

Tilburg University

The Sorry Clause

Srivastava, Vatsalya

Publication date:
2016

Document Version
Early version, also known as pre-print

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Srivastava, V. (2016). *The Sorry Clause*. (CentER Discussion Paper; Vol. 2016-008). CentER, Center for Economic Research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Discussion paper

THE SORRY CLAUSE_

by
Vatsalya Srivastava

January 21, 2016

TILEC Discussion Paper No. 2016-004
CentER Discussion Paper No. 2016-008

ISSN 2213-9532
ISSN 2213-9419
<http://ssrn.com/abstract=2737286>

The Sorry Clause*

Vatsalya Srivastava[†]

Tilburg University, CentER, TILEC

January 21, 2016

Abstract

When players face uncertainty in choosing actions, undesirable outcomes cannot be avoided. Accidental defections caused by uncertainty, that does not depend on the level of care, require a mechanism to reconcile the players. This paper shows the existence of a *perfect sorry equilibrium* in a game of imperfect public monitoring. In the sorry equilibrium, costly apology is self-imposed in case of accidental defections, making private information public and allowing cooperation to resume. Cost of the apology required to sustain this equilibrium is calculated, the efficiency characteristics of the equilibrium evaluated and outcomes compared to those from other bilateral social governance mechanisms and formal legal systems. It is argued that with the possibility of accidental defections, other social mechanisms have limitations, while formal legal systems can generate perverse incentives. Therefore, apologies can serve as a useful economic governance institution.

Keywords: Apology, Sorry, Imperfect Public Monitoring, Uncertainty, Social Norms, Economic Governance, Legal Institutions, Courts, Incentives.

JEL Classification: D08, K04, Z01.

*I am grateful to Jens Prüfer, Eric van Damme, Michael Powell, John Barkley Rosser, Jr., Mark Ramseyer, Alex Teytelboym, Avner Greif and all participants of ISNIE (2015, Harvard), ESNIE (2015, Corsica) and CLEEN (2015, Tilburg) conferences for their insightful comments and feedback. All errors are my own.

[†]CentER, TILEC, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; v.srivastava@uvt.nl

1 Introduction

The word sorry, is an almost universally acknowledged expression of apology. As a regretful acknowledgment of an offence or failure, it is used to express regret, seek forgiveness, reconcile differences and enable cooperation in the future in case of breach of a social contract. An apology is tendered in a multitude of different circumstance to mitigate ill-will and resolve conflicts: from individuals who bump into someone in a crowd, to national governments who have kidnapped aboriginal children in the past.¹ The concept of apology is time honoured and finds expression in many different cultures (Allen, 2006; Al-Zumor, 2011; Bar-Siman-Tov, 2003). If the wide variety of *Sorry greeting-cards* available were not evidence enough, there are firms now who apologise on behalf of their clients for a fee:

*"Tianjin and the central city of Xi'an now boast new, successful apology companies. And apology call-in shows on state radio have also started to appear . . . On behalf of clients, the apologisers write letters, deliver gifts and make explanations."*²

The ubiquitous usage of apologies, its important role in society and complexities have been duly acknowledged and investigations have spawned a vast literature in a host of different disciplines. There is evidence for the role that apologies play in business (Folkes, 1984; Hearit, 1994; Abeler et al., 2010). A vast empirical literature in social and child psychology investigates the various aspects of apologies (Schlenker, 1980; Schlenker and Darby, 1981; Darby and Schlenker, 1982; Ohtsubo and Watanabe, 2009; Ohtsubo et al., 2012) and its impact on forgiveness and justice (Witvliet et al., 2008; Wallace et al., 2008). Other fields of study like linguistics (Cohen, A. D., Olshtain, 1985; Reiter, 2000), history (Trouillot, 2000; Gibney, 2008) and politics (Lind, 2011; Nytagodien and Neal, 2004) have also explored the effect of apologies. Economic experiments have also been used to verify the efficacy of apologies in repairing damaged reputation and enabling reconciliation (Fischbacher and Utikal, 2013; Ho, 2012).

In a departure from the positive exploration and documentation of apologies in other subjects, the legal literature has taken a normative stand of calling for a greater role for apologies in the judicial system of the U.S.A (Wagatsuma and Rosett, 1986; Shuman, 2000; Petrucci, 2002). The virtues of apologies have been evaluated and extolled in keeping down crime and recidivism rates (Haley, 1982, 1995, 1998; Robbennolt, 2003). Its role in providing cathartic relief to affected parties in wide variety of cases including, medical malpractice (Keeva, 1999), divorce (Schneider, 2000) and civil rights abuse (White, 2005), has also been found to be positive. These studies posit apology as a redressal mechanism outside the

¹<https://www.aadnc-aandc.gc.ca/eng/1100100015644/1100100015649>

²<http://www.nytimes.com/2001/01/03/world/tianjin-journal-for-a-fee-this-chinese-firm-will-beg-pardon-for-anyone.html>

formal judicial process. Their argument being that the justice system stands to benefit from the restorative, conciliatory effects that apologies engender if they are incorporated into the judicial process.

Nonetheless, despite its pervasiveness in society, offering an apology or determining when or how much to apologise is no simple matter:

On May 7, 1999, during the NATO bombing of Yugoslavia, five US guided bombs hit the People's Republic of China embassy in the Belgrade district of New Belgrade, killing three Chinese reporters and outraging the public. President Clinton promptly expressed deep regrets and tendered an apology. However, the apology was dismissed as being too casual, delivered while Mr. Clinton was outdoors and wearing a polo shirt. The claim of much of the Chinese media was that *"it was not really an apology at all."* (Stamato, 2008)

The efficacy of an apology, seems to be determined by norms that are also likely to provide a measure of their sincerity. The economic literature on social norms maintains that formal legal institutions are not indispensable to covenants of cooperation (that enable productive economic activities) (Ostrom et al., 1992; Ostrom, 2000; Posner, 1997). The norms that allow for such cooperation, reflect a social contract, an implicit agreement among the members of a society. There is plethora of evidence supporting the role of apology in governing interaction and maintaining social organisation, making it one such social contract and an therefore, an instrument of economic governance.

This paper is an attempt to analyse apologies as a social norm. A game-theoretic framework, of imperfect public monitoring, is used to rationalise the existence and usage of apologies. The questions that structure this analysis are: What in the nature of socio-economic interactions necessitates the need for apology? Is apology an equilibrium outcome? What is the cost of such an apology? What are the characteristics of such an equilibrium? How do the welfare outcomes compare to that in other modes of community governance, like ostracising defaulting members? What are some of the additional constraints that courts face compared to apologies? In answering these questions, this paper contributes to the literature on imperfect public monitoring games by establishing the existence of a sorry equilibrium. In a sorry equilibrium, while the equilibrium strategy is not public ³, a costly apology allows private information to be made public. It also contributes to the legal discussion on apologies by providing a more rigorous analytical framework which enables delineation of the trade-offs inherent in the use of apologies, along with inferences about the limitations of other governance mechanisms.

The remainder of the paper is organised as follows. In the next section, some of the empirical facts that have led to legal scholars emphasising the role of apologies are presented. In section 3 the relevant extant literature on social interaction and economic governance

³It conditions on private information.

is reviewed while the characteristics of apologies are identified. The model is outlined in Section 4; laying out all the assumptions. In section 5 the equilibrium strategy is proposed and it is shown that it constitutes an equilibrium; cost of apology in equilibrium and some comparative statics are discussed. In section 6, some characteristics of the sorry equilibrium are shown. The welfare outcomes of different community enforcement mechanisms are compared to apologies in Section 7. In section 8 a model of formal court governance is outlined, its limitations identified and some policy suggestions are made. Section 9 concludes by summarising the results and making suggestions for future research.

2 USA, India and Japan

The criminal justice system in the United States of America has 6.94 million people or 1 in every 35 American adults under its purview, of which approximately 2 million are incarcerated (Glaze and Herberman, 2012). In fact, the U.S is next only to Seychelles in its prison population rate.⁴ These statistics are staggering, but the socio-political impact of high incarceration rates and the huge burden they put on the taxpayer notwithstanding, these number do not reflect the full scale and scope of the impact that the legal system has on the economy at large.⁵ The tort system in the US has also been under attack for being costly and inefficient (Shuman, 2000). While some claim that it lets corporations get away with huge excesses, others claim that it imposes too high a cost on businesses (Krauss, 2012).

In particular, American courts have gained notoriety in their handling of "second-order law": class actions, sexual harassment claims, or medical malpractice (Ramseyer and Rasmussen, 2010).⁶ Their penchant for awarding large punitive damages in many high profile cases is well documented (Wagatsuma and Rosett, 1986; Ramseyer and Rasmussen, 2010). The unpredictability that stems from such cases can have major repercussions for the economy: *"They can profoundly affect the social relations and economic structure within a coun-*

⁴Number of people incarcerated per capita: http://www.prisonstudies.org/highest-to-lowest/prison_population_rate?field_region_taxonomy_tid=All.

⁵There are also major concerns about the composition of prison population. In 2006, about 60 percent of jail inmates reported having had symptoms of a mental health disorder in the twelve months prior to their imprisonment (James and Glaze, 2006). There are other long term impacts of imprisonment. Many people are returned to jail for non-payment of fines and fees (Evans, 2014). According to another study, individuals who do manage to find work after release earn less on average than their counterparts who have never been incarcerated. Among formerly incarcerated men in the study, 'two-thirds of whom were employed before being incarcerated' hourly wages decreased by 11 percent, annual employment by nine weeks, and annual earnings by 40 percent as a result of time spent in jail or prison (Western and Pettit, 2010).

⁶The authors distinguish between first-order law and second order law. In this classification first order law is pertains to the typical disputes over automobile accidents and contract claims.

try, but not because they are common or because they protect property rights. They affect social relations and economic structure because despite their scarcity they discourage investment and cause firms to take precautions of little social value."(Ramseyer and Rasmusen, 2010, p. 6)

The Indian Judicial System, servicing the world's second most populous country is another massive legal institution. There are about 20 million pending cases in lower courts and 3.2 million such cases in the high courts (Hazra and Micevska, 2004). More recent estimates put the total number closer to 30 million.⁷ 70% of the incarcerated population is under-trial and many for those awaiting trials have been in jail longer than a formal sentence would have required them to be (Krishnan and Raj Kumar, 2010). The costs of such a logjam in the courts, both social and economic, are so large that while often spoken of, hardly any comprehensive and reliable estimates are available.

In India, the limited capacity of the legal system is a binding constraint on the number of cases that can be handled by the courts. In the US, a relatively more efficient legal system with a propensity to grant multi-million dollar damages and commit people to supervision has probably created perverse incentives whereby, it is individually rational for citizens to access the courts even when it is not socially optimal to do so.⁸

The nature of the problems facing the legal systems of the two countries and their sources are disparate, but it is apparent that both the systems require improvements. In contrast to the U.S and India, in the Japanese legal system the *attitude of the accused and their willingness to apologise and confess is crucial to the decision of whether to prosecute or not* (Haley, 1982; Stephens, 2008). In fact, relying on the willingness of the accused to apologise after the offence, at one time in Japan "33% of all cases involving non-traffic related offences were suspended by the prosecution" (Haley, 1982, p.271). Further, even when prosecution did proceed and the defendant was found guilty, the court suspended jail sentences in more than two-thirds of such cases. Apology is therefore used in the Japanese legal system as an informal sanction which reduces the likelihood that a dispute will be taken to court (Haley, 1998).

The Japanese legal system provides a great example of apologies being effective in a wide-variety of scenarios; resolving disputes and fostering less acrimonious attitudes. It has also been associated with the low crime and recidivism rates in Japan (Haley, 1998). But lest the case of Japan be considered an isolated cultural anomaly, there are other examples from all over the world. In the U.S., evidence from a study on settlement offers at the time

⁷<http://timesofindia.indiatimes.com/india/Courts-will-take-320-years-to-clear-backlog-cases-Justice-Rao/articleshow/5651782.cms>.

⁸Optimal here refers to social optimal where (marginal cost)=(marginal benefit) given the law and institutional capacity. This is not to deny the role of supply side activities in altering the optimal level.

of an accident indicates the relevance of apologies. 73% of the respondents would accept the settlement offered if a full apology was tendered compared to 52% when no apology was offered (Robbennolt, 2003). Since 1989, community conferences in New Zealand have dealt with more than half of the juvenile offenders in an alternative system that relies on getting the victim and the offender to meet and utilises apologies for dispute resolution and victim rehabilitation (Scheff, 1998). In Australia the same system is used for both adult and juvenile crimes. In fact, it is claimed that tort plaintiffs often profess that what they really wanted was an apology and brought a suit only when it was not forthcoming (Shuman, 2000).

