's best choice of action if he was the stronger party. Then the retaliation-induced equilibrium outcome

is $(s_2^R, s_1^R, \tilde{s}_2^R) = (\frac{c}{2a}, \frac{b}{2a}, \frac{b-c}{2a})$. At the retaliation-induced equilibrium, payoffs for parties 1 and 2 are $\frac{b^2}{4a}$ and $\frac{b(2c-b)}{4a}$, respectively.

The second branch is characterized by fair retaliation ($\frac{b}{2a} \le s_2$). Party 1, victim of the original wrong, imposes retaliation on party 2 in the same degree as the harm suffered. Given proportionality in retaliation, no getting-even punishment in stage 3 follows. In this case, since $\arg\max_{s_2} -as_2^2 + cs_2 = \frac{c}{2a}$, the unconstrained choice for party 2 is $s_2 = \frac{c}{2a}$. Recall that party 2 is the less aggressive party as $c < b$ is assumed. Hence, $s_2 = \frac{c}{2a} < \frac{b}{2a}$ is implied. But this violates the upper bound constraint for the fair retaliation branch. Thus the upper bound constraint must be binding and $s_2^R = \frac{b}{2a}$. With this choice, the less aggressive party (party 2) preempts excessive retaliatory action by party 1, by adopting a level of initial aggression consistent with the expected level of reaction by the other party. The induced equilibrium outcome takes the form $(s_2^R, s_1^R, \tilde{s}_2^R) = (\frac{b}{2a}, \frac{b}{2a}, 0)$. The same potential payoff for party 2, $\frac{b(2c-b)}{4a}$, obtains in the fair retaliation and excessive retaliation scenarios.

This result rests on the fact that, even though the weaker party undertakes the initial aggression, it is the stronger party who ultimately determines the harm level from the retaliatory interaction. The less aggressive party has a multitude of options, all yielding identical payoff. When confronting party 1, the less aggressive first mover (party 2) can expect a level of retaliation consistent with the more aggressive predisposition of his opponent, regardless of the actual gravity of the initial offence. Party 2 can allocate s_2 and $\tilde{s}_2$ in different ways, but all alternative allocations yield the same payoff.[29]

[29] In fact, an infinite number of choices that party 2 can take all lead to the same payoff.

A second observation concerns the parties' participation in the retaliatory exchange. Since the retaliatory game is started by an initial aggression, the participation constraint to this interaction should be considered. While initial aggression by the strong party is always feasible, initial aggression by a weaker party requires special conditions. A less aggressive party will engage in such violent interaction only when the expected payoff of the game is positive. The weaker party will not engage in initial aggression if $\frac{b(2c-b)}{4a}$ is negative. For this participation constraint to be met, c must be greater than $b/2$ (recall that c is smaller than b by assumption). This further means that the difference in the strength and predisposition to violence between the parties should not be too large for the participation constraint to be met.

3.3 Comparing MLM and LML games.

Table 1 summarizes payoffs for the kind-for-kind retaliation, in which excessive or disproportionate retaliation is subsequently sanctioned with a punishment aimed at reestablishing the balance of harm inflicted on the two parties. The payoffs for the two cases of *MLM* and *LML* retaliation are as follows.

	Payoff for Party 1 (M)		**Payoff for Party 2 (L)**	**Total Payoff**
The *MLM* game	$\frac{b^2}{4a}$	$>$	$\frac{b(2c-b)}{4a}$	$\frac{bc}{2a}$
The *LML* game	$\frac{b^2}{4a}$	$>$	$\frac{b(2c-b)}{4a}$	$\frac{bc}{2a}$

Table 1: *Kind-for-Kind Retaliation*

A few conclusions should be drawn from this Table. First, the total payoffs for both parties are the same, independently of the sequence of moves between more aggressive and less aggressive parties. Neither the *MLM* game nor the *LML* game are

efficient, as the total payoffs for these games ($\frac{bc}{2a}$) are less than the total payoff under social optimum ($\frac{b^2+c^2}{4a}$). The possibility of retaliatory action by the second mover subject to *ex post* scrutiny of proportionality improves upon the scenario of unregulated revenge considered in Section 2. Recall that the subgame perfect Nash equilibrium with unregulated revenge generates total negative payoffs for the parties, $(b-a)+(c-a)<0$. The payoffs under this game thus represent a Pareto improvement over the case of unregulated revenge.

