The Sorry Clause*

Vatsalya Srivastava†
Tilburg University

January 12, 2018

Abstract

This paper shows the existence of a *sorry equilibrium* in a game of imperfect public monitoring. In this equilibrium, a self-imposed costly apology tendered after an accidental defection allows for continued cooperation between players. The cost of such an apology cannot be too high or too low. Efficiency of this sorry equilibrium is evaluated and its welfare outcomes compared to other informal governance mechanisms and the formal legal system. With the possibility of accidental defections, it is shown that informal mechanisms have limitations, while formal legal systems can generate perverse incentives. The analysis shows that apologies serve as a useful economic governance institution.

Keywords: Apology, Sorry, Imperfect Public Monitoring, Uncertainty, Social Norms, Economic Governance, Legal Institutions, Courts, Incentives.
JEL Classification: D02, D80, K40, Z13.

---

*I am very grateful to Jens Prüfer for invaluable guidance. This paper has benefitted from comments by Eric van Damme, Avinash Dixit, Cédric Argenton, Abhilash Maji, Michael Powell, John Barkley Rosser, Jr., Mark Ramseyer, Alex Teytelboym, Avner Greif, Clemens Fiedler and the participants of ISNIE (2015, Harvard), ESNIE (2015, Corsica), CLEEN (2015, Tilburg), Games (2016, Maastricht), EEA (2016, Geneva), TILEC and Economics Department seminars. All errors are my own.

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

The word sorry is an almost universally acknowledged expression of apology. As a regretful acknowledgment of an offense or failure, an apology is used across cultures to seek forgiveness, reconcile differences and enable future cooperation. They are tendered in a multitude of different circumstances 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] They are also used extensively in some legal systems. For instance, in Japan, apologies are used by the prosecution to suspend "33% of all cases involving non-traffic related offenses" (Haley, 1982, p.271). The widespread use of apologies is not lost on businesses either. If the wide variety of *sorry greeting cards* available were not evidence enough, there are firms that apologize on behalf of their clients for a fee:
"Tianjin and the central city of Xi′an now boast new, successful apology companies ... On behalf of clients, the apologizers write letters, deliver gifts and make explanations."[2]

This paper treats apologies as a social institution[3] and an instrument of economic governance[4] to develop a better understanding of when and how much to apologize, and to compare it with other informal mechanisms and formal court based systems. A game-theoretic model of imperfect public monitoring in an infinitely-repeated prisoner's dilemma is constructed for this purpose. This model captures a situation where not all players can control the action(s) that they play, generating the possibility of mistakes and therefore the need for an apology. It is shown that in such a setup, the use of a costly apology can sustain an equilibrium with cooperation. In such an equilibrium, a player apologizes by self inflicting a cost in case of an accidental defection in order to prevent future payoffs from being adversely affected.

The setup of imperfect public monitoring used here also eschews costly signaling (Ho, 2012; Mungan, 2012) or lying aversion (Fischbacher and Utikal, 2013), used in models of apologies by economists, in favor of a costly state verification model (Gale and Hellwig, 1985; Townsend, 1979). The difference from a canonical state verification setup is that the cost of an apology itself allows verification of the state. This allows welfare comparisons with other commonly used informal governance mechanisms like standard (and limited) grim-trigger and ostracism.

This paper contributes to the literature by developing a framework that enables delin-

[1] `https://www.aadnc-aandc.gc.ca/eng/1100100015644/1100100015649`

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

[3] Greif (2006) requires a social institution to have rules governing actions in a central transaction in addition to auxiliary rules that ensure that rules of the central transaction are enforced. Social institutions are therefore "a system of rules, beliefs, norms and organizations that together generate a regularity of (social) behavior"(p.30).

[4] Economic governance refers to "the structure and functioning of the legal and social institutions that support economic activity and economic transactions ..." (Dixit, 2009).

eation of the trade-off inherent in the use of apologies along with identifying the limitations of other governance mechanisms. The cost of an apology that allows for players to continue cooperating is calculated. It is shown that this cost cannot be too high, or an apology will not be used. It cannot be it be too low either, or it will incentivize intentional defection. Surprisingly, this cost depends on the benefit gained from the accidental defection by the offender. This clearly distinguishes apologies from compensation, which depends on harm caused to the offended. This distinction is in line with the depiction of apology as a mechanism that *"paradoxically restores social order without amending the transgression"* by Tavuchis (1991). It is also shown that the cost of an apology that emerges from the proposed equilibrium strategy is the lowest possible one that can sustain continued cooperation in any equilibrium that relies on a costly apology to avoid future punishment.

In comparison to apologies, it is shown that other governance mechanisms have certain limitations. Infinite trigger strategies and ostracism do not allow for continued cooperation. Even limited trigger strategies that allow for a reversion to cooperation, require that victims have to punish themselves to punish the offender. Further, strategies with punishment periods, face the additional problems that players might commit mistakes during the punishment period and this might require the punishment to restart, generating coordination problems. Apologies, invert the burden of resolution of an deviation and allows an offender to self-identify and punish himself to allow cooperation to continue. However, under certain parameter configurations, these other strategies do perform better in terms of social welfare. In light of this discrepancy and the benefits that apologies offer, it is postulated that certain pro-social preferences may have been socialized to reduce the social welfare loss from apologies and encourage their widespread usage.[5] The constraints that courts face in distinguishing mistakes from intentional defection is also investigated. It is found that while courts can deter infractions in some cases, in others, either litigation costs are too high or they impose damages that may generate perverse incentives.

The equilibrium strategy proposed to account for apologies is a *public* strategy that induces a *public perfect equilibrium* (PPE) (Abreu et al., 1990; Fudenberg and Maskin, 1986; Fudenberg et al., 1994). The continuation payoffs in this case are conditional on the observed public actions in the prisoners dilemma and on (the cost of) a public apology. This implies that after any observed defection, an appropriately costly apology *resets* the trigger for any punishment, allowing both players to continue cooperating. This voluntary imposition of cost to provide additional information on which punishment can be conditioned, distinguishes it from T-PPE (Kandori and Matsushima, 1998) and semi-public equilibria (Compte, 1998)

[5]Socialization is a term used by sociologists and anthropologists for the lifelong process of inheriting and disseminating norms, customs and values; providing an individual with the skills necessary for participating within their own society. Tabellini (2008) is a great example of an economic model endogenizing socialization.

used in the theoretical literature. This also makes it distinct from the types of punishment allowed in the experimental literature (Aoyagi and Frechette, 2009; Ambrus and Greiner, 2012; Fudenberg et al., 2012) on games of imperfect public and private monitoring. The voluntary nature of an apology makes it similar to information revelation in the literature on imperfect private monitoring (Compte, 1998; Kandori and Matsushima, 1998). However, in such games communication is used to glean information from noisy private signals in the absence of a public signal by making continuation payoffs of player independent of disclosed information. In the case of an apology, a player provides additional public information to avoid avoid punishment and improve continuation payoffs.

From a policy perspective, the paper shows that apologies can resolve some of the limitations that courts face in an imperfect public monitoring setting. This provides an analytical framework for the argument made by legal scholars that the judicial system of the USA stands to benefit from the restorative and conciliatory effects that apologies engender (Petrucci, 2002; Shuman, 2000; Wagatsuma and Rosett, 1986). The virtues of apologies have been evaluated and extolled in keeping down crime and recidivism rates (Haley, 1982, 1995, 1998; Robbennolt, 2003). Apologies have also been found to have a positive role in providing cathartic relief to affected parties in a wide variety of cases including, medical malpractice (Keeva, 1999), divorce (Schneider, 2000) and civil rights abuse (White, 2005). The particular cases of the USA and India are discussed in this regard.

The remainder of the paper is organized as follows. In the next section, the relevant literature on social interactions and economic governance is reviewed and the characteristics of apologies are identified. The model is outlined in section 2. In section 3 a strategy with apology is proposed and it is shown that it constitutes a PPE. The cost of an apology in such an equilibrium is calculated and comparative statics are presented. In section 4, efficiency of the proposed equilibrium is evaluated. The welfare outcomes of different community enforcement mechanisms are compared to apologies in section 5. In section 6 a model of formal court governance is outlined and its limitations identified. In section 7 some empirical facts comparing the legal systems in the USA, India and Japan are presented and policy implications of the theory are discussed. Section 8 concludes by summarizing the results and making suggestions for future research. All the proofs are provided in Appendix A.

1 Characterising Apology

In it's the day-to-day usage, apologies are used rather frequently, both in long-standing relationships as well as in interactions with relative strangers. Therefore, any model of apologies must capture the following characteristics:

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

(2) In order for apologies to be an equilibrium outcome there must be scope for failures/offenses in equilibrium. Given the usage of apologies in a wide-variety of circumstances, the cause of such failures must be sufficiently general to affect many different activities.

This paper therefore builds on models pioneered by Green and Porter (1984) in which, punishments are played out in equilibrium due to an exogenous source of uncertainty. Further, 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 (Darby and Schlenker, 1982; Schlenker, 1980; Schlenker and Darby, 1981).