3 Characterising Apology

The existence of pro-social preferences, norms and social enforcement mechanisms (replying on monitoring and punishments) can allow bi-lateral or multi-lateral interactions to overcome the problems inherent in social dilemmas (Abreu, 1988; Greif, 1993; Sobel, 2006; Dixit, 2009). The mechanism that allows for sustained cooperation in such a setting has usually been modelled in the literature as self-reinforcing equilibria, made possible by repeated interactions between agents⁹. In such models of cooperation without external enforcement, the punishment never gets played out in the equilibrium, the threat exists merely to ensure that players cooperate in equilibrium. This is also true for the role of legal sanctions in law and economics literature, where sanctions are usually put in place to deter cheating (Hermalin et al., 2007).

The equilibrium outcomes of such models, however, are at odds with observations of the day-to-day usage of apologies. They are used rather frequently, both in long-standing relationships as well as in interactions with comparative strangers. The first 2 unique characteristics of an apology that distinguish it from other forms of social enforcement:

(1) The frequent usage of apologies suggests that it constitutes an equilibrium outcome and is not to be relegated to the off-equilibrium path to merely act as deterrence.

(2) In order for apologies to be an equilibrium outcome there must be scope for failures/offences in equilibrium. Moreover given the usage of apologies in a wide-variety of circumstances, the cause of such failures must be sufficiently general to find manifestation in many different types of events.

Studies in social psychology posit that an apology is a plea for the offended to not take a negative event as representative of the intentions and character of the offender (Schlenker and Darby, 1981; Schlenker, 1980; Darby and Schlenker, 1982).

(3) An apology is an attempt to establish a lack of intent on part of the offending party,

⁹In situations where reputations are important, case specific ex-post arrangements (Bull, 1987) in finitely repeated interactions can also lead to cooperative outcomes.

along with being an acknowledgment of harm or offence caused. So, while the harm caused is not disputed, the contention is that any such harm was accidental and not the result of an intentional action. There must therefore be a separation in intent and outcome in the game.

The possibility of allowing for *mistakes* in selection of actions to explain the need for apologies has already been explored in the literature (Fischbacher and Utikal, 2013). It has been shown that in an experimental setting, players use apologies if their intentions cannot be easily inferred. In such situations, the evidence shows that "*An apology is a strong and cheap device to restore social or economic relationships that have been disturbed*" (Fischbacher and Utikal, 2013). However, this particular attempt at explaining the usage of apologies relies on players having other regarding preferences. Such an approach requires additional assumptions about preferences and limits the possibility of exploring why such preferences might have been socialised in the first place.¹⁰

The *separation of intent and outcome* has been also been treated in a more general setting in the economics literature, as games of imperfect public monitoring (Rubinstein, 1979; Fudenberg and Maskin, 1986; Fudenberg et al., 1994). This literature does not rely on additional behavioural assumptions; instead the offended player is responsible for statistically distinguishing between an accidental offence and a deliberate one. There are however limitations to the solutions provided to the problem of accidental defection in this literature. The complications of statistical identification aside, the equilibria proposed either rely on mixed strategies (strategies are not specified, instead, payoffs that can be generated by some mixed strategies are the focus) (Fudenberg and Maskin, 1986; Fudenberg et al., 1994) or players who do not discount (Rubinstein, 1979).

An apology is fundamentally different from these approaches; it requires the offending player to take on the burden of resolution. Apologies, in most social conventions can only be offered, not taken. This implies that an apology is a self-imposed sanction. As an apology must be offered, the offending player can condition his actions on private information. An apology therefore, may be offered only when the player knows that the offence was not deliberate.¹¹ This is a departure from public strategies hitherto used in games of imperfect monitoring.

(4) A village elder can order it or friends can goad for it, but eventually an apology has to be offered by the offender to the offended. This distinguishes it from more conventional punishment that can be met out by some authority or an outside agent. In fact, an apology has a redeeming quality built into it. This characteristic is most visible in a social dictum

¹⁰Therefore it is instructive to identify the usefulness of apologies without pro-social preferences that support its usage.

¹¹In fact, it has been shown that apologising after a deliberate offence can often make the situation worse (Fischbacher and Utikal, 2013; Struthers et al., 2008).

regarding apology that emphasises the need to feel sorry, when an apology is offered.

This *need to feel sorry* can be understood as the cost that needs to be incurred for the apology to serve its purpose.

(5) Apologising is costly. This observation while not self-evident is a critical one. An apology performs the purpose of placating the affected player. If there was no cost attached to apologising, the party causing the harm could always apologise and consequently there is no reason why the affected party should take any apology seriously. So there must be some psychological, social or even monetary costs attached to apologising. This argument is akin to claiming, that in general, apologising is not cheap talk.

This is not to deny that in certain cases, apologies may well be cheap talk. However, cheap talk is useful (and a cheap talk apology likely to be honest), only if the interests of the two players are aligned to a substantial degree (Crawford and Sobel, 1982; Farrell and Rabin, 1996). So, if the objective of the strategic interaction between players is akin to a coordination game, then a cheap talk apology might indeed be useful (Ohtsubo and Watanabe, 2009; Ho, 2012). But the objective of this paper is not to investigate all possible uses of the word sorry in the English language, but to establish the role of apologies as a resolution and reconciliation mechanism. As social dilemmas do not afford the luxury of useful cheap talk, apologies are in general considered to be a costly self-imposed sanction.

Furthermore, even seemingly cheap talk apologies might have costs associated with them. If apologies are useful tools to mitigate conflicts, then it is entirely feasible that preferences favouring their use be socialised in members of a society. This might imply that there may be psychological costs attached with apologising.¹²

(6) There are different levels of sorry. The social psychology literature differentiates between at least 5 different levels of apologies (Schlenker and Darby, 1981). The easiest way to think of this claim is to imagine the varying levels of sincerity with which sorry can be said. A simple sorry, or a profuse apology or even buying a sorry greeting card to convey the apology. This implies that the cost incurred for an apology can change to reflect some underlying differences in the circumstances.

Definition 1 *Apology is a self-inflicted cost, undertaken by the offending party to convince the offended party that any mistakes committed were accidental.*

These accidental infractions or mistakes are caused by the *residual stochasticity* inherent in actions. While the law and economics literature on torts and liability emphasises the level of *due care* (Brown, 1973), *residual stochasticity* refers to that left over component of uncertainty that does not depend on the level of care taken by an individual.

¹²Apologising might involve shame, disappointment, embarrassment and other negative emotions. So, while the model developed here does not rely on pro-social preferences, they might reflect a part or the entirety of the cost of apology posited here.

Some of the previous attempts at modelling apologies have focused on its role as a signal (Ohtsubo and Watanabe, 2009; Ho, 2012). A signalling model necessitates additional assumptions about the differences in preferences across types of players, the choice of which must be contrived. The differences across types might even be superfluous to an understanding of apologies, as the additional insights from a signalling model can be inferred from a simpler model.¹³ This simpler model also allows for a comparison with other social norms and courts, existing models of which do not employ signals to distinguish between types.¹⁴

4 The Model

The setting for this model is an infinitely repeated bilateral game, where the players select their actions simultaneously in every repetition.¹⁵ The stage game for this bilateral interaction is taken to be a simple prisoner’s dilemma with the following actions and payoffs:

Table 1: Stage Game Pay-off Matrix

	Cooperate (C)	Defect (D)
Cooperate (C)	h, h	l, w
Defect (D)	w, l	d, d

Where: $w > h > 0 > d > l$ and $2h > w + l$

The time between two repetitions is taken to be discrete $t \in \{0, 1, 2 \dots \infty\}$. The rate of time preference for each player is assumed to be constant $\delta \in (0, 1)$. The cost of an apology is captured by s . The cost of the apology is borne by the player who apologises. This cost is not borne to compensate the other player, but is considered as utility lost from this two-player system.

The choice of a player is determined by intention. So if a player intends to cooperate, he chooses C. But, the uncertainty inherent in playing the action ensures that the player

¹³If there are two types of players, one prefers to apologise (the nice type) and the other who doesn’t (the bad type). The good type is modelled in subsequent sections. The second type is not explicitly modelled, however, off equilibrium actions of the good type can be inferred to signify those of the bad type.

¹⁴The model developed in the next section is further different from these papers as; unlike (Ho, 2012) cost of apology is endogenous to the model and unlike (Ohtsubo and Watanabe, 2009), the need for apologies is motivated by mistakes.

¹⁵The game can be extended to depict multi-lateral community interaction, wherein a player might not be playing the same player in every repetition of the game, but all potential players in the community know of the outcome of previous iterations of the stage game, with some probability.

cannot ensure that C is actually played out. In effect, the distinction between intention and outcome emphasised earlier is captured by the difference between *chosen action* and *the action that gets played*. This also has consequences for the information available to the other player. The other player can observe the action that gets played out, but has no information about the intention behind it.

The assumptions made to incorporate the role of uncertainty in the model are as follows:

(1) Player 2 can play the action that he intends to play. For instance, if player 2 intends to Cooperate, he can play Cooperate with certainty.

(2) Player 1 cannot ensure with certainty which action will play out. This uncertainty is captured by the parameter p . So, if player 1 chooses to play C, then C gets played with probability p and D gets played with probability $(1 - p)$. Alternately, if player 1 chooses to play D, then D gets played with probability p and C gets played with probability $(1 - p)$.

(3) $1 > p > 0.5$. The second part of the inequality reflects that an action chosen by player 1 is more likely to be played than not.

(4) Player 2 can observe the action that gets played by player 1, not the action that player 1 chooses to play.

These assumptions can be incorporated in the stage game to reflect expected payoffs in every period t :

Table 2: Expected Payoffs in the Stage Game

	Cooperate (C)	Defect (D)
Cooperate (C)	$ph + (1 - p)w, ph + (1 - p)l$	$pl + (1 - p)d, pw + (1 - p)d$
Defect (D)	$pw + (1 - p)h, pl + (1 - p)h$	$pd + (1 - p)l, pd + (1 - p)w$

Notation: E_{CD}^1 is the expected payoff of player 1, when he chooses to play C and player 2 plays D:

$$E_{CD}^1 = pl + (1 - p)d$$

E_{CD}^1 is reflected in the first row, second column of the payoff matrix. The first row, first column contains E_{CC}^1 and E_{CC}^2 .

(5) Sustained cooperation in this game is desirable only if $E_{CC}^2 > E_{DD}^2$, so it is assumed that $p > \frac{w-l}{h+w-d-l}$ ¹⁶.

Timing with-in the stage game in every period t is as follows:

Stage 1: The prisoner's dilemma is played out and payoffs are realised.

Stage 2: Player 1 decides whether to apologise or not.

¹⁶ $0.5 < \frac{w-l}{h+w-d-l}$ as $0 < (h - d) < (w - l)$.

In the next section it will be shown that a strategy with apology constitutes a Nash equilibrium of this game.¹⁷

5 Analysis

The game described here is one of imperfect public monitoring. Both players know the *action that was played* (public information). One of the players (player 1 in this instance) has additional information about the action that was *chosen* (private information). The equilibrium strategy proposed here is a pure strategy that conditions on both public and private information available in the game.¹⁸ This is a departure from the usual practice of using public strategies for games of imperfect public monitoring.¹⁹ The proposed equilibrium strategy is as follows:

Player 1 (Stage 1) - Choose to play C at $t = 0$. In $t \geq 1$:

- Choose to play C if:
 - (C, C) is played out in all $\hat{t} < t$, or
 - Apology costing s^* has been offered in stage 2 $\forall \hat{t} < t$ in which (D, C) was played out.
- Else, choose to play D .

Player 1 (Stage 2) - In $t \geq 0$:

- Offer an apology costing s^* if C has been chosen in stage 1 and (D, C) gets played out.
- Else, do not apologize

Player 2 - Choose to play C at $t = 0$. In $t \geq 1$:

- Play C if:

¹⁷In games of imperfect monitoring, Sub-game perfection (SPE) is not a very useful concept. Further, as apology is conditioned on private information, the equilibrium proposed here cannot be a Public Perfect equilibrium (PPE). However, it will be shown that the equilibrium derived here shares some of the characteristics of SPE and PPE. Further discussion in section 6.

¹⁸A pure strategy defines a specific move or action that a player will follow in every possible attainable situation in a game. Such moves may not be random, or drawn from a distribution, as in the case of mixed strategies.

¹⁹Public strategies that condition only on public information (information available to all players) available in the game.