Second, regardless of the order of moves, the payoff for the more aggressive party is higher than the payoff for the less aggressive party. Hence, there is a more-aggressive-party advantage. This is quite intuitive since this regime of retaliation allows the more aggressive party to influence the total level of reciprocal violence, either directly by means of initial aggression, or subsequently by means of excessive retaliation. If confronted by a more aggressive party, the weak first mover expects retaliatory action of the other party to maximize the optimal level of revenge for the more aggressive party, regardless of the gravity of the initial offence. It is therefore rational for the weaker party to preempt the more aggressive party by adopting the more aggressive party's desired level of harm as its own, or to wait and get even at a later stage.

The only instance in which the more-aggressive-party advantage is not present is when the weaker party, being the first-mover, can control participation in this game and abstain from an aggression on its stronger opponent. This relates to the final observation that for substantial differences between predispositions to violence of the parties, the less aggressive party rationally refrains from imposing an initial harm on a more aggressive party. In particular, when the less aggressive party controls participation, the retaliatory exchange can only be observed if $0 < b/2 < c < b$.

4. Fixed Retaliation and the "Measure-for-Measure" Principle

In this section, we consider a second regime of retaliation, characterized by a more direct rule of fixed measure-for-measure retribution. Under such a regime the

original victim is allowed to impose retaliation duplicating the harm in the same objective measure and modality as the originally suffered harm. Since retaliation is strictly regulated and imposition of talionic penalties is constrained *ex ante*, retaliation cannot exceed the original harm, rendering an *ex post* test of proportionality unnecessary under such a regime.

This regime of retaliatory justice also has historical analogues. Anthropologists and legal historians have amply documented the transition from discretionary revenge to norms of proportional retribution. Across different cultures and legal traditions, practices of unregulated revenge and mutual aggression are initially constrained by norms of kind-for-kind retribution and subsequent rules of measure-for-measure retaliation. These norms were eventually codified, establishing a single talionic multiplier for almost all cases of wrongdoing. The establishment of sanctions based on fixed 1:1 retaliation characterizes this regime.[30] The incorporation of retaliatory practices into bodies of written law during the ninth and eight century BCE is best exemplified by the Biblical *lex talionis*.[31] Unlike prior practices carried out in the absence of customary or codified rules, the *lex talionis* created an express and well-defined punitive rule. The victim (or his family) was entitled – and, at the same time, obligated – to perpetrate literal talionis in the measure indicated by the law: "[t]hou shalt give life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound for wound, stripe for stripe."[32] The talionic rules of this period serve two main purposes. First, they create an upper limit to retaliatory justice: only one life for a life can be vindicated, no more. Second, they serve as minimum punishment for the criminal: no less than the law requires.[33] In later times, upper and lower limits began to diverge, with legitimate criminal penalties falling somewhere between those two boundaries.

---

[30] Interestingly, this limit is applicable independently of the level of social undesirability of the crime and the probability of detection of the wrongdoer. The generality of the 1:1 constraint, however, had several advantages over the discretionary imposition of retaliatory penalties. Contrasting the dynamics of the two legal regimes shows that the Biblical *lex talionis* introduced a stabilizing constraint in the (otherwise unstable) dynamics of discretionary retaliation. See Parisi, Francesco (2001), "The Genesis of Liability in Ancient Law," 3 *American Law and Economics Review*, 3: 82.

[31] Exodus 21:23-24; Leviticus 24:17-22.

[32] Exodus 21:23-25.

[33] In this way the lex talionis served at the same time for an "upper" and a "lower" limit to punishment. In later times, upper and lower limits began to diverge, with legitimate criminal penalties falling somewhere between those two boundaries.

We proceed to consider this form of retaliatory punishment, modifying our previous game by including a fixed retaliation constraint. As for the previous cases, in stage 1 the first party chooses its level of initial aggression. In stage 2, the second party chooses its level of retaliation, subject to a fixed retaliation constraint allowing the initial victim to impose measure-for-measure harm on its aggressor. The measure of retaliation cannot exceed the measure of the original harm.