(3) An apology is an attempt to establish a lack of intent on part of the offending party, along with an acknowledgement of harm or offense caused. While the harm caused is not disputed, the contention is that any such harm was accidental, not intentional. Thus, there must be a separation in intent and outcome in the game.

The possibility of *mistakes* in selection of actions explaining 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 some players having other-regarding preferences (lying averse). Such an approach requires additional assumptions about preferences and limits the possibility of exploring why such preferences might have been socialized in the first place.

The *separation of intent and outcome* has also been treated in a general setting in the economics literature, as games of imperfect public monitoring (Abreu et al., 1990; Fudenberg and Maskin, 1986; Fudenberg et al., 1994; Rubinstein, 1979). The setup in Rubinstein (1979) requires the offended player to statistically distinguish between an accidental offense and a deliberate one. Porter (1983) requires that the offended player defect for T periods following a defection to deter intentional defection (by changing the continuation payoffs). The offended player thus has to punish himself by defecting in these T periods to be able to punish the offender. The complications (and cognitive costs) of statistical identification and the undesirable consequences of punishment by Nash reversion on the offended player aside, the folk theorem result in this literature focuses on the set of possible payoffs, rather than specifying the strategy used by players (Abreu et al., 1990; Fudenberg and Maskin, 1986; Fudenberg et al., 1994).

There is also experimental literature that investigates players' strategies and outcomes in infinitely (Aoyagi and Frechette, 2009; Fudenberg et al., 2012) and finitely (Ambrus and Greiner, 2012) repeated games of imperfect public monitoring. In an infinitely repeated prisoner's dilemma, Fudenberg et al. (2012) find that players tend to use more cooperative strategies when the payoffs allow for a cooperative equilibrium (players cooperate until a

defection). They also find the players to be more lenient (don't always immediately punish a defection) when they only observe actions of other players with noise compared to when there is no noise. In a similar setup, Aoyagi and Frechette (2009) find that cooperation decreases in the game as the noise increases (though cooperation is higher than the theoretical maximum at high noises). In a finitely repeated public goods game, Ambrus and Greiner (2012) conclude that even though costly punishment can increase contributions it may not be good for social welfare due to the possibility of *unfair* punishments. Both these experiments reveal the uncertainty that can be created by imperfect public monitoring when the offended player is responsible for punishing deviance.

An apology is different from these approaches as it enables the offending player to take on the burden of resolution and avoiding the outcome where both players suffer in a punishment period. In line with this idea, in most social conventions apologies can only be offered, not taken. This implies that if an offending player can apologize, then the offended player need not punish every observed defection and can instead punish only those defections for which an apology is not tendered. This use of a self-imposed sanction to avoid future sanctions is a departure from the punishment strategies hitherto used in games of imperfect public monitoring.

(4) An apology has to be offered by the offender to the offended. This distinguishes it from conventional punishments that can be met out by some authority or an outside agent. This characteristic is most visible in the social dictum that emphasizes 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) Apologizing 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 apologizing, the player causing harm can always apologize and consequently there is no reason why the affected player should accept an apology. As apologies are taken seriously, there must be some psychological, social or even monetary costs attached to apologizing. This argument is akin to claiming, that in general apologizing is not cheap talk.

This is not to deny that in certain cases, apologies maybe cheap talk. However, cheap talk is useful (and a cheap talk apology likely to be honest) only if the interests of the 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 apology might indeed be useful (Ho, 2012; Ohtsubo and Watanabe, 2009). But the objective of this paper is not to investigate all possible uses of the word sorry. It is 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. This self imposed costly sanction, in the spirit of costly state space verification models (Gale and Hellwig, 1985; Townsend, 1979), can be used to

infer if the defection was a mistake or not, negating the need for an audit.

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 favoring their use be socialized in members of a society. This might imply that there maybe psychological costs and benefits attached with apologizing (refer Section 5.5).

(6) There are different levels of being sorry. The social psychology literature differentiates between at least five 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 an apology can be offered. This implies that the cost incurred for an apology can change to reflect some underlying differences in 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. *Residual stochasticity* refers to that left over component of uncertainty that does not depend on the level of care taken by an individual. This contrasts with the emphasis on the level of *due care* in work pioneered by Brown (1973). It is closer in flavor to the more recent work on product liability (Jos Ganuza et al., 2016; Polinsky and Shavell, 2010). Although apologies fundamentally differ from the reputation based market mechanisms posited as the alternative to courts in these papers.[6] This emphasis on uncertainty is also a point of departure from the work of Mungan (2012), where the cost of remorse has been emphasized as the motivation for apologies.[7]

Some of the previous attempts at modeling apologies have focused on its role as a signal (Ho, 2012; Mungan, 2012; Ohtsubo and Watanabe, 2009). A signaling model necessitates additional assumptions about the differences in preferences across types of players, the choice of which must be contrived. Further, the differences across types might even be superfluous to an understanding of apologies, as the additional insights from a signaling model can be inferred from the simpler model used here.[8] The setup used in this paper relies on the simple idea that mistakes motivate the need for apologies to develop a clearly specified model, distinguishing it from previous works (Mungan, 2012; Ohtsubo and Watanabe, 2009).

[6]Apologies allow the offender to undertake the entire cost of punishment, whereas other strategies require that the offended player also suffer a cost in punishing the offender.

[7]Mungan (2012) also proposes a costly apology to distinguish between sincere and insincere apologies from offenders who suffer from remorse ex-post committing a crime.

[8]If there are two types of players: one prefers to apologize after mistake (the nice type) and the other doesn't (the bad type). The good type is modeled in subsequent sections. The second type is not explicitly modeled, however off equilibrium actions of the good type can be inferred to capture those of the bad type.

Additionally, this model endogenizes the cost of an apology and allows for comparisons with other informal governance mechanisms and courts (existing models of which do not employ signals). Such comparisons are not addressed by the existing economic literature on apologies (Fischbacher and Utikal, 2013; Ho, 2012).

2 The Model

The setting for this model is an infinitely repeated bilateral game, where the players select their actions simultaneously in every repetition.[9] 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: $t \in \{0, 1, 2 \ldots \infty\}$. 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^t, the value of which need not be stationary across repetitions.[10] The cost of the apology is borne by the player who apologizes. This cost is not borne to compensate the other player, but is considered as utility lost from this two-player system.

A player's *choice* of action is determined by intention. If a player intends to cooperate, he chooses to play C. But the uncertainty in playing the action means that the player cannot ensure that C is actually played out. In effect, the distinction between intent and outcome emphasized earlier is captured by the difference between *chosen action* and *the action that gets played*. This has consequences for the information available to the other player: the action that gets played can be observed, but there is no information about the chosen action. 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.

[9]The game can be extended to depict multi-lateral community interaction, wherein a player might play different players across repetitions, but all potential players in the community know of the outcome of previous iterations of the stage game with some probability.

[10]The self-inflicted cost, undertaken by the offending party to convince the offended party that any mistakes committed were accidental. For example, buying a sorry greeting-card.

(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 $(1-p)$.[11]

(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. The expected payoffs in the stage game in every period t would be as in Table 2.

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^1_{CD}: expected payoff of player 1, when player 1 chooses to play C and player 2 plays D, where $E^1_{CD} = pl + (1-p)d$. The first row, first column contains E^1_{CC} and E^2_{CC}. E^1_{CD} captures the payoff in the first row, second column.

(5) Sustained cooperation in this game is desirable only if $E^2_{CC} > E^2_{DD}$, so it is assumed that $p > \frac{w-l}{h+w-d-l}$. [12]

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 realized.

Stage 2: Player 1 decides to apologize at cost s^t (where $s^t \geq 0$).

3 Analysis

The game described here is one of imperfect public monitoring. Both players know the *action that played out* and if an apology was made (public information). The set of actions that can be played out by both players is {C,D}. Given the structure of the game, a strategy will prescribe an action choice in stage 1 for both players and the cost of apology in stage 2 to be undertaken player 1. Letting x^t denote the action played out by player 1, y^t denote the action of player 2, with s^t as the cost of apology in t, the public history of the game at t is $h^t = \{(x^0, y^0, s^0), (x^1, y^1, s^1), \ldots, (x^{t-1}, y^{t-1}, s^{t-1})\}$. The pure strategy equilibrium proposed

[11] See Appendix B for a treatment of the symmetric case where neither player can ensure with certainty which action will play out.

[12] $0.5 < \frac{w-l}{h+w-d-l}$ as $0 < (h-d) < (w-l)$.

here conditions only on public history of the game and has a stationary cost of apology s^*, such that s^t can take only two values, either s^* or 0.

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

- Choose to play C if h^t is such that $\forall \hat{t} < t$:
  - $x^{\hat{t}} = C$ and $y^{\hat{t}} = C$; or
  - $x^{\hat{t}} = D$ and $y^{\hat{t}} = C$ and $s^{\hat{t}} = s^*$.
- Else, choose to play D.