- (C, C) is played out in all $\hat{t} < t$, or
 - Whenever (D,C) is played out in any $\hat{t} \leq t$, there is an apology costing s^* in stage 2.
- Else, play D.

This strategy is a combination of a form of self-inflicted sanctions (cost of apology) and a trigger strategy. In equilibrium, both players choose to cooperate, player 1 apologises (at the cost s^*) if an accidental defection happens and both players continue to cooperate. The problem is that the choice of apologising is based on private information of player 1.²⁰ It will be shown in this section that the cost of apology required for the strategy to constitute an equilibrium can be determined and used to overcome over this problem. Further, the proposed strategy will constitute an equilibrium if each player is better off choosing this strategy than defecting from it, given that the other player does not deviate. Therefore, equilibrium conditions must ensure that it is incentive compatible for both the players to choose to play C at stage 1 and that apologising when an accidental defection happens, is incentive compatible for player 1 in stage 2.

The assumption of un-observability of intentions creates the need for determining the cost of apology at the margin. This is because an defection is intentional; player 1 can choose to defect and apologise if (D,C) gets played out. This possibility creates the need to determine the appropriate cost of the apology for a given level of uncertainty (as captured by p) and value of the payoffs. These costs should be low enough so that player 1 is willing to incur them, but high enough to deter intentional defection.

Player 1 (Stage 2): The first of the incentive compatibility constraints that has to be met in equilibrium for player 1, is that if player 1 chooses to play C, but D gets played, the payoff from apologising should be higher than the payoff from not apologising.

$$w + \sum_{t=1}^{\infty} \delta^t E_{DD}^1 \leq (w - s) + \sum_{t=1}^{\infty} \delta^t (E_{CC}^1 - (1 - p)s) \quad (1)$$

The left hand side (LHS) of equation 1 is the payoff of player 1 from not apologising once D is played out. The right hand side (RHS) of the equation is the payoff from apologising in the same case. The timing of the game ensures that decision to apologise or not (stage 2) is made after the payoffs are realised (in stage 1) and are therefore known with certainty. Further, the continuation payoff in the LHS is determined by the equilibrium strategy for player 2 which is to play D if there is no apology forthcoming after a (D,C) outcome. Player 1 can't do any better than choosing to play D. Equation 1 simplifies into the first incentive compatibility constraint (IC):

²⁰Player 2 can therefore not verify if player 1 had chosen to cooperate or defect before offering the apology.

IC 1: $s^* \leq \frac{\delta(E_{CC}^1 - E_{DD}^1)}{(1-\delta p)}$, For player 1 to be better off apologising while choosing to play C.

This condition establishes the upper bound for the value of s^* . The cost of an apology must be therefore less than the discounted value of the difference between the cooperative outcome and the non-cooperative outcome. This result is quite intuitive and ensures that the cost of apology cannot be higher than the benefit from it.

Player 1 (Stage 1): The proposed strategy requires that player 1 must be better off choosing to play C and apologising if D gets played, than choosing to play D and apologising if D gets played (assuming IC 1 holds).

$$\sum_{t=1}^{\infty} \delta^t (E_{DC}^1 - ps) \leq \sum_{t=1}^{\infty} \delta^t (E_{CC}^1 - (1-p)s) \quad (2)$$

The LHS of equation 2 is the payoff of player 1 from choosing to play D and apologising when D gets played. The RHS is the payoff from choosing to play C and apologising when D gets played, The major difference is that D is played with probability p in the former and $(1-p)$ in the latter. Equation 2 simplifies into the second incentive compatibility constraint (IC):

IC 2: $s^* \geq \frac{\delta(E_{DC}^1 - E_{CC}^1)}{(2p-1)}$, for the proposed strategy to constitute an equilibrium.

This condition specifies the lower bound for the value of s^* . The cost of an apology should therefore be higher than the gains that player 1 might get from choosing to play D and apologising.

IC 1 and IC 2 capture the trade-offs with respect to the cost of an apology. As these two conditions cannot be contradictory, it must be the case that in equilibrium the following condition should hold:

Lemma 1 (Equilibrium Condition 1) $\frac{(E_{DC}^1 - E_{CC}^1)}{(2p-1)} \leq \frac{\delta(E_{CC}^1 - E_{DD}^1)}{(1-\delta p)}$, for player 1 to play the proposed strategy in equilibrium.

Player 2: In addition to the condition posited in Lemma 1, the parameter for the rate of time preference δ , has to be large enough such that player 2 does not have an incentive to deviate ²¹

$$E_{CD}^2 + \sum_{t=1}^{\infty} \delta^t E_{DD}^2 \leq \sum_{t=1}^{\infty} \delta^t E_{CC}^2 \quad (3)$$

²¹As player 1 cannot choose his actions with certainty: if δ is large enough that player 2 does not deviate due to the threat of the grim trigger, it will be large enough to deter player 1 from choosing to defect in stage 1.

Equation 3 is based on the trigger strategy of player 1. So, if player 2 plays D in any repetition, player 1 will choose to play D in all subsequent repetitions. This leads to the second equilibrium condition.

Lemma 2 (Equilibrium Condition 2) $\delta \geq \frac{(E_{CD}^2 - E_{CC}^2)}{(E_{CD}^2 - E_{DD}^2)}$, for player 2 to play the proposed strategy in equilibrium.

For the proposed strategy profile to constitute an equilibrium, both of the equilibrium conditions must hold.

5.1 Existence

Proposition 1 (Existence) For every p , there exists a δ that satisfies both the equilibrium conditions. Therefore the proposed strategy profile constitutes a Nash Equilibrium in pure strategies.²²

The intuition for the existence of this equilibrium is straightforward. Both the players are better off when both of them choose to play C in every repetition. The cost of apology in equilibrium, is so high that player 1 incurs the cost only if the defection was accidental. Therefore, if player apologizes at a cost of s^* , player 2 is informed that the defection was not deliberate, apology is accepted and player 1 *pardoned*. This ensures that the grim outcome (both players choosing to play D), which serves to deter both players from intentionally defecting, is not triggered accidentally. Apology therefore, offers a resolution mechanism that paves the way for the players to continue choosing to cooperate despite accidental defections.

The apology described by this equilibrium is essentially a kind of truth claim, the veracity of which depends on the cost incurred to make it²³. It also captures all of the characteristics of an apology identified in section 3. Apologies are used in equilibrium and their need is necessitated by residual stochasticity embedded in the stage game. The separation in intent and outcome is sufficiently general to affect a wide variety of activities and explains the frequency of the usage of apologies in diverse contexts and situations. It is self-inflicted, costly and its cost can vary to reflect the underlying payoffs.

²²Proof in the Appendix.

²³Such costs can be incurred in a variety of ways: public declaration accepting the offence caused; spending time and effort to convince player 2; acts of self-sacrifice: a farmer might burn a part of crops, another might physically hurt himself etc.

5.2 Cost of Apology

IC 1 and IC 2 specify the range of s^* , instead of an exact amount. However an exact cost of apology in equilibrium can be determined.

Definition 2 (*Cost of Apology in Equilibrium*): s^{**} is the cost of apology that is incurred by player 1 on the equilibrium path.

Proposition 2 (Cheapest Apology) In equilibrium, $s^{**} = \frac{\delta(E_{DC}^1 - E_{CC}^1)}{(2p-1)} = (w - h)$.

Proof: As an apology is a self-inflicted cost, player 1 would choose the lowest possible value of s^* that it is just enough to convince player 2 that defection was accidental. Therefore, the value of s^* in equilibrium must equal its lower bound derived in IC 2.

This result implies that player 1 needs to give up the entire extra payoff that he gets from accidentally defecting to convince player 2 that the defection was not intentional. Further, while the need for apology is necessitated by uncertainty, the cost of the apology does not depend on the extent of the uncertainty. The other interesting aspect of the cost of apology in equilibrium is that it does not depend on the harm caused to player 2 by player 1's mistake. This is contrary to the common prescription in the law and economics literature (Hermalin et al., 2007).²⁴

Apology vs Compensation: Compensation might be part of an apology and apology might accompany a compensation. But, the equilibrium here highlights the differences between an apology and a compensation mechanism. In a game with no uncertainty, a player using a grim trigger strategy continues to cooperate if the other player has cooperated in all previous periods. In a game with uncertainty, an apology, in equilibrium, ensures that the grim outcome is not triggered, by assuring the other player that the offending player had chosen to cooperate. It is therefore, a forward looking strategy that ensures future cooperation (Ho, 2012). A compensation, on the other hand is meant to be a payment for the damage caused in the accidental defection. It therefore, depends on the harm caused and does not come with an implicit promise of continued cooperation (although an equilibrium can be constructed where it may dissuade intentional defection).

5.3 Comparative Statics

Equilibrium Condition 1 (EQ 1) and 2 (EQ 2) are the constraints that describe the relation between the two parameters p and δ in equilibrium. It would be reasonable to expect that an increase in p , or the certainty with which the action chosen by player 1 gets played would

²⁴The damage payments in law and economics literature are compensation mechanisms, where player 1 pays player 2. An apology as described here, however, does not involve any transfers.

allow a smaller δ to sustain the equilibrium. EQ 2 is in line with this expectation.²⁵ The lesser the chances of player 1 making a mistake, the more potent is the threat of the grim-trigger for player 2. However, EQ 1 mandates that a higher p require larger δ to support the equilibrium.²⁶

This result is counter-intuitive as lower chances of an accidental defection are expected to reduce the importance of the promise of future returns. It results from the constraint posed by IC 1. This constraint implies that if p is higher, ceteris paribus, player 1 has a greater incentive to choose to play C and not apologise. In equilibrium this increased incentive to not apologise is offset by greater concern for larger future returns accrued from apologising.

6 Technical Analysis

The equilibrium described in section 5 requires that the offending player offer an apology, such that the players can continue to choose to cooperate after an accidental defection. However, it is not the only possible equilibrium strategy in which the cost of apology can be incorporated. There can be other strategies involving an apology that may also constitute a Nash equilibrium for the given game.²⁷ For instance, a strategy that requires that player 1 apologise only once for every 2 mistakes and consequently requires that player 2 continue to cooperate after the first mistake knowing that if a second mistake occurs, player 1 will apologise. There can be others that spread the cost of an apology over multiple periods; others still might require that cost of apology be incurred probabilistically (a lottery for apologies).

Definition 3 *A Sorry Equilibrium is a Nash equilibrium of a game of the type described here that requires that the defecting player undertake the (required) cost of an apology, so that both players can continue choosing to cooperate after an accidental defection.*

6.1 Efficiency and Equilibrium Payoffs

The possibility of many different sorry equilibria, raises the question of efficiency. In this context, the measure of efficiency of an equilibrium is how close the equilibrium payoffs are to the best possible outcome of both players choosing to play C in each repetition without any additional cost (in this case that is $p(h + h) + (1 - p)(w + l)$ in every t). The sorry equilibrium analysed in the previous section, imposes a cost of $(w - h)$ in equilibrium, but it

²⁵Proof in Appendix

²⁶Proof in Appendix

²⁷This isn't a claim that they do exist, but a conjecture that they might.

utilizes the simplest possible strategy that involves the use of apologies in this setting. This simplest possible sorry equilibrium has some unique properties.

Proposition 3 (Efficiency) *A sorry equilibrium with s^{**} as the cost (incurred) for each defection, is the most efficient possible sorry equilibrium.* ²⁸

The intuition for this result is that in every t , if the net present expected cost of apology is less than $(w - h)$, then the equilibrium will collapse. The proof relies on the fact that every possible sorry equilibrium is subject to IC 1 and IC 2. This fact implies that the set of equilibria payoffs for all possible sorry equilibrium can be characterised.

Proposition 4 (Sorry Equilibrium Payoffs) *The set of expected equilibrium payoffs in every period of any sorry equilibrium is $\{\pi_1, \pi_2\} = \{ph + (1 - p)(w - s^*), ph + (1 - p)l\}$.* ²⁹

These results show that even though only one possible sorry equilibrium has been analysed here, the results can be generalised to any other sorry equilibrium in a similar setting. Here after, the sorry equilibrium analysed in section 5 is referred to as the sorry equilibrium.

6.2 Stationarity and Perfection

In a game with imperfect monitoring, there are no proper sub-games. This is because, at least one of the players will be uncertain of which information node is he at. Therefore, the notion of sub-game perfection (SPE) is not of much use. On the other hand, the sorry equilibrium posited here, does rely solely on public strategies, so public-perfection (PPE) does not hold either. However, the sorry equilibrium does share some characteristics with SPE and PPE.