As before, to understand the mechanics of this retaliation game we consider two alternative situations. In the first scenario, the more aggressive party (party 1) moves first, undertaking an aggression on the other, less aggressive party (party 2). We refer to these alternative scenarios as the *ML* game. In the second scenario, the order of moves is reversed. We shall refer to this alternative scenarios as the *LM* game to signify the sequence of a *less* aggressive first mover and a *more* aggressive second retaliator.

In both scenarios the first mover's choice (initial aggression choice) is unconstrained, but the second party's reaction is directly constrained by the rule of fixed retaliation, imposing a maximum ceiling to the measure of retaliatory harm. Note that the rule of fixed retaliation is germane to the principle of proportional retribution which governed stage 3 in the previously considered group of games. In the present scenario, however, proportionality has *ex ante* effects, operating as a constraint on the retaliatory reaction of the second mover, rather than operating *ex post*, as a test of proportionality to reestablish the balance between the harmful behaviors of the parties.

4.1 The ML game

In the *ML* scenario the more aggressive party (party 1) moves first, harming the other, less aggressive party (party 2). After suffering aggression from party 1, party 2 can retaliate and impose in kind harm up to the level of the harm originally imposed by the other party. We investigate the *ML* retaliation-induced equilibrium by backward induction.

In stage 2, given any choice of party 1 in stage 1 (s_1), party 2 is allowed to choose at most s_1, and is confronted with the following problem:

$$\underset{s_2}{Max} \quad P_2 = -as_1^2 + cs_2 \quad s.t.\ s_2 \le s_1 .$$

Since $\frac{\partial P_2}{\partial s_2} = c > 0$, P_2 is increasing in $[0, s_1]$. Hence $s_2^R = s_1$ and the less aggressive party matches the more aggressive party's harmful behavior.

In stage 1, knowing s_2^R, party 1 faces the following problem:

$$\underset{s_1}{Max} \quad P_1 = -as_2^{R2} + bs_1 = -as_1^2 + bs_1 .$$

Hence $s_1^R = \frac{b}{2a}$ is chosen. The retaliation-induced equilibrium outcome is $(s_1^R, s_2^R) = (\frac{b}{2a}, \frac{b}{2a})$. The retaliation-induced equilibrium payoffs are $P_1^R = \frac{b^2}{4a}$ and $P_2^R = \frac{b(2c-b)}{4a}$. Note that the sum of the payoffs for the two parties is $\frac{bc}{2a}$.

4.2 The LM game

In the *LM* scenario the less aggressive party (party 2) moves first, harming the other, more aggressive party (party 1). Also in this case, once the initial harm is inflicted, the victim is allowed to impose retaliatory harm, not to exceed the harm originally suffered. Again, we investigate the *LM* retaliation-induced equilibrium by backward induction.

In stage 2, given an initial aggression by party 2 in stage 1 (s_2), party 1 is allowed to retaliate at most s_2, and is confronted with the following problem:

$$\underset{s_1}{Max} \quad P_1 = -as_2^2 + bs_1 \quad s.t.\ s_1 \le s_2 .$$

Since $\frac{\partial P_1}{\partial s_1} = b > 0$, $s_1^R = s_2$. Hence, the more aggressive party always retaliates at the maximum allowable level, matching the harm level originally imposed by the less aggressive party in stage 1.

In stage 1, knowing s_1^R, party 2 faces the following problem:

$$\underset{s_2}{Max} \quad P_2 = -as_1^{R2} + cs_2 = -as_2^2 + cs_2 .$$

Hence $s_2^R = \frac{c}{2a}$ is chosen. The retaliation-induced equilibrium outcome is $(s_2^R, s_1^R) = (\frac{c}{2a}, \frac{c}{2a})$. The payoffs in such equilibrium are $P_2^R = \frac{c^2}{4a}$ and $P_1^R = \frac{c(2b-c)}{4a}$. Note that the sum of the payoffs is again $\frac{bc}{2a}$.