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

- $s^t = s^*$ if $x^t = D$, $y^t = C$ and if h^t is such that $\forall \hat{t} < t$:
  - $x^{\hat{t}} = C$ and $y^{\hat{t}} = C$; or
  - $x^{\hat{t}} = D$ and $y^{\hat{t}} = C$ and $s^{\hat{t}} = s^*$.
- Else, $s^t = 0$.

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

- Play C if h^t is such that $\forall \hat{t} < t$:
  - $x^{\hat{t}} = C$ and $y^{\hat{t}} = C$; or
  - $x^{\hat{t}} = D$ and $y^{\hat{t}} = C$ and $s^{\hat{t}} = s^*$.
- Else, play D.

This is a *public* strategy as both players condition their actions only on publicly observed outcomes. It relies on two distinct kind of sanctions: a self-inflicted punishment (cost of apology) and a trigger strategy. In equilibrium, both players choose to cooperate. In case of an accidental defection player 1 apologizes (at the cost s^*), following which both players continue to cooperate. This happens despite player 2 not being able to verify if player 1 had chosen to cooperate or defect before offering the apology. It will be shown in this section that there is a cost of apology for which this strategy constitutes *Perfect Public Equilibrium* (PPE). In such an equilibrium, the strategy will constitute a Nash equilibrium of the game for each t and public history h^t. Therefore, in a PPE players must not only be playing mutual best responses on the equilibrium path, but also in each public history off the equilibrium path. This requirement makes PPE analogous to a subgame perfect equilibrium for games of imperfect public monitoring. This is because, even though imperfect observability ensures

that there are no proper subgames of the repeated game, each public history can be treated as if it gives rise to a distinct subgame (Fudenberg et al., 1994).

The continuation of cooperation in equilibrium depends on the cost of apology s^*. This is because even if a defection by player 1 is intentional; player 1 can apologize if (D,C) gets played. 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 in stage 1. The equilibrium conditions must therefore ensure that choosing to play C in stage 1 is incentive compatible for both the players and that apologizing when an accidental defection happens is incentive compatible for player 1 in stage 2.

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 apologizing should be higher than the payoff from not apologizing.

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

The left hand side (LHS) of equation 1 is the payoff of player 1 from not apologizing once D is played out. The right hand side (RHS) of the equation is the payoff from apologizing in the same case. The timing of the game ensures that decision to apologize 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 cannot do any better than choosing to play D. Equation 1 simplifies into the first incentive compatibility constraint (IC):

IC 1: $s^* \leq \frac{\delta(E^1_{CC} - E^1_{DD})}{(1-\delta p)}$, player 1 must be better off apologizing 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 apologizing if D gets played, than choosing to play D and apologizing if D gets played (assuming IC 1 holds). As player 1 apologizes in either case, the continuation payoff will not differ across the two cases. Therefore, the cost of apology must be such that even a single-shot deviation must not be incentive compatible.

$$(E^1_{DC} - ps) \leq (E^1_{CC} - (1-p)s) \tag{2}$$

Equation 2 compares the payoffs of player 1 in a single repetition of the game. LHS of equation 2 is the payoff of player 1 from choosing to play D and apologizing when D gets

played. The RHS is the payoff from choosing to play C and apologizing 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: $s^* \geq \frac{(E^1_{DC}-E^1_{CC})}{(2p-1)}$, 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 payoff that player 1 gets from choosing to play D and apologizing. 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:

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

The constraint on δ by Lemma 1 is relaxed as h and/or $(d+l)$ increases. The equilibrium can therefore be supported by larger range of δ if the benefits from cooperating and/or the payoff loss in the game by defection of one or both players are large.

Player 2: In addition to the condition 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.[13]

$$E^2_{CD} + \sum_{t=1}^{\infty} \delta^t E^2_{DD} \leq \sum_{t=0}^{\infty} \delta^t E^2_{CC} \tag{3}$$

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

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

The constraint on δ by Lemma 2 is relaxed as $(d-l)$ increases relative to $(w-h)$. The equilibrium can therefore be supported by larger range of δ as the gains from cheating when the other player cooperates decreases relative to the the saving from cheating when the other player cheats.

3.1 Existence

Proposition 1 (Existence) *For every p, there exists a $\delta(p)$ such that the proposed strategy profile constitutes a Perfect Public Equilibrium for all $\delta \geq \delta(p)$.*

[13]Player 1 cannot choose actions with certainty. So, if δ is large enough such that player 2 does not deviate due to the threat of the grim trigger, player 1 will also be deterred from choosing to defect in stage 1.

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 will not choose to intentionally defect in stage 1. Therefore, if player apologies at a cost of s^*, in equilibrium, player 2 is informed that the defection was not deliberate. Player 2 then responds by choosing to cooperate in the next repetition signifying that the *apology was accepted* and player 1 *forgiven*. 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 allows 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 it.[14] It also captures all of the characteristics of an apology identified in section 1. 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.

3.2 Cost of Apology

Thus far, the cost of apology is captured by parameter s, implying that player 1 cannot choose the cost of apology. In this sub-section s is assumed to be a choice variable, whereby player 1 can choose the cost he incurs to offer an apology. Further, while IC 1 and IC 2 specify the range of s^*, an exact amount is not prescribed. However, the exact cost of apology that will be chosen by player 1 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, $s^{**} = \frac{(E^1_{DC} - E^1_{CC})}{(2p-1)} = (w - h)$.*

This result states 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 aspect of the cost of apology in

[14]Such costs can be incurred in a variety of ways: public declaration accepting the offense caused; spending time and effort to convince player 2; acts of self-sacrifice: a player might give up something valuable etc.

equilibrium is that it does not depend on the harm caused to player 2 (l). This is in contrast to the law and economics literature that focuses on harm caused (Hermalin et al., 2007).[15]

Apology vs Compensation: Compensation might be part of an apology and an apology might accompany a compensation. But, the equilibrium here highlights the differences between an apology and a compensation. 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 need not come with an implicit promise of continued cooperation.

3.3 Comparative Statics

Equilibrium Condition 1 (EC 1)and 2 (EC 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 (probability of action chosen by player 1 getting played) would allow a smaller δ to sustain the equilibrium. EC 2 is in line with this expectation; as lesser the chances of player 1 making a mistake, greater is the threat of the grim-trigger for player 2.

However, EC 1 mandates that a higher p requires a larger δ to support the equilibrium.[16] This result is counter-intuitive as lower chances of an accidental defection requires an increase in the importance of the future to support the equilibrium. This anomaly is caused by the constraint posed by IC 1, which requires that if p is higher, ceteris paribus, player 1 has a greater incentive to choose to play C and not apologize. In equilibrium this increased incentive to not apologize is offset by the larger future payoffs accrued from apologizing.

4 Efficiency and Equilibrium Payoffs

The equilibrium described in section 3 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.[17] For instance, a strategy that requires player 1 to apologize only once every 2 mistakes. Such a strategy would consequently also require

[15]Damages in law and economics literature are compensation mechanisms, where player 1 pays player 2. An apology as described here, does not involve any transfers.

[16]Proof in Appendix A.3.

[17]This isn't a claim that they do exist, but a conjecture that they might.

that player 2 continue to cooperate after the first mistake knowing that if a second mistake occurs, player 1 will apologize. 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. In general, in a game of asymmetric (one player cannot control actions that are played out) imperfect public monitoring in a prisoner's dilemma, equilibrium characterized by a costly apology can be classified as a distinct sub-set of the possible equilibrium of the game.

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

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 analyzed in the previous section, imposes a cost of $(w-h)$ in equilibrium utilizing the simplest possible strategy that involves the use of apologies in this setting. This simplest possible sorry equilibrium generates the lowest cost apology that is possible.

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

The intuition for this result is that in every t, if the net present expected cost of apology is less than $(w-h)$, then player 1 will choose to defect and the cooperative equilibrium will collapse. It must be noted that the cost of an apology need not be stationary. For instance, the cost of an apology might increase in the number of infractions caused. However, as the cost of apology must be sufficiently high to deter defection in every period and low enough to be offered by player 1, the expected net present value of total cost incurred due to a defection must be subject to IC 1 and IC 2. IC 2 requires that the total cost incurred by player 1 for every defection must be greater than or equal to s^{**}, and this cost depends on the payoff parameters w and h. Therefore the total cost of a defection in any sorry equilibrium, whether it is spread over multiple periods or costs for multiple defections are clubbed together, can be no cheaper than s^{**}. This also implies that the set of equilibria payoffs for all possible sorry equilibrium can be characterized.

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

This proposition encapsulates the idea that the expected net present value of the total costs that must be incurred by player 1 in any sorry equilibrium must conform to IC 1 and 2.

These results show that the equilibrium analyzed here, while not unique, has some features that make it worth investigating from amongst the family of all possible sorry equilibria. Hereafter, the equilibrium analyzed in section 3 is referred to as "the sorry equilibrium".

5 Welfare

5.1 Social Welfare in the Sorry Equilibrium

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:

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

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.