Proposition 5 (Stationarity) *In the Sorry Equilibrium, payoffs are stationary, ie. payoffs are the same starting from any period t .* ³⁰

Furthermore, in the sorry equilibrium s^* is a public outcome that can be used to infer private information. So, while the equilibrium strategies used are not all strictly public, s^* allows for replication of another property of SPE and PPE.

Proposition 6 (Perfection) *In the sorry equilibrium, for each t and history h^t , the strategies are a Nash equilibrium from that point on.* ³¹

This implies that strategy prescribed in the Sorry Equilibrium constitutes a Nash Equilibrium for off-equilibrium actions, irrespective of the t at which it is evaluated at. It is a characteristic that it shares with SPE and PPE.

²⁸Proof in Appendix.

²⁹Proof in Appendix.

³⁰Proof in appendix

³¹Proof in appendix

7 Welfare

7.1 Social Welfare in the Sorry Equilibrium

The existence of the sorry equilibrium establishes that an apology can resolve the problems posed by residual stochasticity. However, if such a resolution is not socially beneficial, it might not be desirable. The net benefit to society can be evaluated by calculating the net social welfare in a sorry equilibrium, assuming a simple utilitarian welfare function.

Considering that the costs incurred to undertake an apology do not accrue to the offended parties, but are assumed to be lost to society, the present value of social welfare is:

$$\sum_{t=1}^{\infty} \delta^t (E_{CC}^1 + E_{CC}^2 - (1-p)s^{**}) \quad (4)$$

Lemma 3 (SW-SE) *Social Welfare in the Sorry Equilibrium* = $\frac{(1+p)h+(1-p)l}{1-\delta}$.

It is apparent that p affects social welfare in the sorry equilibrium, which is increasing in p , ceteris paribus. The net social welfare also depends on the harm caused to player 2 (l) due to accidental defections. This condition also illuminates that a sorry equilibrium is not socially desirable where l is very large compared to h . This is in contrast to the cost of an apology, which does not depend on the harm caused.

In cases where an accidental defection might result in great harm like debilitating injury, death or large loss of property, the sorry equilibrium might not be socially desirable. However, the desirability of any outcome can only be determined when compared to alternatives. Some of the other community enforcement mechanisms provide useful benchmarks.

7.2 Social Welfare under Grim-Trigger Strategy

The grim-trigger strategy equilibrium is an often-used theoretical device to enforce cooperate in a social dilemma. The strategy prescribes that in case of any deviation from the cooperative (C,C) outcome, both players start defecting (D,D) forever. It is a simple strategy that works very well to enforce cooperation when both the players can choose their actions with certainty. In this setting, however, accidental deviations happen and set off the grim trigger.

Lemma 4 (SW-GT) *Social welfare with Grim-trigger strategy* = $\frac{2p(h-d)}{1-\delta p} + \frac{2pd+(1-p)(w+l)}{1-\delta}$. ³²

Proposition 7 (SE and GT) *Social welfare in the sorry equilibrium is higher than with grim-trigger strategy if $h(1 + \delta p) + \delta p 2d > w(1 - \delta p)$.* ³³

³²Proof in Appendix.

³³Proof in Appendix.

The higher weight on h is on account of the continued interaction in a sorry equilibrium. w appears in the condition due to accidental defections without provision for an apology. ³⁴ d captures the cost of the trigger strategy. Further, a higher p or δ , ceteris paribus, makes the sorry equilibrium more attractive. A higher p has the dual effect of increasing the probability of the cooperative outcome being played out and reducing the expected cost of apology. A higher δ , implies more patient players, this also favours the sorry equilibrium, due to greater importance for continued cooperation.

Despite the fact that a grim trigger strategy might lead to higher net social welfare under certain parameter configurations, it does suffer from limitations. Player 2, the player who can choose his actions with certainty is worse off playing a grim trigger strategy than the sorry equilibrium ³⁵. This creates an odd situation where a stricter punishment mechanism makes the defector (accidental or otherwise) better off, while the victim suffers more. Further, a grim trigger strategy does not provide a viable alternative to the sorry equilibrium if the cost of accidental defection (l) is very high (Refer Lemma 4).

7.3 Social Welfare under Ostracism

Ostracism is a practice that involves exclusion from social acceptance by general consent. It was practiced in many parts of the ancient world (Masten and Prüfer, 2014, p.379) and is still a norm in some communities across the world (Williams, 1997, 2002). It relies on the threat of exclusion from the community to enforce acceptable behaviour. If adapted to reflect the stage game modelled in section 4, any player who defects would be ostracised from the community. It is also pertinent to point out that being ostracised from a community is likely to be very costly. Ostracism is not only costly to the individual being ostracised; the social costs of ostracism also include the cost to the community from losing a potentially economically productive member. It hinders social interaction and prevents the ostracised individual from participating in economic activities. ³⁶ These costs of ostracism are captured by the parameter O_s and lumped together in the period in which a player defects. ³⁷

Lemma 5 (SW-O) *Social welfare under Ostracism* = $\frac{p2h+(1-p)(w+l-O_s)}{1-\delta p}$. ³⁸

³⁴In the sorry equilibrium, the cost of apology ensures that the payoff for player 1 is h instead of w after an accidental defection Refer Lemma 3.

³⁵This is because $E_{DD}^2 \leq E_{CC}^2$ in the stage game in accordance with Model Assumption (5).

³⁶For instance if the only doctor in a community were to be ostracised, cost to the community could be very high.

³⁷All interaction ceases after a player defects and the game is not played.

³⁸Proof in Appendix.

Proposition 8 (SE and O) *The net social welfare in the sorry equilibrium higher than under ostracism if $h(1 + \delta p) + O_s(1 - \delta) > w(1 - \delta) - l\delta(1 - p)$.*³⁹

The costs involved in the sorry equilibrium are the cost of apology and the cost of accidental defection (due to continued interaction). Therefore, the social welfare from ostracism is higher if w is much larger than h (cost of apology is very high) and/or l is very large and/or O_s is very small. Conversely welfare in the sorry equilibrium is higher if O_s is very large. Given this, barring a situation in which l constitutes a very large loss or the payoff from defection (w) being inordinately large; O_s is likely to be high enough to ensure that a sorry equilibrium leads to a better outcome. This is because O_s reflects the lifetime costs of ostracism, which include the social costs of isolation, the resulting humiliation and the foregone economic opportunities. Further, even if there were no costs of ostracism ($O_s = 0$), the social welfare in the sorry equilibrium would be higher: if $(w - h)$ or l or both would be low.

7.4 Limited Trigger Strategies and Contribute Tit for Tat

If the damage from accidental defection (l) is very high, ostracism ensures that such damages are not suffered repeatedly. However, it shares the limitation of the grim-trigger strategy in its inability to distinguish between deliberate and accidental defections. In both of these cases therefore, once a defection happens, cooperative interaction ceases. There are however other alternatives in the game theory literature that allow for a return to cooperation after a defection.

One example is that of a *Limited* grim-trigger (LGT) strategy. The strategy specifies that in the instance of a defection, both players defect for the following \bar{t} periods and then return to cooperating. This avoids the problem of the standard grim trigger by *limiting* the number of punishment periods⁴⁰. However, it still retains the odd feature of punishing the victim more than in the sorry equilibrium.

The strategy of *Contribute Tit for Tat* (CTFT) (Wu and Axelrod, 1995) resolves this problem. The strategy requires that both players begin by cooperating, they continue doing so unless there is a unilateral defection. After a deviation, the victim defects until a cooperation from other player. The defector cooperates after a defection. After a period of the victim defecting and the offender cooperating, both players continue to cooperate (Sugden, 1986). In this case, the offender is punished (cooperates, payoff l) and the victim is better off (defects, payoff w), before they start cooperating again. However, both CTFT and LGT are

³⁹Proof in Appendix

⁴⁰It is clear that the limited grim trigger would have better welfare consequences than the standard grim-trigger and therefore the sorry equilibrium (under certain parameter configurations).

complicated to implement and require more coordination (could be costly) than the sorry equilibrium. This is because, uncertainty in action selection can cause accidental defections from the prescribed strategy in the punishment phase too. So, if player 1 is to defect for 2 periods (as punishment in LGT), but cooperates instead, the punishment needs to restart. Similarly, in CTFT, as an accidental defection by player 1 might be followed by another accidental defection (instead of cooperation), making it difficult to define the punishment period ex-ante. The sorry equilibrium strategy on the other hand, is simpler and cheaper to implement.

7.5 Apologies and Behavioural Compensation

Thus far it has been assumed that compensation is not an integral part of an apology. However, given the possible benefits of continued cooperation under the sorry equilibrium (eg. learning, h and/or p increases over time), it can be argued that societies might work towards making it more viable. This could be achieved by inculcating preferences that allow for mitigation of the harm suffered by receiving an apology, a kind of *joy of receiving preference*.⁴¹ The economic literature on pro-social preferences propounds: "Societies go to great lengths to instil such preferences in children during their process of socialisation in families, school, and religious establishments, and continue the process in adults." (Dixit, 2009, p.11)

There exists anecdotal evidence to support this argument. The legal literature on apologies, posits apologies as "contributing to the psychological health and well-being of the people involved" (Keeva, 1999). There is also some evidence that "apology is a therapeutic balm" and "helps reduce the victim's anger" (Shuman, 2000). Incorporating these claims into the sorry equilibrium, the cost of apology that is lost to society is: $(1 - k)(w - h)$, where $k \in [0, 1]$ is the proportion of the cost that is directly or indirectly transferred to the other player. While such a mechanism is not essential for the existence of the sorry equilibrium, as shown in section 5 it certainly improves its social welfare consequences, potentially making it more viable in a wider variety of situations.

8 Courts

The laws that courts are charged with implementing vary considerably across jurisdictions and are often very complicated, riddled with many clauses, caveats and exceptions. This makes modelling a formal legal system which fits all observable cases, almost impossible. Nonetheless, replicating a generic simplified form of court enforcement might yield useful

⁴¹This also relates to the behavioural cost of making an apology mentioned in section 3. Such costs imply that even seemingly costless apologies, have a cost associated with them.

insights; assisting in identifying limitations, complications and the possibility of perverse incentives. This section evaluates the limitations of a formal legal system when faced with residual stochasticity.

The efficacy of legal system can be assessed by its ability to deter infractions. The tools available to the courts to enforce cooperative behaviour are: incarceration, compensation (transfer) and fines. Incarcerating offenders shares some of its characteristics with ostracism, so the investigation in this section will be limited to compensation and fines. An ideal court system would therefore, have sanctions that deter both players from deliberately defecting and when an accidental defection does occur, initiate some action to allow for future cooperation. The easiest way to have such a court system would be to have an authority that can read intentions and inflict sanctions of its own accord. However, as this is highly unlikely, hence the need for a model.

8.1 Court Model

The model of courts in this section uses the stage game defined in section 4. The additional assumptions required are borrowed from previous work on modelling courts (Masten and Prüfer, 2014) and are as follows:

(1) Courts cannot take *Suo-Motu* notice of a defection, a suit must be brought before a court by a plaintiff, accusing the defendant.⁴²

(2) If the court rules in favour of the plaintiff, the defendant shall be required to pay damages I . These damage payments will be considered transfers, unless mentioned otherwise.

(3) The parameter $\tau \in [0, 1]$ reflects the probability with which the court will rule in favour of the plaintiff. It can also be thought of as the probability with which the plaintiff is able to satisfy the burden of proof (Masten and Prüfer, 2014).

(4) Courts (like players) cannot know the action that a player chose to play. The role of intent in law is rather complicated and difficult to fully capture in a stylised setting. The simplifying assumption here shall be that while courts cannot distinguish between deliberate and accidental harm; there is a difference in the probability of being convicted between the two. In the following analysis the probability of the court punishing (finding the player guilty) an intentional defection is $\tau \in [0, 1]$, and the probability of an accidental defection being punished is $\hat{\tau} \in [0, 1]$. The intuition behind this difference is that probability of satisfying the burden of proof would change if the defection was accidental.

(5) Both the defendant and the plaintiff are assumed to incur the same litigation cost c in the event of a suit being filed.

(6) The courts are restricted to making type-2 "False-negative" errors.

⁴²Suo motu is a Latin term meaning on its own motion

The last assumption assures that a defecting player has no incentive to file a suit.⁴³ Further, a cooperating player who files a suit against a defector has an expected payoff of $(\tau I - c)$, while the defector will have an expected payoff $(-\tau I - c)$. These payoffs imply that for a cooperating player to have an incentive to file a suit, $\tau I > c$. If the court deems that the minimum damages to deter infractions is T , then the damages awarded must be $I = \max\{T, \frac{c}{\tau}\}$.