4.3 Comparing ML and LM Games

Table 2 summarizes payoffs for the *ML* and *LM* situations where the victim of a wrong can impose retaliation in-kind not to exceed the harm initially imposed by the aggressor.

	Payoff for Party 1 (M)		**Payoff for Party 2 (L)**	**Total Payoff**
The *ML* game	$\frac{b^2}{4a}$	$>$	$\frac{b(2c-b)}{4a}$	$\frac{bc}{2a}$
	$\vee$		$\wedge$	$=$
The *LM* game	$\frac{c(2b-c)}{4a}$	$>$	$\frac{c^2}{4a}$	$\frac{bc}{2a}$

Table 2: *Measure-for-Measure Retaliation*

Although neither retaliation game is efficient, the total payoffs for both games are the same. Further, both *ML* and *LM* games generate equilibria which constitute a Pareto improvement over the subgame perfect Nash equilibrium under unregulated revenge.

Second, note that payoffs for the parties depend on the order of their moves. More specifically, the payoff for the more aggressive party is larger when it moves first than when it moves last. Likewise, the payoff for the less aggressive party is larger when it

moves first than when it moves last. Hence, the regime of fixed retaliation creates a first-mover advantage.

Third, although the payoff for the more aggressive party is always positive, the payoff for the less aggressive party can be negative. Negative payoffs may be present when the more aggressive party moves first and the less aggressive party engages in a retaliatory exchange when its taste for retaliation is substantially lower than the other party: $2c < b$. This is because the weak and less aggressive party, if subject to an initial aggression, would engage in retaliation and would rationally match the level of harm imposed by the stronger initial aggressor. The resulting level of mutual aggression exceeds what the weaker party would have chosen as a first mover.

Further note that when the weaker party is the first mover, it can control the initial level of aggression and indirectly determine the level of retaliation it will endure. Participation in this situation is always assured, regardless of the parties' different predispositions to violence.

5. Unregulated Revenge, "Kind-for-Kind," and "Measure-for-Measure" Retaliation

In this section we sum up our previous findings, comparing the relationship between the regimes (a) mutual aggression in the state of nature, (b) kind-for-kind retaliation, subject to an *ex post* test of proportionality; and (c) measure-for-measure retaliation.

Both kind-for-kind retaliation and measure-for-measure retaliation regimes are improvements over mutual aggression. Although individual payoffs for strong and weak parties vary under different regimes, total payoffs for the two parties are identical under both kind-for-kind and measure-for-measure regimes. Hence the two regimes yield an equal improvement in the social aggregate payoff, compared to the benchmark case of mutual aggression in the state of nature. This may suggest that the historical transition

from kind-for-kind retaliation to measure-for-measure retaliation was driven by distributional concerns, rather than efficiency considerations.[34]

The two regimes of retaliation lead to a socially optimal level of aggression and retaliation if parties are identical. This is consistent with the general result according to which reciprocity constraints lead to optimal levels of cooperation between symmetric players.[35] The optimality result does not hold when our regimes of retaliation are applied to heterogeneous players. When asymmetries are involved, in both cases of kind-for-kind and measure-for-measure retaliation the total payoff falls short of the maximal payoff in a social optimum.

Different regimes of retaliation generate different payoffs for the parties when asymmetries are involved. Regardless of the regime, when there is a conflict, the weaker party is always worse off than the stronger party.

Additionally, depending on the circumstances some of the differences between the parties' payoffs depend on the order of moves while others depend on the parties' aggressiveness and relative strength. In the measure-for-measure regime the payoffs for the parties depend on the order of their moves. More specifically, the payoff for the more aggressive party is larger when it moves first than when it moves last. Likewise, the payoff for the less aggressive party is larger when it moves first than when it moves last. Hence, the regime of fixed retaliation creates a first mover advantage in contrast to the more-aggressive party advantage observed under the kind-for-kind regime.