5.2 Social Welfare under Grim-Trigger Strategy

The grim-trigger 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.[18] In this setting, however, accidental deviations happen and set off the grim trigger.

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

[18]It is pertinent to point out that for $p = 1$, the sorry equilibrium is the same as a grim trigger strategy equilibrium.

Proposition 5 (SE and GT) *Social welfare in the sorry equilibrium is higher than with grim-trigger strategy if:* $2p\delta(h-d) > (1-\delta p)s^{**}$.

The LHS of this condition captures the difference between the continuation payoffs after a (accidental) defection across the two cases. In the case of the sorry equilibrium, the players continue to cooperate (captured by h), whereas in the grim trigger case both players defect (captured by d). The big advantage offered by the sorry equilibrium is continued cooperation. So, as the value of cooperation (h) increases relative to the costs of mutual defection (d), the social welfare under the sorry equilibrium gets larger.[19] It is also obvious that as the cost of apology decreases, i.e. as the difference between w and h reduces, social welfare in the sorry equilibrium increases. In addition to that a higher p and/or δ , ceteris paribus, also make the sorry equilibrium more attractive. A higher p has a dual effect. It increases the probability of the cooperative outcome being played out and thereby increasing the value of continuation payoffs from continued cooperation under the sorry equilibrium. It also simultaneously reduces the expected cost of apology in the sorry equilibrium. A higher δ implies more patient players. Such players would place a relatively greater weight on the value of the continuation payoff than the cost of apology.

Despite the fact that a grim trigger strategy might lead to higher net social welfare under certain parameter configurations, it does suffer from certain limitations. Player 2, the player who can choose his actions with certainty is worse off playing a grim trigger strategy than the sorry equilibrium.[20] This creates a situation where a stricter punishment mechanism makes the victim suffer more. In general, the prospect of reducing ones own payoff to punish a defector makes a threat more credible. However, in this case due to residual stochasticity, the threat cannot prevent all defections and the grim outcome has to played out, making the victim worse off. 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).

5.3 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 modeled in section 2, 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;

[19] A increase in h would, ceteris paribus, also decrease the cost of apology.

[20] This is because $E^2_{CC} > E^2_{DD}$ in the stage game.

the social costs of ostracism also includes 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.[21] These costs of ostracism are captured by the parameter O_s and lumped together in the period in which a player defects.[22]

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

Proposition 6 (SE and O) *Social welfare in the sorry equilibrium is higher than under ostracism if:* $2p\delta h + (1-p)\delta(w+l) + (1-\delta)O_s > (1-\delta p)s^{**}$.

The LHS of this condition captures the difference between the continuation payoffs in the sorry equilibrium. This payoff is determined by the value of continued cooperation (h), the cost of accidental defection while choosing to cooperate ($w+l$) and avoidance of the costs of ostracism (O_s). It is obvious that an increase in O_s would make the sorry equilibrium more attractive. Also, if the benefits from continued cooperation (h) increase and/or the cost of defection for player 2 (l) decreases, the social welfare in the sorry equilibrium increases. However, an increase in the payoff from defection (w) increases the cost of apology more than it increases the value of the continued cooperation.[23] Additionally, an increase in p, again has the dual effect of increasing the value of continued cooperation (in both the sorry equilibrium and under ostracism) and simultaneously decreasing the expected cost of apology. As the increase in continued cooperation in both cases are the same, the reduction in the expected cost of apology makes the sorry equilibrium relatively more lucrative. An increase in δ, on the other hand has an ambiguous effect. Its effect depends on the relative sizes of the payoffs (h, w and l) and the cost of ostracism. This is because while it increases the value of continued cooperation and decreases the weight on the cost of apology; it also increases the weight on the cost of accidental defection (while cooperating) and on the cost of ostracism.

Given all of 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 the cost of apology and/or the cost of defection for player 2 (l) are low.

[21] Example: if the only doctor in a community was ostracised, cost to the community could be very high.

[22] All interaction ceases after a player defects and the game is not played.

[23] This is because $(1-\delta p) > (1-p)\delta$.

5.4 Limited Trigger Strategies and Contrite 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, cooperation ceases. There are however other alternatives in the literature that allow for a return to cooperation after a defection.

One example is 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. Such a strategy would be akin to the sort of trigger strategy with "price wars" between colluding oligopolists suggested in Porter (1983). This avoids the problem of the standard grim trigger by limiting the number of punishment periods. A few other such lenient variants of the grim trigger, like Tit-for-2-Tats (punish by defecting once after two defections) and 2-Tits-for-2-Tats (punish by defecting twice after two defections) were found to be frequently used by players in an experimental setting by Fudenberg et al. (2012). However, all of them retain the undesirable feature of punishing the victim more than in the sorry equilibrium.

The strategy of *Contrite Tit for Tat* (CTFT) (Wu and Axelrod, 1995) resolves this problem. This 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 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 will: either differ across different cases of defection or the length of the punishment period will have to be conditional on the actions played out in the punishment period. 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.

5.5 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 and/or l decreases 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*.[24] The economic literature on pro-social preferences propounds: *"Societies go to great lengths to instil such preferences in children . . . socialization 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 3 it certainly improves its social welfare consequences, potentially making it more viable in a wider variety of situations.

6 Courts

The laws that courts are charged with implementing vary considerably across jurisdictions and are often very complicated, riddled with caveats and exceptions. This makes modeling a formal legal system which fits all observable cases, almost impossible. Nonetheless, a simple model of court enforcement might yield useful insights; assisting in identifying limitations, complications and the possibility of perverse incentives.This section evaluates the limitations of a formal legal system when dealing with players who face residual stochasticity.

The efficacy of a legal system, in this case, 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 sanction to deter both players from deliberately defecting and when an accidental defection does occur, initiate some action to allow for future cooperation. This ideal court would be able to read intentions and inflict sanctions of its own accord. However, this ideal system is highly unlikely. Hence the need for this model.

[24]This also relates to the behavioural cost of making an apology mentioned in section 1. Such costs signify that even seemingly costless apologies have a cost associated with them.

6.1 Court Model

The model of courts in this section uses the stage game described in section 2. The assumptions of the model are based on Masten and Prüfer (2014) and are as follows:

(1) A court cannot take *Suo-Motu* notice of a defection, a suit must be brought before the court by a plaintiff, accusing the defendant.[25]

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

(3) The (exogenous) parameter $\tau \in [0, 1]$ reflects the probability with which the court will rule in favor of the plaintiff.[26] 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.

(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.

The last assumption ensures that a defecting player has no incentive to file a suit.[27] 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 $(-\tau I - c)$. Therefore, for a cooperating player to have an incentive to file a suit, $\tau I \geq c$. If the court deems the minimum damages to deter infractions to be T, then the damages awarded must be $I = max\{T, \frac{c}{\tau}\}$.

6.2 Damages and Litigation Costs

The courts as described here, cannot distinguish between accidental and deliberate defections. Therefore, damages awarded cannot account for the difference across the two cases. Nonetheless, in order to dissuade deliberate defections by player 1 the court will calculate damages to make the payoffs from choosing to cooperate higher than deliberately defecting.

$$ph + (1-p)(w - \tau T - c) \geq (1-p)h + p(w - \tau T - c) \tag{5}$$

The equation assumes that player 2 cooperates, this is because a case would be brought before the court only by a cooperation player. The lower bound for the required transfer as determined by such a court can be obtained by rearranging this inequality.

$$T \geq \frac{w - h - c}{\tau} \tag{6}$$

[25] Suo motu is a Latin term meaning on its own motion, it is used where a court acts on its own cognizance.

[26] This assumption will be altered in section 6.3.

[27] Courts do not make type 1 errors. So if a defecting player sues he will merely incur a cost of $-c$.

The transfer amount is set such that the payoff from cooperating be higher than from defecting.[28] Given this T, if $\frac{c}{\tau} > T$, such that $I = \frac{c}{\tau}$, then the cost of litigation, $2c$ would be higher than the cost of an apology in the sorry equilibrium.[29] Therefore, in the rest of this section it is assumed that c is small enough such that $I = T = \frac{w-h-c}{\tau}$.[30]

If the cost of litigation is small: $I = T$. In this case, the damages awarded by the court will be sufficient to deter intentional defection by player 1. The total cost of such a decision, including damages and litigation costs ($2c$) would be larger than the cost of an apology.[31] However, as damages are transferred from one player to another and litigation costs small, the net impact on social welfare would be smaller than in the sorry equilibrium.

6.3 Damages and Outcomes

Results in the previous sub-section refer to a specific case where punishment does not vary across deliberate and accidental defections. This section will relax this assumption and consider cases where this condition is not met. The role of intent in law is rather complicated and difficult to fully capture in a stylized setting. *"... there is a large class of crimes for which showing mens-rea (a guilty mind) requires showing that the defendant had some intention when engaging in the actus-reus (a guilty act)"* (Yaffe, 2010). This is the reason for which most legal systems, for instance distinguish between murder and manslaughter, punishing the former more severely. The uncertainty in this process of proving intent, will be captured here by the difference in the probability of being convicted between an accidental defection and a deliberate one. This is a simplifying assumption, as instead of allowing for uncertainty across multiple levels of punishment (as in the case of murder and manslaughter), this model assumes that there is only one level of punishment available to courts.