8.2 Damages and Litigation Costs

The courts as described here, cannot distinguish between accidental and deliberate defections. Therefore, the damages awarded by such a court cannot account for the difference in probability of conviction across the two cases. Nonetheless, the court will calculate damages to try and dissuade deliberate defections by making the payoffs from choosing to cooperate higher than deliberately defecting.

$$ph + (1 - p)(w - \tau I - c) \geq (1 - p)h + p(w - \tau I - c) \quad (5)$$

The lower bound for the required transfer as determined by such a court can be obtained by rearranging this inequality.

$$T \geq \frac{w - h - c}{\tau} \quad (6)$$

The transfer amount is set such that the payoff from cooperating be higher than from defecting.⁴⁴ Given this T , if $\frac{c}{\tau} > T$, such that $I = T$, then the cost of litigation, $2c$ would be higher than the cost of an apology in the sorry equilibrium.⁴⁵ Therefore, in the rest of this section it is assumed that c is small enough such that $I = T = \frac{w-h-c}{\tau}$.⁴⁶

8.3 Damages and Outcomes

The magnitude of I must be sufficient to deter a deliberate defection. It is clear that I is sufficient to deter player 1 from intentionally defecting iff:

$$ph + (1 - p)(w - \tau I - c) \geq (1 - p)h + p(w - \tau I - c) \quad (7)$$

⁴³As the courts do not make type 1 (false-positives) errors, if a defecting player files a suit he will merely incur a cost of $-c$.

⁴⁴This is the same T that a court which does not take uncertainty into account, ie. assumes that players can choose their actions with certainty, would choose.

⁴⁵ $\frac{c}{\tau} > \frac{w-h-c}{\tau}$, therefore $2c > (w - h)$.

⁴⁶Where, $T = \frac{w-h-c}{\tau}$, the smallest transfer required to maintain cooperation. This result would obviously also hold for any larger T .

The equation assumes that player 2 cooperates, this is because a case brought before the court, only if the other player is cooperating. The appropriate I can be obtained by re-writing equation 7.

$$I^* \geq \frac{(2p-1)}{p(\hat{\tau} + \tau) - \hat{\tau}}(w - h - c) \quad (8)$$

The difference between equations 6 and 8 stems from assumption (4). It highlights the fact that while courts cannot explicitly distinguish between accidental and deliberate defections, there is a difference in the likelihood of getting convicted in the two cases.

The Best Case Scenario: The best case scenario transpires if $\tau = 1$ and $\hat{\tau} = 0$. This implies that a deliberate defection is punished with certainty, while an accidental defection is not punished. In this case equation 8 reduces to:

$$I^* \geq \frac{(2p-1)}{p}(w - h - c) \quad (9)$$

Now, as $\frac{(2p-1)}{p} < 1$, I as chosen by the court will be sufficient to dissuade deliberate defection.

The General Case: If τ and $\hat{\tau}$ are not assigned extreme values, then the positive results of the best case scenario no longer hold.⁴⁷

Lemma 6 (Small Damages) *If $\tau > \hat{\tau}$, then damages I imposed by the court would be too small to deter intentional defections by Player 1.*⁴⁸

The intuition behind this result rests on the fact that when player 2 cooperates, player 1 is better off choosing to defect (when there are no damages). In the best case scenario, defecting deliberately would result in certain punishment and was therefore not a lucrative option. In this case however, the player might be punished even if he chooses to cooperate, while there is no certainty that deliberate defection will be punished. Therefore, the damages required to deter deliberate defections would be higher in this case.

Therefore, if damage I is imposed by courts, player 1 will deliberately defect if player 2 cooperates. However, this might in turn have an effect on the preferred action of player. It might be useful to evaluate the impact of court awarded damages in equilibrium.

⁴⁷If $\tau < \hat{\tau}$, then accidental defections would be more likely to be punished than deliberate defections. And while it might be the case that accidental defectors might end up getting convicted more often than deliberate ones, this outcome is unlikely to be consistent with commonly held ideas of justice. Moreover, the results of this section do not hold for this case. This creates an interesting paradox, whereby if the objective of a court is to deter deliberate defections, then accidental defections should be punished more vigorously (with a higher probability). This of course would create perverse incentives that would reduce willingness to partake in any activity involving uncertainty in choosing actions. However, evaluating such issues is beyond the purview of this paper.

⁴⁸Proof in Appendix.

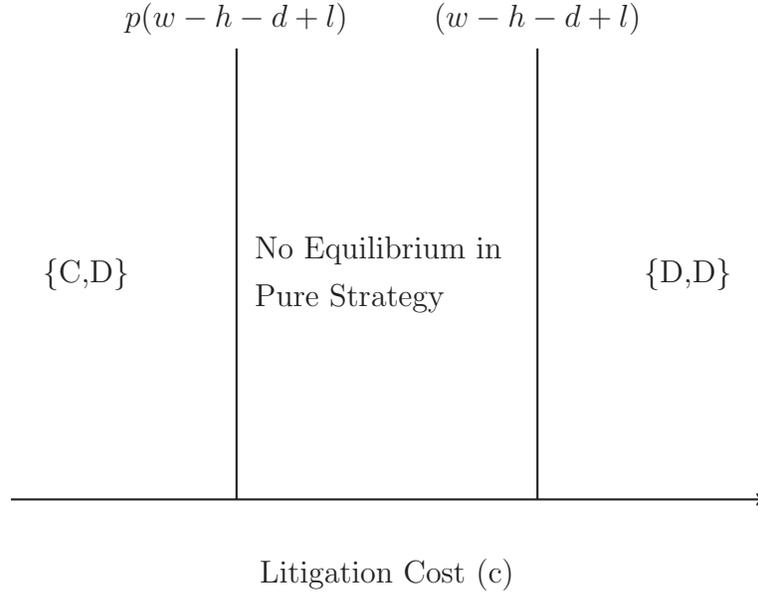


Figure 1: Equilibria in Pure Strategies

Proposition 9 (Equilibrium in One shot Game) *Pure strategy Nash Equilibria of the one shot game:*.⁴⁹

- *If $p(w - h - d + l) \geq c$: Player 2 cooperating and Player 1 choosing to defect is a Nash equilibrium.*
- *Else if $p(w - h - d + l) < c$ and $(w - h - d + l) \leq c$: both players choosing to defect, constitutes a Nash equilibrium.*
- *Else if $p(w - h - d + l) < c$ and $(w - h - d + l) \geq c$: There is no equilibrium in pure strategies.*

This rather odd set of outcomes is the result of the fact that a cooperating player can get damage payments from a defecting player. So, when litigation costs are low player 1 is better off cooperating due to the low cost of getting damages from player 2. on the other hand, player 2 is better off paying the damages with probability τp (incurring c and getting payoff of w with probability p) than cooperating and suffering harm when player 1 defects accidentally. As the cost of litigation rises, player 2 is better off reducing the probability of going to court by cooperating (probability of going to court reduces to $(1 - p)$ in this case). However, if player 2 cooperates, player 1 is better off choosing to defect, thereby increasing his chances of getting damages. Therefore there is no equilibrium at medium level of litigation costs. When the cost of litigation is very high, both players are better off choosing to defect, thereby reducing their probability of going to court.

⁴⁹The proposition assumes that $(w - h - d + l) > 0$. Proof in Appendix.

None of these equilibrium outcomes are desirable. While this exposition of the legal system might seem too simplistic, it happens to reflect some of the elements of the Indian judicial system. Historically, the Indian system has been given in to low compensation and fines (Srinivasan and Eyre, 2007; Galanter, 1985). While there can be no denying that there are other institutional factors at play, insufficient damages, as this model shows, allows for intentional defection thereby, increasing the probability of infractions. Given limited institutional capacity, higher number of infractions would severely limit the efficacy of courts and contribute to the log-jam of the sort that exists in Indian Courts (Hazra and Micevska, 2004).

The most straightforward solution to the problem encountered in previous case, is for courts to award larger damages. However, as the appropriate damages, I^* , depend on the values of τ and $\acute{\tau}$; how large the damages must be, is difficult to determine if the exact values of the parameters are not know.

Lemma 7 (Large Damages) *The smaller the difference between τ and $\acute{\tau}$, the larger the damage payments that are required.* ⁵⁰

As the probability of being convicted rises for an accidental defection and the difference between τ and $\acute{\tau}$ reduces, the damages required to deter deliberate defection will rise. This is because the expected payoff (without any damages) from choosing to defect is greater than choosing to cooperate for player 1 (when player 2 cooperates). Therefore, if the difference between τ and $\acute{\tau}$ is very small, it makes choosing to play D more profitable for player 1. In fact the smaller the difference, the closer a court mechanism is to a grim-trigger strategy or ostracism. ⁵¹

8.4 Large Damages

Instituting larger damage payments to tide over this problem may create additional problems. Very large transfers entail large-scale redistribution and can potentially give rise to other perverse incentives. They can have serious repercussions: individuals or their businesses can go bankrupt, hospitals shut down and companies become averse to experimenting with new products or services. The damages imposed on player 1 might be large enough to prevent the possibility of any future interaction between the players. They might also encourage investment in legal precautions (hire a large team of lawyers). Large damages therefore, could

⁵⁰Proof in Appendix.

⁵¹However, as is obvious, when there is no difference between accidental and deliberate defections in the probability of being punished ($\tau = \acute{\tau}$) then I decided by the court would be sufficient to deter deliberate defections. This would ofcourse be a knife edge solution and therefore has not be discussed in greater detail.

foster either socially unproductive investments or affect the economy directly by limiting the willingness of people to participate in activities that involve some uncertainty.

These claims, while not explicitly derived, reflect some elements of the U.S judicial system. In fact, if an assumption made earlier is relaxed and courts are allowed to make type 1 (false positives) errors too, these problems get exacerbated. The large damages imposed maybe larger than the discounted value of the payoffs from continued interaction and players therefore might have the incentive to sue even if the outcome is (C,C) or even when they themselves have defected. The existence of a legal system that awards large damages therefore, might paradoxically create disincentives for continued cooperation.

An additional side-effect of such a system is its effect on the cost of litigation, c . As large damages have very serious consequences, courts are likely to try to reduce chances of errors. This would imply an increased role for lawyers, lengthy court proceedings and a higher burden of proof (might also have an effect on τ). All of this would adversely affect the welfare effects of the legal system.

8.5 Policy Discussion

This section has presented the difficulties that courts face in deterring intentional defection. However, even if courts can overcome the problems highlighted here; there is another potential limitation in the formal legal system. Damage payments (sufficient to deter intentional defection) are enforced by the courts and received by the offended player with a probability of $\hat{\tau}$. So with a probability of $(1 - \hat{\tau})$, the offended player receives no compensation or validation of his claim that an offence has been committed. In such cases, it is likely that the interaction will cease to continue. Further, it is likely to lead to festering anger and resentment. It is for the reasons of the sort pointed out here that legal scholars have advocated a more active use of apologies in legal systems.

An important function of any legal system is to act as a grievance redressal mechanism that enables continued, productive interaction. Residual stochasticity in choosing actions requires mechanisms that can distinguish between the deliberate and the accidental. Costly apologies offer one such mechanism. It reduces the cost of litigation (apologies rely on self-identification), the need for large damage payments (avoiding undesirable side-effects) and allows for continuation of interaction (rehabilitation and reconciliation).

The sorry equilibrium posited in this paper does not rely on nature of the payoffs involved. The payoffs need not be monetary and could be emotional, psychological too. The model does not distinguish between the two and implicitly assumes that they are interchangeable and valued equally. While, this assertion is not valid in all circumstances, it is important to point out that this assumption implies that an emotionally costly apology works even when the payoffs from the stage game are purely monetary. This is of particular importance as

the courts (specially in the USA) (Wagatsuma and Rosett, 1986) are given in to evaluating the monetary equivalents of emotional distress and degradation, producing absurdly large punitive damage judgements.

Recidivism is another issue of great concern for legal systems. But having a justice system (USA) that sends a lot of people to prisons, coupled with the social stigma associated with incarceration (Rasmusen, 1996), can create a population of people who might have no other option, but to return to their old vocation. The large number of under trials in India are also likely to suffer a similar fate. Japan on the other hand, with a judicial system that uses apologies actively, has very low recidivism rates (Haley, 1982). Allowing court sanctioned apologies for minor misdemeanours, might help better reintegrate perpetrators of accidental defections in both, USA and India.