One corollary follows from the above difference. Given the first mover advantage, the participation constraint for the weaker party is always fulfilled when it is the first mover in a measure-for-measure retaliation regime. This is not so in the kind-for-kind regime. As shown in Table 3 below, the participation constraint for the weaker party in such system may not be satisfied, even when the weaker party moves first.[36] This

[34] Parisi (2001) suggests that kind-for-kind retaliation regimes might lead to explosive spirals of violence when parties have different conceptions of fair retaliation. In this sense, the transition from kind-for-kind retaliation to measure-for-measure retaliation may be explained by the need to avoid such an escalation of violence. Parisi, Francesco (2001), "The Genesis of Liability in Ancient Law," 3 *American Law and Economics Review.*

[35] See Fon, Vincy and Parisi, Francesco (2003), Reciprocity-Induced Cooperation, *Journal of Institutional and Theoretical Economics* __-__.

[36] This is indicated by the fact that the payoff for party 2, the weaker party, can be positive or negative under kind-for-kind retaliation.

means that fewer instances of retaliatory conflict may emerge under the kind-for-kind regime compared to the measure-for-measure alternative.

	Payoff for Party 1 (M)		**Payoff for Party 2 (L)**	**Total Payoff**
The *LML* game (Kind-for-Kind)	$\frac{b^2}{4a}$	$>$	$\frac{b(2c-b)}{4a}$ $(>0 \text{ or } <0)$	$\frac{bc}{2a}$
	$\vee$		$\wedge$	$=$
The *LM* game (Measure-for-Measure)	$\frac{c(2b-c)}{4a}$	$>$	$\frac{c^2}{4a}$ (>0)	$\frac{bc}{2a}$

Table 3: *Comparing Retaliation Regimes*

Table 3 above compares individual and total payoffs under the two regimes of retaliation when the weak or less aggressive party moves first. We limit the comparison to this subset of situations because, as indicated above, no differences between the two regimes are ascertainable when the stronger or more aggressive party moves first (i.e., the individual payoffs are the same under the *MLM* and the *ML* games).

As mentioned before, in the kind-for-kind *LML* retaliation game, the payoff for the weaker or less aggressive party may be negative. This implies that under the kind-for-kind regime, the weaker party may abstain from undertaking an initial aggression, while initial aggression may always be rational in a measure-for measure regime. This difference is justified by the fact that, in the 2-stage *LM* game, the weaker party chooses the level of harm and indirectly controls the level of retaliation, thus being able to choose the privately optimal level of initial aggression. In the kind-for-kind regime, the weaker party lacks control over the level of mutual harm, since the equilibrium level of harm is determined by the preference of the stronger party. Thus, under such regime, the weaker party may find it rational to abstain from engaging in an initial aggression, even when aggression would have been rational in a measure-for-measure regime. Further, the weaker party can achieve a better payoff in the *LM* game than in the *LML* game. This

suggests that the transition from kind-for-kind to measure-for-measure retaliation benefits the less aggressive and weaker members in the group.

Differences in participation constraints among the various cases are also quite instructive. First, recall that neither party voluntarily chooses to engage in a game of mutual aggression, since negative payoffs are expected. However, in the state of nature, parties only control their strategies and neither party single-handedly controls the outcome. Thus, in spite of the negative expected payoffs, mutual aggression dominates in equilibrium. The remaining cases of regulated retaliation between asymmetric parties provide mixed participation incentives. The participation constraint is always satisfied under both regimes of retaliation when symmetric parties are involved. This is consistent with the intuition that if aggression yields a net benefit to one party, it would also be beneficial to its opponent, given the parties' identical preferences. Particularly, since the first regime of proportional retribution creates a more-aggressive-party advantage, the less aggressive party, if substantially weaker, may avoid participation, if given an opportunity to do so. Likewise, under the second regime of fixed retaliation the first mover has a strategic advantage. This may create incentives for the disadvantaged second-mover to avoid being attacked by a more aggressive party, if given an opportunity to escape the conflict.

6. Conclusions

Vindictiveness and retaliation may be as important as honesty for the evolution of cooperation. Negative reciprocity can achieve results that cannot be achieved with positive reciprocity alone. For example, the presence of positive reciprocity norms could not easily correct unilateral aggression problems. In our analysis, this can be seen by the fact that players without a taste for retaliation would be quite ineffective at constraining other players' unilateral aggression. Put differently, positive reciprocity and negative reciprocity have different domains of application. In the presence of cooperative first movers, positive reciprocation would provide an effective response, but in the face of an aggressive first mover positive reciprocity would provide a quite inadequate response. Retaliation with like aggression becomes necessary. Positive reciprocity will not help in

case of aggression by others, just like negative reciprocity cannot do much to reward positive cooperation from others. This indicates that positive and negative reciprocity are complementary strategies that provide best strategic attitudes in different sets of social interactions.