As in assumption 4, courts do not know the action that players intended to play. However, there is likely to be a difference across deliberate and accidental defection in the probability of meeting the burden of proof. Therefore in this section, the modified assumption 3:

Modified assumption 3: (3a) The probability of punishing a deliberate defection remains $\tau \in [0, 1]$. However, the probability of an accidental defection being punished is $\acute{\tau} \in [0, 1]$. Further, it is assumed that given the lack of information about the intent of the accused, courts are unable to account for the difference across τ and $\acute{\tau}$.[32]

[28] This 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.

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

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

[31] $\frac{w-h-c}{\tau} + 2c > (w - h)$, as $\frac{w-h-c}{w-h-2c} > 1 > \tau$.

[32] This assumption does not require that courts be unaware that this difference exists, but that they are

Damages under Assumption 3(a): Now, the magnitude of I must be sufficient to deter a deliberate defection by player 1. It is clear that I is sufficient to deter player 1 from intentionally defecting iff:

$$ph + (1-p)(w - \acute{\tau}I - c) \geq (1-p)h + p(w - \tau I - c) \tag{7}$$

The appropriate I can obtained by re-writing equation 7.

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

The difference between equations 6 and 8 stems from assumption (3a). 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 scenarios transpires if $\tau = 1$ and $\acute{\tau} = 0$. This implies that a deliberate defection is punished with certainty, while an accidental defection is not punished. In this case equation 7 reduces to equation 9. I chosen by the court will dissuade deliberate defection by player 1.[33]

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

The General Case: If τ and $\acute{\tau}$ are not assigned extreme values, then the positive results of the best case scenario need not hold. Assuming that $I^* = \frac{(2p-1)}{p(\acute{\tau}+\tau)-\acute{\tau}}(w-h-c)$, whether I^* is sufficient to deter intentional defection depends on the relative sizes of τ and $\acute{\tau}$.

Lemma 6 (Adequate Damages) *: Adequacy of I as deterrent depends on τ and $\acute{\tau}$*

(i) If $\tau < \acute{\tau}$, then $I < I^*$.

(ii) If $\tau > \acute{\tau}$, then $I > I^*$.

The intuition behind this result rests on the fact that when there are no damages, player 1 is better off choosing to defect if player 2 cooperates. In the best case scenario, defecting deliberately would result in certain punishment and therefore was not a lucrative option. However, if $\tau < \acute{\tau}$, the expected damages to be paid are higher in the case of accidental defection. This would incentivise deliberate defection and therefore require a higher I^* to ensure that player 1 chooses to cooperate. The reverse is true for when $\tau > \acute{\tau}$.

(i) When $\tau < \acute{\tau}$: If damage I is imposed by courts, player 1 will deliberately defect if player 2 cooperates. This might in turn have an effect on the preferred action of player 2. Therefore, the impact of court awarded damages must be evaluated in equilibrium.

---

unaware of the magnitude of the difference between τ and $\acute{\tau}$.

[33] As $\frac{(2p-1)}{p} < 1$, $\tau = 1$ and $\acute{\tau} = 0$; $\frac{(w-h-c)}{\tau} > \frac{(2p-1)(w-h-c)}{p}$.

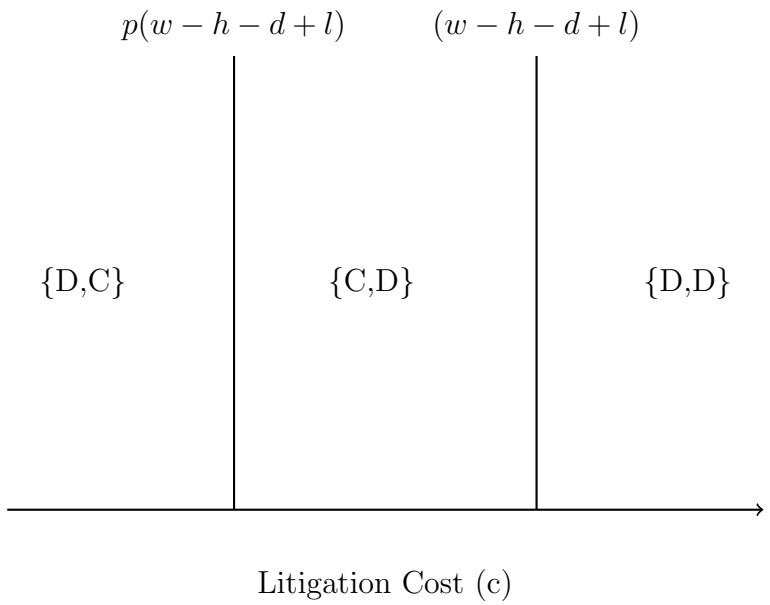

Figure 1: Equilibria in Pure Strategies

Proposition 7 (Equilibrium in One shot Game) *Pure strategy Nash Equilibria of the one shot game:*[34]

- *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* $(w-h-d+l) \geq c$*: Player 2 defecting and Player 1 choosing to cooperate is a Nash equilibrium.*

This set of outcomes are a result of residual stochasticity and the fact that a cooperating player can get damages from a defecting player. So, when litigation costs are low player 1 is better off choosing to defect as the benefits of defection compensate for the low damages and cost of litigation incurred when defecting. On the other hand, player 2 is better off cooperating and getting damages with probability τp (while incurring small c) rather than defecting and giving up the possibility of receiving damages. As the cost of litigation rises, player 1 is better off choosing to cooperate due to possibility of receiving damages from player 2. However, the low damages are insufficient to deter player 2 from defecting when player 1 chooses to cooperate. When the cost of litigation is very high, both players are better off choosing to defect. Player 1 is better off reducing the probability of going to court $(1-p)$ and the damages are still insufficient to deter player 2 from intentionally defecting.

[34]The proposition assumes that $(w-h-d+l) > 0$.

So, if the legal system is such that an accidental defection is more likely to be punished than an intentional one, none of the equilibrium outcomes are desirable. The courts might respond to such a situation by awarding larger damages (refer 6.4).

(ii) When $\acute{\tau} < \tau$ **:** Court imposed damages I, can deter intentional defection by player 1, as player 1 will choose to cooperate if player 2 cooperates. However, difference between I and I^* depends on the difference between $\acute{\tau}$ and τ.

Lemma 7 (Larger Damages) *For* $\tau > \acute{\tau}$*: As the difference between* τ *and* $\acute{\tau}$ *increases,* I^* *decreases and the difference between* I *and* I^* *increases.*

This result states that as the difference between τ and $\acute{\tau}$ reduces, the damages required to deter deliberate defection will reduce. This is as, in the absence of damages, player 1 is better choosing to defect if player 2 cooperates. But as the difference between τ and $\acute{\tau}$ increases and an intentional defection is more likely to be punished choosing to cooperate becomes relatively more profitable for player 1. This generates a paradox. As the legal system gets better at distinguishing between accidental and intentional defections, it imposes progressively higher damages than required. As intentional defections are deterred, accidental defections are punished with much larger damages than they need to be.

6.4 Large(r) Damages and Unresolved Disputes

The simplest solution to the problem encountered in proposition 7, is for courts to award large damages. However, instituting larger damages may create other problems. Large transfers entail large-scale redistribution and can potentially give rise to perverse incentives. Large damages could foster socially unproductive investments and/or affect the economy directly by limiting the willingness of people to participate in activities that involve some uncertainty (innovation). The problems from awarding large damages get further exacerbated if an assumption made earlier is relaxed and courts are allowed to make type 1 (false positives) errors. The damages imposed can be larger than the perceived value of continued interaction, giving the players an incentive to sue even if the outcome is (C,C) or even when they themselves have defected. A perception of larger damages being imposed on accidental defections than required (as shown in lemma 7) might also lead to similar problems. Awarding large(r) 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. 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 (increase in c), lengthy court proceedings and a higher burden of proof (might also have an effect on τ).

Further, even if adequate damages are imposed by the courts, they are only received by the offended player with a probability of $\acute{\tau}$. So, with probability of (1 - $\acute{\tau}$), the offended player receives no compensation or validation of his claim that an offense has been committed. In such cases, it is likely that the interaction will cease to continue in the future.

7 USA, India and Japan

The criminal justice system in the United States of America has 1 in every 35 American adults under its purview, of which approximately 2 million are incarcerated (Glaze and Herberman, 2012). In fact, the USA is next only to Seychelles[35] in its prison population rate.[36] The tort system in the USA 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 Rasmusen, 2010).[37] Their penchant for awarding large punitive damages in many high profile cases is well documented (Ramseyer and Rasmusen, 2010; Wagatsuma and Rosett, 1986). 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 country ... 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.[38] 70% of the incarcerated population is under-trial and many for those awaiting trials have been in jail longer than a sentence would have required them to be (Krishnan and Raj Kumar, 2010). The costs of such a logjam in

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

[36]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). Among formerly incarcerated men in the study, 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). All this is made worse by the fact that many people are returned to jail for non-payment of fines and fees (Evans, 2014).