These observations serve to juxtapose the potential outcomes generated in courts to those generated in the sorry equilibrium. Such a comparison in addition to exposing the limitations of courts, provides a justification for the use of apologies in the Japanese legal system. Given the problems facing the legal system in both the USA and India, it may be prudent to heed calls in the legal literature for seriously looking into the Japanese experience with apologies (Petrucci, 2002). Studies in psychology study have gathered evidence to show that a costly apology is considered sincere across countries (Ohtsubo et al., 2012). The study was carried out in Chile, China, Indonesia, Japan, the Netherlands, South Korea and the U.S. In fact, there are successful programs for victim offender mediation in the USA and these can form the basis for further exploration into incorporating apologies as part of the formal legal system. In India, the *Lok Adalats* (people's courts), a non-adversarial system, has been in operation for sometime now. They provide a great opportunity to integrate formal apologies into the legal structure. Promoting apologies as a legally sanctioned method of conflict resolution, might also have the beneficial effect of alleviating the cost and time constraints on both the judicial systems.⁵² The usage of apologies in Japan offers insights into how apologies can be a low cost judicial tool for offences which are not grievous in nature.

There are, of course, limits to how useful an apology can be. In addition to the conditions explicitly assumed or derived, the model in section 4 makes certain implicit assumptions. Clarifying them is useful in demarcating the limits of usefulness of apologies. The model assumes that even though player, suffer harm (l), his ability to participate in future repetitions of the game is not compromised. In cases involving debilitating injury, torture, death among others, this assumption is not satisfied. It also assumes that the gains from defection

⁵²In one California County in 2002 there were more than 12,000 guilty pleas entered by people who did not have a lawyer. National standards in the U.S, limit felony cases to 150 a year per attorney. Yet felony caseloads of 500, 600, 800 or more are common (Source: National Legal Aid & Defender Association (NLADA)). In India, the courts are understaffed, with too few judges and 26 % of the pending cases being more than 5 years old NCMS (2012).

(*w*) can be identified and that a portion of it could be given up. There maybe cases like those involving adultery, certain types of fraud, among others where the gains from defection cannot be given up. The requirement of repeated interaction is also a rather strict condition. However, further investigation into the behavioural consequences of an apology might also work in one-shot games.⁵³

Apologies alone will not provide the answer to all the problems facing the legal system in the two countries. But, there are additional benefits from reforming the system to fit in apologies, reforms that allow apologies to be complementary to the legal process. For instance, the Japanese legal system is less ambiguous and has clearly defined rules for determining damages (in most cases). This makes opportunistic, frivolous suits less likely and apologies more likely to be accepted⁵⁴ (and therefore made). In Canada, the Ontario Apology Act came into force on April 23, 2009. It mandates that evidence of an apology is not admissible as evidence of fault or liability in any civil, administrative, or arbitration proceeding. In making such changes a legal system is likely to improve its reliability and efficiency, along with making itself more humane.

9 Conclusion

”An apology is the superglue of life. It can repair just about anything.”

Lynn Johnston

Apologies are frequently used in everyday interactions the world over to mitigate and resolve conflict situations. They have even been put to use in seemingly intractable situations, like in the case of the *Truth and Reconciliation Commission* in post-apartheid South Africa. These and many other examples instigate the need for an economic analysis of apology, to establish the extent to which apology is the ‘superglue of life’.

In order to evaluate the usefulness and efficacy of apologies, it might be illustrative to consider a counter-factual: A world in which everyone can choose their action with certainty. In this world, the legal system and other social governance mechanisms like standard grim trigger and ostracism are very effective. In fact, they are so effective in deterring defection that this world would not need any ‘glue’, as nothing ever ‘breaks’. On the contrary, a world where players face uncertainty in choosing actions, undesirable outcomes cannot be avoided. Accidental defections caused by residual stochasticity require a mechanism to reconcile the players, to *glue together what might have broken*. This world, in so far as it more closely resembles the world we live in, can benefit from using apologies.

⁵³As opposed to behavioural antecedents of an apology, that are used to motivate its existence.

⁵⁴The plaintiff will know in which cases he is likely to get large damages and is therefore less likely to go to court in expectation of a windfall gain

Given the observed ubiquitous use of apologies, either our society exists in a constant state of dis-equilibrium or apology must be an equilibrium outcome. The sorry equilibrium posited in section 5, provides evidence for the latter. It shows that a strategy that conditions on private information, can constitute an equilibrium in a game of imperfect public monitoring. Further, it is show that this equilibrium has stationary payoffs and shares the notion of perfection with sub-game perfect equilibria and public-perfect equilibria. Additionally, the equilibrium so described is also shown to be the most efficient sorry equilibrium and the set payoffs for all sorry equilibria are characterised.

The existence of such an equilibrium, does not preclude the possibility of interactions that may not be worth being 'glued'. A mistake in such a scenario could lead to a debilitating injury, death or payoffs from defection that cannot be given away. Given social welfare concerns, it is possible that in such cases, more indiscriminate social governance mechanisms might be more desirable. A social governance mechanism like ostracism for instance would ensure that the game is never played again. But if continued interaction is valued, then as show in section 7, the sorry equilibrium offers a simple and easy to implement strategy.

The possibility to account for the choice of the player (instead of only the outcome) also ensures that apologies do not suffer some of the limitations, that courts might. A formal legal system cannot always distinguish between deliberate and accidental defections. This creates the possibility of legal systems generating perverse incentives, as shown in section 8. In a legal system more like India's, such incentives are likely to encourage more deliberate defections. Large damages of the sort imposed in the U.S.A can lead to undesirable outcomes and can also create perverse incentives. In general, over reliance of the formal legal systems might have costly undesirable side effects. These findings lend credence to the calls in the legal literate to seriously consider better integrating apologies into the formal legal systems, a la Japan.

The efficacy of apologies notwithstanding, in conducting this investigation the model proposed in section 4 makes certain strong assumptions. For instance, the effectiveness of an apology might reduce when used too frequently. A model that accounts for residual and controlled (determined by the extent of care) stochasticity simultaneously might generate more insight into the margins at which apologies are useful. Alternately, a model that explicitly models the dynamic benefits of continued interaction might be more useful determining the true worth of sustained cooperation. A model that frames the offering and acceptance of apologies in the shadow of the legal system, might offer a clearer picture of how they can be accommodated in the legal system. These potential extensions to this model reveal directions for future research to develop a more complete model of apologies.

A Appendix

A.1 Proof of Proposition 1

For the Sorry Equilibrium to exist, both Lemma 1 (EC 1) and Lemma 2 (EC 2) must be satisfied simultaneously:

$$\text{Consider EC 1: } \frac{(E_{DC}^1 - E_{CC}^1)}{(2p-1)} \leq \frac{\delta(E_{CC}^1 - E_{DD}^1)}{(1-\delta p)}$$

$$\text{As, } \frac{(E_{DC}^1 - E_{CC}^1)}{(2p-1)} = \frac{(pw + (1-p)h - ph - (1-p)w)}{(2p-1)} = (w - h), \text{ EQ 1 reduces to}$$

$$(1 - p\delta)(w - h) \leq \delta(ph + (1 - p)w - pd - (1 - p)l)$$

Further simplification leads to

$$\frac{(w - h)}{w - pd - (1 - p)l} \leq \delta \tag{A.1}$$

Now, for equation A.1 to be met and for $\delta \in (0, 1)$

$$\frac{(w-h)}{w-pd-(1-p)l} < 1, \text{ which implies}$$

$$-h < -pd - (1 - p)l$$

As, $h > 0$ and $d, l < 0$, by assumption, equation A.1 always holds. Therefore, $\forall p$ there is always some $\delta \in (0, 1)$ that satisfies EC 1.

$$\text{Consider EC 2: } \delta \geq \frac{(E_{CD}^2 - E_{CC}^2)}{(E_{CD}^2 - E_{DD}^2)}$$

By Assumption (5) in section 4, $E_{CC}^2 > E_{DD}^2$, therefore:

$$\frac{(E_{CD}^2 - E_{CC}^2)}{(E_{CD}^2 - E_{DD}^2)} < 1 \tag{A.2}$$

As the condition A.2 is always met, there is some $\delta \in (0, 1)$ that always satisfied EQ 2. Therefore, for every given p , there must be some $\delta \in (0, 1)$ that satisfies both EC 1 and EC 2.

A.2 Effect of increase of p on EC 1 and EC 2

Consider inequality A.1: a unit increase in p decreases the denominator of the LHS by $(-d + l)$. This is because $0 > d > l$ by assumption. Now, as the numerator does not change, but the denominator decreases, LHS increases. This in turn requires higher δ to support the inequality and therefore EC 1.

Consider EC 2: a unit increase in p , changes the LHS:

$$\text{The numerator: } w - h - d + l$$

$$\text{Denominator: } 2w - 2d$$

Now, $2w - 2d > w - h - d + l$ as, $w - d > l - h$, increase in the denominator is larger than the numerator. Therefore, the LHS decreases as p increases, allowing smaller δ to support the inequality.

A.3 Proof of Proposition 3

The Pareto efficient outcome of this game is (E_{CC}^1, E_{CC}^2) . In the sorry equilibrium there is a cost of $(w - h)$ is imposed on player 1 and therefore it is bounded away from efficiency. However, there can be other equilibria in which a costly apology (of the type described here) can be used to sustain cooperation in an infinitely repeated game.

Consider: A strategy in which the cost that has to be undertaken when offering an apology is s . And let the Net Present Expected (NPE) cost that incurs (to be undertaken when offering an apology) to player 1 in $t = T$, due to a defection in $t = T$, be x :

So, if the equilibrium strategy being considered mandates an apology every 2 defections, then $x = \frac{NPE(s)}{2}$, where $NPE(s)$ is the net present expected value of s in T . If the equilibrium strategy instead, requires the cost of the apology for a defection in T to be undertaken in $T + 1$ and $T + 2$, then (assuming no further defections in these 2 periods), $x = (\delta s + \delta^2 s)$. x is therefore the total discounted (expected) cost of each defection that player 1 makes in the game. In the sorry equilibrium posited in this paper, x is the same as s .

If x is so defined, then in any sorry equilibrium, it must be sufficient to deter even a single deliberate defection. As x is the cost of every defection, to accomplish this, it must meet 2 conditions:

$$w + \sum_{t=1}^{\infty} \delta^t E_{DD}^1 \leq (w - x) + \sum_{t=1}^{\infty} \delta^t (E_{CC}^1 - (1 - p)x) \quad (\text{A.3})$$

Such that it is not too costly: if player 1 chooses to play C, but D gets played, the payoff from apologising must be higher than the payoff from not apologising. And,

$$\sum_{t=1}^{\infty} \delta^t (E_{DC}^1 - px) \leq \sum_{t=1}^{\infty} \delta^t (E_{CC}^1 - (1 - p)x) \quad (\text{A.4})$$

Such that it is not too cheap: player 1 must be better off choosing to play C and apologising, than choosing to play D and apologising if D gets played.

But conditions A.3 and A.4 are the same as equations 1 and 2 in section 5. Thus, every sorry equilibrium must have an x such that it meets the upper and lower bounds of s^* . As x is the cost imposed on each defection in all possible sorry equilibrium, the most efficient, lowest cost sorry equilibrium, must have, $x = s^{**}$.

A.4 Proof of Proposition 4

As shown in A.3, all possible sorry equilibrium must the cost of defection, x equal to s^* . Therefore, payoffs in every sorry equilibrium must be such that player 1 incurs a cost s^*

after every defection. The set of every sorry equilibrium is the same as the set of payoffs in the sorry equilibrium posited in this paper. Therefore, in every sorry equilibrium, expected payoffs for both player in every t :

$$\{\pi_1, \pi_2\} = \{ph + (1 - p)(w - s^*), ph + (1 - p)l\}$$

A.5 Proof of Proposition 5

The payoff for a stationary strategy is the infinite sum of the (expected) payoffs achieved at each stage. Given the strategy for the sorry equilibrium, the expected payoffs of player 1 in each repetition of the stage game is:

$$t = 0: ph + (1 - p)(w - s^{**})$$

$$t = 1: ph + (1 - p)(w - s^{**})$$

$$t = 2: ph + (1 - p)(w - s^{**})$$

$$t = 3: ph + (1 - p)(w - s^{**})$$

and so on ... (Where $s^{**} = (w - h)$)

The expected payoffs of player 2 in each repetition of the stage game is:

$$t = 0: ph + (1 - p)l$$

$$t = 1: ph + (1 - p)l$$

$$t = 2: ph + (1 - p)l$$

$$t = 3: ph + (1 - p)l$$

and so on ...

Therefore, the expected payoff in equilibrium, for each player is same in every repetition. As payoffs in every repetition are the same and the game is infinitely repeated, it does not matter is payoffs are calculated starting at $t = 0$ or $t = T$.