This leads us to suggest that there may be an important relationship between the evolution of vindictiveness and the sustainability of peaceful cooperation in human societies. A population endowed with attitudes of positive reciprocity but not ready for negative reciprocity could easily fall prey of invaders with unilateral aggression strategies. In evolutionary terms, positive reciprocity without a complementary attitude for negative reciprocity would not be evolutionarily stable.

In this paper, after considering a simplified model of mutual aggression and unregulated revenge, we examined two alternative regimes of retaliation. Under the first regime of kind-for-kind retaliation, individuals engage in private retaliation, subject to an *ex post* test of proportionality. The measure of retaliation is discretionary and depends on subjective circumstances and vindictive predisposition of the parties. Excessive retaliation is corrected *ex post*. Whenever appeasement between the parties is not achieved at the retaliation stage, balance in the relationship between the parties is reestablished at a third stage. In this stage excessive retaliation by the original victim is sanctioned with the imposition of a getting-even punishment on the overly vindictive party. Under the second regime of measure-for-measure retaliation, a victim's reaction is directly constrained by a rule of fixed retaliation, which limits punishment to the measure of the harm originally suffered. Under this regime, excessive retaliation is prevented, rendering the getting even stage unnecessary.

Our economic model identified the attributes of these regimes of retaliation. Both regimes of retaliation represent improvement over the alternative regime of mutual aggression in the state of nature. Interesting differences between the two regimes of retaliation emerge in the case of asymmetric parties. In the kind-for-kind regime with an *ex post* test of proportionality, the overall level of reciprocal violence is ultimately determined by the preference of the more aggressive party. This creates a more-aggressive-party advantage. No such advantage is found under the second regime of retaliation, characterized by fixed measure-for-measure punishment. In this regime the

equilibrium level of reciprocal harm is unilaterally determined by the initial aggressor, since the victim can only replicate the harm that it originally suffered. Thus, with asymmetric parties the preference of the active aggressor rather than the retaliator will be satisfied. This creates a first mover advantage. These differences vanish in two sets of circumstances. First, the two regimes yield no differences in the retaliation-induced outcomes when symmetric parties are involved. Second, no advantage will be present if the disadvantaged party can exit the game, for example by refraining from attacking a more aggressive opponent. The model further shows the limits of the various retaliatory regimes when heterogeneous parties are involved. This may explain the success and diffusion of norms of retaliation among homogeneous groups and the gradual abandonment of such retaliatory regimes when differences among groups and individuals over time became more sizeable.

These results reveal that norms of proportional retribution and practices of fixed retaliation can increase social value by avoiding the subgame perfect Nash equilibrium of mutual aggression and by encouraging parties to converge towards a more desirable level of peaceful coexistence. Given a vengeance motive and preference for retribution in human nature, peaceful and cooperative behavior is no longer dominated by strategies of unilateral aggression and can become part of a Nash equilibrium even when there is no repeat interaction. The fear of proportional retaliation can support better social outcomes, effectively constraining the levels of mutual aggression that would otherwise dominate in equilibrium. These results support recent theories providing an evolutionary explanation of negative reciprocity in human behavior. These theories suggest that retaliatory attitudes develop because they pay off. Human attitudes for revenge and retaliation operate as a trigger device that allows players credibly to pre-commit to carry out retaliation in case of unjust harm. These human traits allow players to avoid the undesirable outcome of mutual aggression and unregulated revenge that dominates in the absence of such emotional or cultural constraints.

Vincy Fon
Assistant Professor of Economics
George Washington University
601 Funger Hall
2201 G Street, N.W.
Washington, D.C. 20052
vfon@gwu.edu

Francesco Parisi
Professor of Law
George Mason University School of Law
3401 North Fairfax Drive
Arlington, VA 22201
parisi@gmu.edu