[37]The authors distinguish between first-order law and second-order law. In this classification, first order law pertains to the typical disputes over contract claims.

[38]`http://timesofindia.indiatimes.com/india/Courts-will-take-320-years-to-clear-backlog-cases-Justice-Rao/articleshow/5651782.cms.`

the courts, both social and economic, are so large that while often spoken of, hardly any comprehensive and reliable estimates are available.

The nature and source of the problems facing these two legal systems are disparate, but it is apparent that both the systems require improvements. In contrast to the USA and India, in the Japanese legal system the *attitude of the accused and their willingness to apologize 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 apologize after the offense, at one time in Japan *"33% of all cases involving non-traffic related offenses 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 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 idiosyncrasy, there is anecdotal evidence from all over the world. In the USA, evidence from a study on settlement offers at the time of an accident reports: 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. A psychology study gathered evidence to show that a costly apology is considered sincere across countries (Ohtsubo et al., 2012).[39] In fact, it has been claimed that tort plaintiffs (in the USA) often profess: *"what they really wanted was an apology and brought suit only when it was not forthcoming"* (Shuman, 2000). From a firm's perspective, there is evidence to show that even a cheap talk apology, following an unsatisfactory purchase leads to better outcomes than monetary compensation (Abeler et al., 2010).

7.1 Policy Discussion

Section 6 presented a stylized model of the difficulties that courts may face in deterring intentional defection when faced with residual stochasticity. The results of proposition 7 show that insufficient damages allow for intentional defection, increasing the probability of infractions. This seems to reflect some of the elements of the Indian judicial system which, historically has been given in to low damages and fines (Galanter, 1985; Srinivasan

[39] Across 7 countries: Chile, China, Indonesia, Japan, the Netherlands, South Korea and the USA.

and Eyre, 2007). 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). Also, lemma 7 and the affect of large damages (refer section 6.4), reflect some elements of the American judicial system. Damages that may deter intentional defection can be perceived to be too high, leading to criticism of the legal system for being excessively costly and inefficient (Shuman, 2000).

Further, unresolved cases in which the plaintiff doesn't get a conviction (as will happen with probability (1-$\acute{\tau}$), are likely to lead to festering anger and resentment. Legal scholars have argued that such emotional anguish and the other issues mentioned here can be avoided by a more active use of apologies in legal systems (Keeva, 1999; Petrucci, 2002; Schneider, 2000; Shuman, 2000; Wagatsuma and Rosett, 1986).

Recidivism is another issue of great concern for legal systems. A legal 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).

Residual stochasticity in choosing actions requires mechanisms that can distinguish between the deliberate and the accidental. Costly apologies offer one such mechanism. They reduce the cost of litigation (self-identification), the need for large damages (no undesirable side-effects) and allow for continued cooperation (reconciliation). The use of apologies in Japan offers insights into how apologies can be a low cost judicial tool for non-grevious offenses. Given the problems facing the legal system in both the USA and India, it may be prudent to allow court sanctioned apologies for minor misdemeanours (Petrucci, 2002). [40]

An intriguing suggestion with regards to integrating apologies into the American legal system is that offenders who wish to apologize be subjected to more stringent punishment (Mungan, 2012). A higher punishment for an apologizer would allow courts to distinguish between sincere and insincere apologies. However, this suggestion relies on the assertion that apologies are a way for remorseful individuals to alleviate the costs of such remorse. However, if residual stochasticity is the motivation for apologies, a higher punishment for accidental defection might be inefficient (would retard learning and innovation).

This paper proposes that apologies need not be intrinsically cheap. Therefore, courts should utilize the pre-existing social institution of apology to alleviate some of the inefficiencies in legal systems. Further, the sorry equilibrium does not rely on nature of the payoffs involved. They need not be monetary (or measured in incarcerated time) and can be psy-

[40]Social welfare depends on harm caused(l). The model assumes that l does not limit a player's participation in future repetitions. In cases of debilitating injury, death etc. this assumption is not satisfied.

chological or social too. This implies that an emotionally costly apology works even when the payoffs from the stage game are purely monetary. This is of particular importance for American courts, which are given in to evaluating the monetary equivalents of emotional distress and degradation, leading to absurdly large damages (Wagatsuma and Rosett, 1986).

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.[41] There are additional benefits from reforms that would 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 accepted (and therefore made).The 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. 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.

8 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. Its ubiquity creates 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 would ever 'break'. 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

[41] In one California County in 2002 there were more than 12,000 guilty pleas entered by people who did not have a lawyer. (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).

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.

Given the pervasive use of apologies, either our society exists in a constant state of disequilibrium or an apology must be an equilibrium outcome. The sorry equilibrium posited in section 3 provides evidence for the latter. It shows that a strategy where a costly apology following a defection *resets* the punishment trigger constitutes a *public perfect equilirium* (PPE) in a game of imperfect public monitoring. It also shows that the cost of an apology that supports such an equilibrium cannot be too high or too low. Further, in the set of all possible PPEs that rely on a costly apology, the equilibrium strategy posited is shown to be the most efficient, enabling the characterization of the entire set of possible payoffs in any sorry equilibrium.

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 or even death. 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 shown in section 5, the sorry equilibrium offers a simple and easy to implement apology.

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 6. These findings lend credence to the calls in the legal literature 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 2 makes certain strong assumptions. For instance, the effectiveness of an apology might depend on the frequency of its usage. Modeling this would need accounting for residual and controlled (determined by the extent of care) stochasticity simultaneously. Alternately, a model that explicitly models the dynamic benefits of continued interaction (learning: reduction in p or l, increase in h) might be more useful in determining the value of sustained cooperation. These potential extensions show much work remains to be done to completely understand apologies.

A Appendix A

A.1 Proof of Proposition 1

For the Sorry Equilibrium to constitute a Nash equilibrium, both Lemma 1 (EC 1) and Lemma 2 (EC 2) must be satisfied simultaneously:

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

As, $\frac{(E^1_{DC}-E^1_{CC})}{(2p-1)} = \frac{(pw+(1-p)h-ph-(1-p)w)}{(2p-1)} = (w-h)$, EC 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 $\delta \in (0,1)$

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

Consider EC 2: $\delta \geq \frac{(E^2_{CD}-E^2_{CC})}{(E^2_{CD}-E^2_{DD})}$. By Assumption (5) in section 2, $E^2_{CC} > E^2_{DD}$, therefore:

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

As the condition A.2 is always met, there is some $\delta \in (0,1)$ that always satisfied EC 2. Therefore, for every given p, there must be some $\delta \in (0,1)$ that satisfies both EC 1 and EC 2. This implies that the proposed strategy constitutes a Nash equilibrium.

For the proposed strategy to constitute a public perfect equilibrium (PPE), it should induce a Nash equilibrium at each date t and history ht. For any history on the equilibrium path, this is trivial. Therefore the histories evaluated here will be off the equilibrium path. Both players choosing to play D in every repetition (trivial) constitutes a Nash equilibrium of the game. As shown here, the proposed strategy also constitutes 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 $\hat{t} < t$, $y^{\hat{t}} = D$: Player 1 will choose to play D in all $T \geq t$. Player 2 will also play D in all $T \geq t$. As both players choose to play D, the prescribed strategy constitutes a Nash Equilibrium.

2. In $\hat{t} < t$, $y^{\hat{t}} = C$, $x^{\hat{t}} = D$ and $s^{\hat{t}} \neq s^*$: Player 2 will choose to play D in all $T \geq t$. Player 1 will also play D in all $T \geq t$. As both players choose to play D, the strategy prescribed by the equilibrium strategy constitutes a Nash Equilibrium.

3. In $\hat{t} < t$, player 1 chooses to play D, $y^{\hat{t}} = C$, $x^{\hat{t}} = D$, $s^{\hat{t}} = s^*$: Player 2 cannot know if player 1 has chosen to play D. But as apology as cost s^* is offered by player 1 following every defection, both players will play their equilibrium strategies as if $t = 0$. Therefore the strategy of both players at t constitutes a Nash equilibrium $\forall T \geq t$ (as shown in the first part of this proof).

4. In $\hat{t} < t$, $y^{\hat{t}} = C$, $x^{\hat{t}} = D$, $s^{\hat{t}} > 0$: The strategy of both players at t constitutes a Nash equilibrium $\forall T \geq t$.