A.6 Proof of Proposition 6

Perfection requires that each $t = T$, the strategy prescribed by the equilibrium strategy constitutes a Nash equilibrium starting at T for each possible history for the game till T . For any history that on the equilibrium path, this is trivial. Therefore the histories evaluated here will be off equilibrium path.

Both players choosing to play D in every repetition (trivial) and the Sorry Equilibrium, both constitute a Nash equilibrium of the game. All deviations from the equilibrium strategy of the game possible at some $\hat{t} < T$ and the strategy prescribed at T by the equilibrium strategy are:

1. In atleast one $\hat{t} < T$, player 2 plays D: Player 1 will choose to play D in all $t > T$, as

(C,C) has not been played in all previous repetitions. Player 2 will also play D in all $t > T$. As both players choose to play D, the strategy prescribed by the equilibrium strategy constitutes a Nash Equilibrium.

2. In atleast one $\hat{t} < T$, (D,C) gets played and player 1 does not apologise or apologises at a cost $s \neq s^*$: Player 2 will choose to play D in all $t > T$, neither as (C,C) has not been played in all previous repetitions nor every (D,C) has been followed by an apology at cost s^* . Player 1 will also play D in all $t > T$. As both players choose to play D, the strategy prescribed by the equilibrium strategy constitutes a Nash Equilibrium.
3. In some $\hat{t} < T$, player 1 chooses to play D, (D,C) gets played, but player 1 follows (D,C) outcome in every $\hat{t} < T$ with an apology at cost s^* : Player 2 cannot know if player 1 has chosen to play D in any repetition. As each (D,C) outcome has been followed with an apology at cost s^* , player 2 takes T as $t = 0$ and continues to cooperate. For player 1, as an apology costing s^* has been offered $\forall \hat{t} \leq T$, he also takes T as $t = 0$ and continues to cooperate. In essence irrespective of the history, the strategy of both players at T (as prescribed by the equilibrium strategy) constitutes a sorry equilibrium (and therefore a Nash equilibrium) $\forall t \geq T$.
4. In atleast one $\hat{t} < T$, player 1 offers an apology at some cost s after a (C,C) outcome: The strategy of both players at T constitutes a sorry equilibrium (and therefore a Nash equilibrium) $\forall t \geq T$.

Additionally, for any history of the game that includes deviations (1) or (2) or both, the prescribed strategy requires that both players choose to defect in all future repetitions. For deviation (3) and (4), it is assumed that when one of the player deviates, the other player does not. For instance, if player 1 deviates as in (3), but player 2 also deviates as in (1), then the prescribed strategy at T is to choose to defect in all $t \geq T$.

A.7 Grim Trigger Strategy

Given this strategy, the expected payoffs in each repetition of the stage game if both player 1 and 2 choose to play C, will be:

$$t = 0: p(h + h) + (1 - p)(w + l)$$

$$t = 1: p^2(h + h) + p(1 - p)(w + l) + (1 - p)K$$

$$t = 2: p^3(h + h) + p^2(1 - p)(w + l) + (1 - p^2)K$$

$$t = 3: p^4(h + h) + p^3(1 - p)(w + l) + (1 - p^3)K$$

and so on ... (Where $K = E_{DD}^1 + E_{DD}^2$)

Proof of Lemma 4

Using the stream of payoffs above and a utilitarian social welfare function, the social welfare from this strategy is:

$$\sum_{t=0}^{\infty} \delta^t p^t (p2h) + \sum_{t=0}^{\infty} \delta^t p^t ((1-p)(w+l)) + \sum_{t=0}^{\infty} \delta^t ((1-p^t)K)$$

As $\sum_{t=0}^{\infty} \delta^t p^t = \frac{1}{1-\delta p}$; $\sum_{t=0}^{\infty} \delta^t ((1-p^t)K) = \sum_{t=0}^{\infty} \delta^t K - \sum_{t=0}^{\infty} \delta^t p^t K$ and $K = E_{DD}^1 + E_{DD}^2 = pd + (1-p)l + pd + (1-p)w$. Rearranging terms gives the same term as Lemma 4 :

$$\frac{2p(h-d)}{1-\delta p} + \frac{2pd+(1-p)(w+l)}{1-\delta}$$

Both the players start by choosing to play C at $t = 0$ as the following 2 conditions are met⁵⁵:

1. As there is always some δ that meets EQ 2, Payer 2 is better off playing C than D.
2. Player 1 also chooses to play C as payoff from choosing C is higher than D

Payoff from choosing C (PC): $\sum_{t=0}^{\infty} \delta^t p^t (ph + (1-p)w) + \sum_{t=0}^{\infty} \delta^t (1-p^t) E_{DD}^1$

Payoff from choosing D (PD): $\sum_{t=0}^{\infty} \delta^t (1-p^t) ((1-p)h + pw) + \sum_{t=0}^{\infty} \delta^t (1 - (1-p^t)) E_{DD}^1$

Now, $PC > PD$ if $\delta \geq \frac{w-h}{w-E_{DD}^1}$. As $h > E_{DD}^1$, there is some $\delta \in (0, 1)$ for which both players choose to play C at $t = 0$.

Proof of Proposition 7

$$(NSW - SE) - (NSW - GT) = \frac{p2h+(1-p)[(w+l)-(w-h)]}{1-\delta} - \left[\frac{2p(h-d)}{1-\delta p} + \frac{2pd+(1-p)(w+l)}{1-\delta} \right]$$

Where $(w-h)$ is the cost of apology in the sorry equilibrium, therefore the total payoff when an accidental defection happens (in sorry equilibrium) is $[(w-l) - (w-h)]$. To determine the condition:

$$\frac{p2h+(1-p)[(w+l)-(w-h)]}{1-\delta} > \left[\frac{2p(h-d)}{1-\delta p} + \frac{2pd+(1-p)(w+l)}{1-\delta} \right]$$

As $(w+l)$ has the same weights on both side, cancelling them out and grouping terms:

$$\frac{p2h-p2d}{1-\delta} - \frac{p2h-p2d}{1-\delta p} > \frac{(1-p)(w-h)}{1-\delta}$$

This is the same as:

$$(p2h - p2d) \left[\frac{1}{1-\delta} - \frac{1}{1-\delta p} \right] > \frac{(1-p)(w-h)}{1-\delta}$$

Simplifying the expression gives:

$$\frac{2p(h-d)\delta(1-p)}{(1-\delta)(1-\delta p)} > \frac{(1-p)(w-h)}{1-\delta}$$

Cancelling out and re-arranging the terms, gives the expression in Proposition 4:

$$h(1 + \delta p) + \delta p 2d > w(1 - \delta p)$$

⁵⁵This paper does not posit grim trigger strategies as an equilibrium for this game.

A.8 Ostracism

The expected payoffs in each repetition of the stage game if player 2 plays C and player 1 chooses to play C, assuming that the long term costs of being ostracised is lumped together in one parameter O_s to be realised only in the period in which a player defects:

$$t = 0: p(h + h) + (1 - p)(w + l - O_s)$$

$$t = 1: p^2(h + h) + p(1 - p)(w + l - O_s)$$

$$t = 2: p^3(h + h) + p^2(1 - p)(w + l - O_s)$$

$$t = 3: p^4(h + h) + p^3(1 - p)(w + l - O_s)$$

and so on ...

Proof of Lemma 5

Using the stream of payoffs above and a utilitarian social welfare function, the social welfare from this strategy is:

$$\sum_{t=0}^{\infty} \delta^t p^t (p2h + (1 - p)(w + l - O_s))$$

As $\sum_{t=0}^{\infty} \delta^t p^t = \frac{1}{1 - \delta p}$, re-arranging terms gives the same term as Lemma 5 :

$$\frac{p2h + (1 - p)(w + l - O_s)}{1 - \delta p}$$

Assuming the cost of ostracism, O_s accrues to the player who has defected, both players start by choosing to play C at $t = 0$ as the following 2 conditions are met ⁵⁶:

1. If player 2 plays D, the game ends with certainty at $t = 0$. Payer 2 is better off playing C than D, as it is assumed that O_s is large enough, so that the following condition is met:

$$\frac{ph + (1 - p)l}{1 - \delta p} \geq p(w - O_s) + (1 - p)(d - O_s)$$

2. Player 1 also chooses to play C as payoff from choosing C is higher than D as the following condition is met:

$$\frac{ph + (1 - p)(w - O_s)}{1 - \delta p} \geq \frac{(1 - p)h + p(w - O_s)}{1 - \delta(1 - p)}$$

This condition reduces to: $\delta \geq \frac{w - O_s - h}{w - h}$, which is always met as $h > 0$ by assumption.

This there is some $\delta \in (0, 1)$ for which both players choose to play C at $t = 0$.

Proof of Proposition 8

$$(NSW - SE) - (NSW - O) = \frac{p2h + (1 - p)[(w + l) - (w - h)]}{1 - \delta} - \frac{p2h + (1 - p)(w + l - O_s)}{1 - \delta p}$$

⁵⁶This paper does not posit ostracism as an equilibrium for this game.

Where $(w - h)$ is the cost of apology in the sorry equilibrium, therefore the total payoff when an accidental defection happens (in sorry equilibrium) is $[(w - l) - (w - h)]$. To determine the condition:

$$\frac{p2h+(1-p)[(w+l)-(w-h)]}{1-\delta} > \frac{p2h+(1-p)(w+l-O_s)}{1-\delta p}$$

Taking $p2h + (1 - p)[(w + l)]$ common:

$$[p2h + (1 - p)[(w + l)]\left[\frac{1}{1-\delta} - \frac{1}{1-\delta p}\right] > \frac{(1-p)(w-h)}{1-\delta} - \frac{(1-p)O_s}{1-\delta p}$$

Simplifying by cancelling out the denominators:

$$[p2h + (1 - p)[(w + l)]\delta(1 - p) > (1 - \delta p)(1 - p)(w - h) - (1 - \delta)(1 - p)O_s$$

Cancelling out and re-arranging the terms, gives the expression in Proposition 8:

$$h(1 + \delta p) + O_s(1 - \delta) > w(1 - \delta) - l\delta(1 - p)$$

A.9 Courts

Damages imposed by a court = $I = \frac{w-h-c}{\tau}$

Damages required to deter deliberate defection = $I^* = \frac{(2p-1)}{p(\acute{\tau}+\tau)-\acute{\tau}}(w - h - c)$

Proof of Lemma 6

Both I and I^* have $(w - h - c)$ in the numerator. Therefore it must be shown that $\frac{(2p-1)}{p(\acute{\tau}+\tau)-\acute{\tau}} > \frac{1}{\tau}$. Assuming that the inequality is correct and cross multiplying the denominators:

$$(2p - 1)\tau > p(\acute{\tau} + \tau) - \acute{\tau}$$

Grouping common terms:

$$(2p - 1 - p)\tau > \acute{\tau}(p - 1)$$

Cancelling out $(p - 1)$, this leads to:

$$\tau > \acute{\tau}$$

But this is true by assumption, therefore $\frac{(2p-1)}{p(\acute{\tau}+\tau)-\acute{\tau}} > \frac{1}{\tau}$ holds true.

Proof of Proposition 7

From equation 7 and Lemma 6 we know that player 1 chooses to play D, if player 2 plays C. So, if it shown that player 2 will indeed play C, then this would constitute a mutual best response and therefore a Nash Equilibria in pure strategies.

To show: If $p(w - h - d + l) \geq c$, player 2 plays C, when player 1 plays D

$$\text{Player 2 will play C if: } E_{DC}^2 + p\tau I \geq E_{DD}^2 - (1 - p)\tau I$$

This is because, if player 1 chooses to play D and player 2 plays C; player 1 defects with probability p and damages are awarded to player 2 with probability $p\tau$. If player 2 also plays D, player 1 cooperates with probability $(1 - p)$ then damages will be awarded to player 1 with probability $(1 - p)\tau$ (accidental cooperation). Inputting the payoffs from the stage game:

$$p(l + \tau I) + (1 - p)h \geq pd + (1 - p)(w - \tau I)$$

Simplifying:

$$\tau I \geq p(d - l) + (1 - p)(w - h)$$

As $I = \frac{w-h-c}{\tau}$:

$$(w-h) - c \geq p(d-l) + (w-h) - p(w-h)$$

This leads to the required condition:

$$p(w-h-d+l) \geq c$$

To show: If $p(w-h-d+l) < c$ and $(w-h-d+l) \leq c$, player 2 plays D, when player 1 plays D

As shown previously, if $p(w-h-d+l) < c$, player 2 is better off playing D, when player 1 chooses to play D. So if it can be shown that player 1 is also better off choosing to play D, when player 2 plays D, then this would constitute a mutual best response and therefore a Nash Equilibria in pure strategies.

$$\text{Player 1 will play D if: } E_{DD}^1 + (1-p)\tau I \geq E_{CD}^1 + p\tau I$$

This is because, if player 2 plays D and player 1 plays C; player 1 cooperates with probability p and wins damages with probability $p\tau$. If player 1 also plays D, player 1 cooperates with probability $(1-p)$ and wins damages will be awarded to player 1 with probability $(1-p)\tau$. Inputting the payoffs from the stage game:

$$pd + (1-p)(l + \tau I) \geq p(l + \tau I) + (1-p)d$$

Using the value of I and simplifying:

$$d(2p-1) + l(1-2p) \geq (w-h-c)(2p-1)$$

Cancelling $(2p-1)$ and re-arranging terms, this leads to the required condition:

$$(w-h-d+l) \leq c$$

To show: If $p(w-h-d+l) < c$ and $(w-h-d+l) > c$, there is no Nash equilibrium in pure strategies.