Additionally, for any history of the game that includes deviations (1) and (2) or both, the prescribed strategy requires that both players choose to defect in all future repetitions. If any of (3) or (4) or both are accompanied by (1) or (2) or both, then both players will choose to defect in all future periods. For instance, if player 1 deviates as in (3), but player 2 deviates as in (1), then the prescribed strategy at t is to choose to defect in all $T \geq t$. ■

A.2 Proof of Proposition 2

As an apology is a self-inflicted cost, player 1 would choose the lowest possible value of s^* that 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. ■

A.3 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:

The numerator: $w - h - d + l$

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.4 Proof of Proposition 3

The efficient outcome of this game is (E^1_{CC}, E^2_{CC}). 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:

The Net Present Expected (NPE) value of the total cost that player 1 incurs in $t = T$ due to a defection in $t = T$, be x. This cost must be sufficient to deter defection by player 1.
Now, consider that the strategy mandates an apology every 2 defections, with s being the cost of apology every time it is offered. Then at every $t = T$, $x = \frac{NPE(s)}{2}$, where $NPE(s)$ is the net present expected value of s in T.
Alternately, If the equilibrium strategy instead requires the cost of the apology for a defection in T to be distributed over two periods such that an apology is offered in $T + 1$, at cost of s and $T + 2$ at a cost of $\hat{s}$. Then, assuming no further defections in these 2 periods, $x = (\delta s + \delta^2 \hat{s})$. x is therefore the total discounted (expected) cost of each defection that player 1 makes in the game. In the sorry equilibrium, 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, it must meet 2 conditions:

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

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

$$\sum_{t=1}^{\infty} \delta^t (E^1_{DC} - px) \leq \sum_{t=1}^{\infty} \delta^t (E^1_{CC} - (1 - p)x) \tag{A.4}$$

Such that it is not too cheap: player 1 must be better off choosing to play C and apologizing, than choosing to play D and apologizing if D gets played.
But conditions A.3 and A.4 are the same as equations 1 and 2 in section 3. 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 equilibria, the lowest cost sorry equilibrium, must have, $x = s^{**}$. x must be sufficient to deter defection in each period and this cost depends on the primitives of the model. So, irrespective of how this cost is distributed across time or probability distributions, the total value of this cost must reflect the value underlying parameters and therefore must always conform to IC 1 and IC 2. ■

A.5 Proof of Proposition 4

As shown in A.4, in all possible sorry equilibria the cost of defection x equals 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.6 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 $K = E^1_{DD} + E^2_{DD}$)

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 $K = E^1_{DD} + E^2_{DD} = pd + (1-p)l + pd + (1-p)w$.

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 met[42]:

1. As there is always some δ that meets EC 2, Player 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 (π_C): $\sum_{t=0}^{\infty} \delta^t p^t (ph + (1-p)w) + \sum_{t=0}^{\infty} \delta^t (1-p^t) E^1_{DD}$

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

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

Proof of Proposition 5

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

Where $(w-h)$ is the cost of apology in the sorry equilibrium. To determine the condition:

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

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:

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

Simplifying the expression gives:

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

[42] This paper does not posit grim trigger strategies as an equilibrium for this game.

Cancelling and re-arranging using $(w-h) = s^{**}$ gives the expression in Proposition 4:

$2p\delta(h-d) > (1-\delta p)s^{**}$. ■

A.7 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 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 [43]:

1. If player 2 plays D, the game ends with certainty at $t = 0$. Player 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 if $O_s > 0$ by assumption. Thus there is some $\delta \in (0,1)$ for which both players choose to play C at $t = 0$. ■

Proof of Proposition 6

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

Where $(w-h)$ is the cost of apology in the sorry equilibrium. 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:

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

[43] This paper does not posit ostracism as an equilibrium for this game.

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, re-arranging terms using $(w-h) = s^{**}$, gives the expression in Proposition 6:

$2p\delta h + (1-p)\delta(w+l) + (1-\delta)O_s > (1-\delta p)s^{**}$. ■

A.8 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, inequalities to hold true, it must be shown that:

(i) If $\tau < \acute{\tau}$: $\frac{(2p-1)}{p(\acute{\tau}+\tau)-\acute{\tau}} > \frac{1}{\tau}$.

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 $(p-1)$, and changing the direct of inequality (as $(p-1) < 0$):

$\tau < \acute{\tau}$

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

(ii) If $\tau > \acute{\tau}$: $\frac{(2p-1)}{p(\acute{\tau}+\tau)-\acute{\tau}} < \frac{1}{\tau}$.

This inequality can be proven using the same method as in (i). Therefore, if $\tau > \acute{\tau}$, then $I > I^*$. ■

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

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

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.

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

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 $(w-h-d+l) > c$, player 2 plays D, when player 1 plays C.

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: $E^2_{DD} - p\tau I \geq E^2_{CC} + (1-p)\tau I$.

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

Using the value of I and simplifying:

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

This condition will always be met as: $p(w-h-d+l) < c$ and $(1-p) < p$ (as $p > 0.5$). ■

Proof of Lemma 7

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

Consider $\frac{1}{p(\acute{\tau}+\tau)-\acute{\tau}}$: Given that $\acute{\tau} < \tau$, in order for the difference between τ and $\acute{\tau}$ to increase, either τ increases or $\acute{\tau}$ decreases or both.

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

If $\acute{\tau}$ decreases (by a unit): Change in denominator is $(p-1)$, but as $p<1$, the denominator increases and value of I^* decreases.
If both change simultaneously, (as $\tau > \acute{\tau}$) both effects reinforce each other and the denominator increases and value of I^* decreases. ■

B Appendix B

B.1 Symmetric Case: Both players make mistakes

In this case neither player can ensure with certainty which action will play out. So, just like player 1 in secion 2, here if either player chooses to play C, then C gets played with probability p and D gets played with probability $(1-p)$. Similarly, if a player chooses to play D, then D gets played with probability p and C gets played with probability $(1-p)$. It is assumed that the mistakes of the players are independent of each other and that players can only observe actions that get played and not the action they choose to play. This leads to a symmetric payoff matrix:

Table 3: Expected Payoffs in the Stage Game

	Cooperate (C)	Defect (D)
Cooperate (C)	E_{CC}, E_{CC}	E_{CD}, E_{CD}
Defect (D)	E_{DC}, E_{DC}	E_{DD}, E_{DD}

Where:

$E_{CC} = p^2h + p(1-p)l + p(1-p)w + (1-p)^2d$
$E_{DC} = p^2w + p(1-p)h + p(1-p)d + (1-p)^2l$
$E_{CD} = p^2l + p(1-p)h + p(1-p)d + (1-p)^2w$
$E_{DD} = p^2d + p(1-p)w + p(1-p)l + (1-p)^2h$

Additionally, assumption (5) can be relaxed in the model from section 2. This because $1 > p > 0.5$ is sufficient to ensure that $E_{CC} > E_{DD}$ in this case. The prescribed equilibrium strategy for both players is similar to the strategy of player 1 in section 3 with some modifications. Both players know the *action that played out* and if an apology was made (public information). The actions that can be played out by both players is {C,D}. If x^t is the action played out by player 1, y^t is the action of player 2, s_1^t and s_2^t is the cost of apology in t, then the public history of the game at t is $h^t = \{(x^0, y^0, s_1^0, s_2^0), (x^1, y^1, s_1^1, s_2^1), \ldots, (x^{t-1}, y^{t-1}, s_1^{t-1}, s_2^{t-1}\}$.
Stage 1 - Choose to play C at $t = 0$. In $t \geq 1$:

- Choose to play C f h^t is such that $\forall \hat{t} < t, (x^{\hat{t}}, y^{\hat{t}}, s_1^{\hat{t}}, s_2^{\hat{t}})$ is:

- $(C, C, s_1^{\hat{t}}, s_2^{\hat{t}})$; or
- $(D, C, s^*, s_2^{\hat{t}})$ or $(C, D, s_1^{\hat{t}}, s^*)$ or (D, D, s^*, s^*)

- Else, choose to play D.

Stage 2 - In $t \geq 0$:

- $s^t = s^*$ if $x^t = D$ and if h^t is such that $\forall \hat{t} < t, (x^{\hat{t}}, y^{\hat{t}}, s_1^{\hat{t}}, s_2^{\hat{t}})$ is:
  - $(C, C, s_1^{\hat{t}}, s_2^{\hat{t}})$; or
  - $(D, C, s^*, s_2^{\hat{t}})$ or $(C, D, s_1^{\hat{t}}, s^*)$ or (D, D, s^*, s^*)

- Else, do not apologize

Given this, the incentive compatibility conditions for both the players are also very similar to the conditions derived in section 3.

Stage 2: Both players must be better off apologizing when a defection happens. In this case however, an apology must be made in the case of both a unilateral and a bilateral defection:
Unilateral Defection: $w + \sum_{t=1}^{\infty} \delta^t E_{DD} \leq (w - s) + \sum_{t=1}^{\infty} \delta^t (E_{CC} - p(1-p)s - (1-p)^2 s)$
Bilateral Defection: $d + \sum_{t=1}^{\infty} \delta^t E_{DD} \leq (d - s) + \sum_{t=1}^{\infty} \delta^t (E_{CC} - p(1-p)s - (1-p)^2 s)$
The resulting condition is however equivalent for the 2 cases:

IC 3: $s^* \leq \frac{\delta(E_{CC} - E_{DD})}{(1-\delta p)}$

Stage 1:: Both players must be better off must be better off choosing to play C and apologizing if D gets played, than choosing to play D and apologizing if D gets played (assuming IC 3 holds).