As shown previously, if $p(w-h-d+l) < c$, player 2 is better off playing D, when player 1 chooses to play D. But if, $(w-h-d+l) > c$, player 1 is better off choosing to play C when player 2 plays D. If player 1 chooses to play C, player 2 plays D if: $E_{DD}^2 - p\tau I \geq E_{CC}^2 + (1-p)\tau I$

$$pw + (1-p)d - p\tau I \geq ph + (1-p)l + (1-p)\tau I$$

Using the value of I and simplifying:

$$c \geq (1-p)(w-h-d+l)$$

But this condition can never be met as, $(w-h-d+l) > c$ and $(1-p) < 1$. So, player 2 is better off playing C, if player 1 chooses to play C. But if player 2 plays C, player 1 is better off playing D (by equation 7). But if player 1 chooses to play D and as $p(w-h-d+l) < c$, player 2 is better off playing D. But if player 2 plays D and $(w-h-d+l) > c$, player 1 is better off choosing to play C. Therefore, there is no Nash Equilibrium in pure strategies.

Proof of Lemma 7

From equation 8: $I^* = \frac{(2p-1)}{p(\hat{\tau}+\tau)-\hat{\tau}}[w - h - c]$

Consider $\frac{1}{p(\hat{\tau}+\tau)-\hat{\tau}}$: In order for the difference between τ and $\hat{\tau}$, either τ reduces or $\hat{\tau}$ or both.

If τ decreases: the denominator decreases and the total value of I^* increases

If $\hat{\tau}$ increases (by a unit): Change in denominator is $(p - 1)$, but as $p < 1$, the denominator decreases and value of I^* increases

If both change simultaneously, (as $\tau > \hat{\tau}$) both effects reinforce each other and the denominator decreases and value of I^* increases.

References

- Abeler, J., Calaki, J., Andree, K., & Basek, C. (2010). The power of apology. *Economics Letters*, 107(2), 233-235.
- Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica: Journal of the Econometric Society*, 383-396.
- Al-Zumor, A. W. Q. G. (2011). Apologies in Arabic and English: An inter-language and cross-cultural study. *Journal of King Saud University-Languages and Translation*, 23(1), 19-28.
- Allen, D. W., & Reed, C. G. (2006). The duel of honor: screening for unobservable social capital. *American Law and Economics Review*, 8(1), 81-115.
- Bar-Siman-Tov, Y. (Ed.). (2003). *From conflict resolution to reconciliation*. Oxford University Press.
- Brown, J. P. (1973). Toward an economic theory of liability. *The Journal of Legal Studies*, 323-349.
- Bull, C. (1987). The existence of self-enforcing implicit contracts. *The Quarterly Journal of Economics*, 147-159.
- Cohen, A. D., & Olshtain, E. (1985). Comparing apologies across languages. *In Scientific and humanistic dimensions of language*, 175-83.
- Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431-1451.
- Darby, B. W., & Schlenker, B. R. (1982). Children's reactions to apologies. *Journal of Personality and Social Psychology*, 43(4), 742.
- Dixit, A. (2009). Governance institutions and economic activity. *The American Economic Review*, 3-24.
- Evans, D. N. (2014). *The Debt Penalty ? Exposing the Financial Barriers to Offender Reintegration*. New York, NY: Research & Evaluation Center, John Jay College of Criminal Justice, City University of New York.
- Farrell, J., & Rabin, M. (1996). Cheap talk. *The Journal of Economic Perspectives*, 103-118.
- Fischbacher, U., & Utikal, V. (2013). On the acceptance of apologies. *Games and Economic Behavior*, 82, 592-608.
- Folkes, V. S. (1984). Consumer reactions to product failure: An attributional approach. *Journal of Consumer Research*, 398-409.
- Fudenberg, D., David, L., and Maskin, E. (1994). The Folk Theorem with Imperfect Public Information. *Econometrica*, 62(5):997-1039.
- Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3), 533-554.

Galanter, M. (1985). Legal Torpor: Why So Little Has Happened in India After the Bhopal Tragedy. *Tex. Int'l LJ*, 20, 273.

Gibney, M. (2008). *The age of apology: Facing up to the past*. University of Pennsylvania Press.

Glaze, L. E. and Herberman, E. J. (2012). *Bureau Of Justice Statistics: Correctional Populations in the United States, 2012*. Washington DC: U.S. Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics. Retrieved from: <http://www.bjs.gov/content/pub/pdf/cpus12.pdf>

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *The American economic review*, 525-548.

Haley, J. O. (1982). Sheathing the sword of justice in Japan: An essay on law without sanctions. *Journal of Japanese Studies*, 265-281.

Haley, J. O. (1995). *Authority without power*. Oxford University Press.

Haley, J. O. (1998). Apology and Pardon Learning From Japan. *American Behavioral Scientist*, 41(6), 842-867.

Hazra, A. K., & Micevska, M. B. (2004). *The problem of court congestion: Evidence from Indian lower courts*. (No. 88) ZEF discussion papers on development policy.

Hearit, K. M. (1994). Apologies and public relations crises at Chrysler, Toshiba, and Volvo. *Public Relations Review*, 20(2), 113-125.

Hermalin, B. E., Avery, W. K., and Craswell, R. (2007). Contract Law. In Polinsky, A. and Shavell, S., editors, *Handbook of Law and Economics, Vol. 1.*, pages 3?138. Elsevier, Amsterdam.

Ho, B. (2012). Apologies as signals: with evidence from a trust game. *Management Science*, 58(1), 141-158.

James, D. J. and Glaze, L. E. (2006). *Bureau of Justice Statistics: Mental Health Problems of Prison and Jail Inmates*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Retrieved from: <http://www.bjs.gov/content/pub/pdf/mhppji.pdf>

Keeva, S. (1999). Does Law Mean Never Having to Say You're SORRY? Going to trial over a case is costly, frustrating?and can perhaps be avoided with a simple apology. *ABA Journal*, 64-95.

Krauss, M. I. (2012). Tort Law, Moral Accountability, and Efficiency: Reflections on the Current Crisis. *Journal of markets & morality*, 2(1).

Krishnan, J. K., & Kumar, C. R. (2010). Delay in process, denial of justice: the jurisprudence and empirics of speedy trials in comparative perspective. *Geo. J. Int'l L.*, 42, 747.

Lind, J. (2011). *Sorry states: Apologies in international politics*. Cornell University Press.

Masten, S. E., & Prfer, J. (2014). On the evolution of collective enforcement institutions: communities and courts. *The Journal of Legal Studies*, 43(2), 359-400.

National Court Management Systems Committee. *National Court Management Systems (NCMS) Policy & Action Plan*, New Delhi: Supreme Court of India. Retrieved from:<http://supremecourtsofindia.nic.in/ncms27092012.pdf>

Nytagodien, R., & Neal, A. (2004). Collective trauma, apologies, and the politics of memory. *Journal of Human Rights*, 3(4), 465-475.

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, 30(2), 114-123.

Ohtsubo, Y., Watanabe, E., Kim, J., Kulas, J. T., Muluk, H., Nazar, G., Wang, F., and Zhang, J. (2012). Are costly apologies universally perceived as being sincere? A test of the costly apology-perceived sincerity relationship in seven countries. *Journal of Evolutionary Psychology*, 10(4), 187-204.

Ostrom, E. (2014). Collective action and the evolution of social norms. *Journal of Natural Resources Policy Research*, 6(4), 235-252.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a Sword: Self-governance Is Possible. *American Political Science Review*, 86(02), 404-417.

Petrucci, C. J. (2002). Apology in the criminal justice setting: Evidence for including apology as an additional component in the legal system. *Behavioral sciences & the law*, 20(4), 337-362.

Posner, R. A. (1997). Social norms and the law: An economic approach. *The American economic review*, 365-369.

Ramseyer, J. M., & Rasmusen, E. (2010). *Comparative litigation rates*. Harvard Law School.

Rasmusen, E. (1996). Stigma and self-fulfilling expectations of criminality. *Journal of Law and Economics*, 519-543.

Reiter, R. M. (2000). *Linguistic politeness in Britain and Uruguay: A contrastive study of requests and apologies (Vol. 83)*. John Benjamins Publishing.

Robbennolt, J. K. (2003). Apologies and legal settlement: an empirical examination. *Michigan Law Review*, 460-516.

Rubinstein, A. (1979). An optimal conviction policy for offenses that may have been committed by accident. In *Applied Game Theory* (pp. 406-413). Physica-Verlag HD.

Scheff, T. J. (1998). Community conferences: Shame and anger in therapeutic jurisprudence. *Rev. Jur. UPR*, 67, 97.

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations* (pp. 21-43). Monterey, CA: Brooks/Cole Publishing Company.

Schlenker, B. R., & Darby, B. W. (1981). The use of apologies in social predicaments. *Social Psychology Quarterly*, 271-278.

- Schneider, C. D. (2000). What it means to be sorry: The power of apology in mediation. *Mediation Quarterly*, 17(3), 265-280.
- Shuman, D. W. (1999). The Role of Apology in Tort Law. *Judicature*, 83, 180.
- Sobel, J. (2006). For better or forever: formal versus informal enforcement. *Journal of Labor Economics*, 24(2), 271-297.
- Srinivasan, M., & Eyre, M. J. (2007). Victims and the criminal justice system in India: Need for a paradigm shift in the justice system. *Temida*, 10(2), 51-62.
- Stamato, L. (2008). Peace and the Culture and Politics of Apology. *Peace Review: A Journal of Social Justice*, 20(3), 389-397.
- Stephens, M. A. (2008). I'm Sorry: Exploring the Reasons Behind the Differing Roles of Apology in American and Japanese Civil Cases. *Widener L. Rev.*, 14, 185.
- Struthers, C. W., Eaton, J., Santelli, A. G., Uchiyama, M., & Shirvani, N. (2008). The effects of attributions of intent and apology on forgiveness: When saying sorry may not help the story. *Journal of Experimental Social Psychology*, 44(4), 983-992.
- Sugden, R. (1986). *The economics of rights, co-operation and welfare*(pp. 58-62). Oxford: Blackwell.
- Trouillot, M. R. (2000). Abortive rituals: Historical apologies in the global era. *Interventions*, 2(2), 171-186.
- Wagatsuma, H., & Rosett, A. (1986). The implications of apology: Law and culture in Japan and the United States. *Law and Society Review*, 461-498.
- Wallace, Harry M., Julie Juola Exline, and Roy F. Baumeister. Interpersonal consequences of forgiveness: Does forgiveness deter or encourage repeat offenses?. *Journal of Experimental Social Psychology* 44.2 (2008): 453-460.
- Western, B. and Pettit, B. (2010). *Collateral Costs: Incarceration's Effect on Economic Mobility*. Washington, DC: The Pew Charitable Trusts.
- White, B. T. (2005). Say you're sorry: court-ordered apologies as a civil rights remedy. *Cornell L. Rev.*, 91, 1261.
- Williams, K. D. (1997). *Social ostracism*. In *Aversive interpersonal behaviors*, pages 133-170. Springer US.
- Williams, K. D. (2002). *Ostracism: The power of silence*. Guilford Press.
- Witvliet, C. V., Worthington, E. L., Root, L. M., Sato, A. F., Ludwig, T. E., & Exline, J. J. (2008). Retributive justice, restorative justice, and forgiveness: An experimental psychophysiology analysis. *Journal of Experimental Social Psychology*, 44(1), 10-25.
- Wu, J., & Axelrod, R. (1995). How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict resolution*, 183-189.