$(E_{DC}^1 - p^2 s - p(1-p)s) \leq (E_{CC}^1 - p(1-p)s - (1-p)^2 s)$

IC 4: $s^* \geq \frac{(E_{DC} - E_{CC})}{(2p-1)}$

The only condition for the equilibrium to exist is that the upper bound on the cost of apology should be larger than the lower bound.

EQ 3: $\frac{\delta(E_{CC} - E_{DD})}{(1-\delta p)} \geq \frac{(E_{DC} - E_{CC})}{(2p-1)}$

Result 1: If $h(p - (1-p)^2) - dp^2 \geq wp(1-p) - l(1-p)^2$, for every p, there exists a $\delta(p)$ such that the proposed strategy profile constitutes a Nash Equilibrium for all $\delta \geq \delta(p)$.

Proof:

$E_{CC} - E_{DD} = (2p-1)(h-d)$ and $E_{DC} - E_{CC} = (2p-1)(p(w-h) + (1-p)(d-l))$:

Inputting these values in EQ 3:

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

Opening brackets and taking δ to LHS:

$\delta \geq \frac{p(w-h)+(1-p)(d-l)}{wp^2 - h(1-p)^2 + d(1-p-p^2) - lp(1-p)}$

For the equilibrium to exist, RHS must be less than 1. This is true if:

$wp^2 - h(1-p)^2 + d(1-p-p^2) - lp(1-p) \geq p(w-h) + (1-p)(d-l)$

Aggregating common terms:

$h(p-(1-p)^2) - dp^2 \geq wp(1-p) - l(1-p)^2$

Now, $p-(1-p)^2 > p(1-p)$ and $p^2 > (1-p)^2$ as $1 > p > 0.5$, this condition can be met for a given p if the benefit from defecting when the other player cooperates $(w-h)$ and the saving from defecting when the other player defects $(d-l)$ are not too large. ■

This is a weaker result compared to proposition 1 where the existence of the equilibrium does not depend on the relative size of the payoff parameters. The additional constraints on the payoffs is intuitive as in the symmetric case, the expected benefit of continued cooperation is lower and the cooperative outcome gets played with a probability of p^2 (compared to p in section 2).

Result 2: In equilibrium, $s^{**} = \frac{(E_{DC}-E_{CC})}{(2p-1)} = p(w-h) + (1-p)(d-l)$.

The cost of apology in equilibrium also changes when both players can make mistakes. It now takes into account uncertainty (p) and the possibility of bad outcomes (d and l), as the other player can also defect. The interesting part of the comparison between the costs of apology across the symmetric and asymmetric case is that the cost is the same if $(w-h) = (d-l)$. Therefore, if the payoffs are such that the benefit from defecting when other player cooperates is larger then the benefit from defecting when the other player also defects ($(w-h) > (d-l)$), then the cost is higher in the assymetric case. Alternately if $(d-l) > (w-h)$ then the cost of apology is higher in the symmetric case as each player always has an incentive to defect when the other player defects in the future and the cost of apology must counteract this incentive.

References

Abeler, J., Calaki, J., Andree, K., & Basek, C. (2010). The power of apology. *Economics Letters*, 107(2): 233-235.

Abreu, D., Pearce, D.,& Stacchetti, E. 1990. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* 58(5): 1041-1063.

Ambrus, A., & Greiner, B. 2012. Imperfect public monitoring with costly punishment: An experimental study. *The American Economic Review* 102(7): 3317-3332.

Aoyagi, M., & Frechette, G. 2009. Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic theory* 144(3): 1135-1165.

Brown, J. P. 1973. Toward an economic theory of liability. *The Journal of Legal Studies* 2(2): 323-349.

Compte, O. 1998. Communication in Repeated Games with Imperfect Private Monitoring. *Econometrica* 66(3): 597-626.

Crawford, V. P., & Sobel, J. 1982. Strategic information transmission. *Econometrica* 50(6): 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* 99(1): 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* 10(3): 103-118.

Fischbacher, U., & Utikal, V. 2013. On the acceptance of apologies. *Games and Economic Behavior* 82: 592-608.

Fudenberg, D., & Maskin, E. 1986. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54(3): 533-554.

Fudenberg, D., David, L., and Maskin, E. 1994. The Folk Theorem with Imperfect Public Information. *Econometrica* 62(5): 997-1039.

Fudenberg, D., Rand, D. G., & Dreber, A. 2012. Slow to anger and fast to forgive: Cooperation in an uncertain world. *The American Economic Review* 102(2): 720-749.

Galanter, M. 1985. Legal Torpor: Why So Little Has Happened in India After the Bhopal Tragedy. *Texas International Law Journal* 20: 273-294.

Gale, D., & Hellwig, M. (1985). Incentive-compatible debt contracts: The one-period problem. The Review of Economic Studies, 52(4), 647-663.

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. 2006. *Institutions and the path to the modern economy: Lessons from medieval trade.* Cambridge, UK: Cambridge University Press.

Haley, J. O. 1982. Sheathing the sword of justice in Japan: An essay on law without sanctions. *Journal of Japanese Studies* 8(2): 265-281.

Haley, J. O. 1995. *Authority without power.* Oxford: 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.

Hermalin, B. E., Avery, W. K., and Craswell, R. 2007. Contract Law. In *Handbook of Law and Economics, Vol. 1.*, edited by Polinsky, A. and Shavell, S., 3-138. Amsterdam: Elsevier.

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`

Jos Ganuza, J., Gomez, F., & Robles, M. (2016). Product liability versus reputation. *The Journal of Law, Economics, and Organization* 32(2): 213-241.

Kandori, M., & Matsushima, H. 1998. Private Observation, Communication and Collusion. *Econometrica* 66(3): 627-652.

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* 85(12): 64-95.

Krauss, M. I. 2012. Tort Law, Moral Accountability, and Efficiency: Reflections on the Current Crisis. *Journal of markets & morality* 2(1): 114-124.

Krishnan, J. K., & Kumar, C. R. 2010. Delay in process, denial of justice: the jurisprudence and empirics of speedy trials in comparative perspective. *Georgetown Journal of International Law* 42: 747-784.

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

Mungan, M. C. 2012. Don't Say You're Sorry Unless You Mean It: Pricing apologies to achieve credibility. *International Review of Law and Economics* 32(1): 178-187.

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

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.

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.

Polinsky, A. M., & Shavell, S. 2010. The uneasy case for product liability. *Harvard Law Review* 123: 1437-1493.

Porter, R. H. 1983. Optimal cartel trigger price strategies. *Journal of Economic Theory* 29(2): 313-338.

Ramseyer, J. M., & Rasmusen, E. 2010. *Comparative litigation rates.* (No. 681) Harvard John M. Olin Discussion Paper Series.

Rasmusen, E. 1996. Stigma and self-fulfilling expectations of criminality. *Journal of Law and Economics* 39(2): 519-543.

Robbennolt, J. K. 2003. Apologies and legal settlement: an empirical examination. *Michigan Law Review* 102(3): 460-516.

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

Scheff, T. J. 1998. Community conferences: Shame and anger in therapeutic jurisprudence. *Revisita Juridicia U.P.R.* 67: 628-635.

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

Schlenker, B. R., & Darby, B. W. 1981. The use of apologies in social predicaments. *Social Psychology Quarterly* 44(3): 271-278.

Schneider, C. D. 2000. What it means to be sorry: The power of apology in mediation. *Mediation Quarterly* 17(3): 265-280.

Selten, R. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4(1): 25-55.

Shuman, D. W. 1999. The role of apology in tort law. *Judicature* 83(4): 180-189.

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.

Stephens, M. A. 2008. I'm Sorry: exploring the reasons behind the differing roles of apology in american and japanese civil cases. *Widener Law Review* 14: 185-204.

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, UK: Basil Blackwell.

Tabellini, G. 2008. The scope of cooperation: values and incentives. *The Quarterly Journal of Economics* 123(3): 905-950.

Tavuchis, N. (1991). *Mea culpa: A sociology of apology and reconciliation.* Stanford: Stanford University Press.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. Journal of Economic theory, 21(2), 265-293.

Wagatsuma, H., & Rosett, A. 1986. The implications of apology: Law and culture in Japan and the United States. *Law and Society Review* 20(4): 461-498.

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 Law Review* 91: 1261-1311.

Williams, K. D. 1997. Social ostracism. In *Aversive interpersonal behaviors*, edited by Kowalski Robin M., 133-170. Boston, MA: Springer US.

Williams, K. D. 2002. *Ostracism: The power of silence.* New York: Guilford Press.

Wu, J., & Axelrod, R. 1995. How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict Resolution* 39(1): 183-189.

Yaffe, G. 2010. Intention in Law. In *A Companion to the Philosophy of Action*, edited by O'Connor T. and Sandis C., 338-344. Oxford, UK: Wiley-Blackwell.