In praise of forgiveness: Ways for repairing trust breakdowns in one-off online interactions

Asimina Vasalou\textsuperscript{a,*}, Astrid Hopfensitz\textsuperscript{b}, Jeremy V. Pitt\textsuperscript{a}

\textsuperscript{a}Electrical and Electronic Engineering, Imperial College London, Exhibition Road, SW7 2BT London, UK
\textsuperscript{b}Toulouse School of Economics (GREMAQ), Manufacture des Tabacs, 21 Allée de Brienne, 31000 Toulouse, France

Received 13 June 2007; received in revised form 22 January 2008; accepted 12 February 2008

Communicated by S. Wiedenbeck
Available online 10 March 2008

Abstract

Online offences are generally considered as frequent and intentional acts performed by a member with the aim to deceive others. However, an offence may also be unintentional or exceptional, performed by a benevolent member of the community. This article examines whether a victim’s decrease in trust towards an unintentional or occasional offender can be repaired in an online setting, by designing and evaluating systems to support forgiveness. We study which of three systems enable the victim of a trust breakdown to fairly assess this kind of offender. The three systems are: (1) a reputation system, (2) a reputation system with a built-in apology forum that may display the offender’s apology to the victim and (3) a reputation system with a built-in apology forum that also includes a “forgiveness” component. The “forgiveness” component presents the victim with information that demonstrates the offender’s trustworthiness as judged by the system. We experimentally observe that systems (2) and (3), endorsing apology and supporting forgiveness, allow victims to recover their trust after online offences. An apology from the offender restores the victim’s trust only if the offender cooperates in a future interaction; it does not alleviate the trust breakdown immediately after it occurs. By contrast, the “forgiveness” component restores the victim’s trust directly after the offence and in a subsequent interaction. The applicability of these findings for extending reputation systems is discussed.

Keywords: Trust; Social dilemmas; Forgiveness; Reputation; Apology; One-off interactions

1. Introduction

Trust is a social lubricant for computer-mediated communication (CMC) enhancing collaboration, cooperation and information exchange (Rocco et al., 2000; Bos et al., 2002; Ridings et al., 2002). It is thus not surprising that a large body of trust research has focussed on the factors that engender trust. For instance, in online auctions, the effect of sellers’ reputation on buyers’ trust, as manifested in their bidding offers, has been repeatedly investigated (e.g. Ba and Pavlou, 2002; Lucking-Reiley et al., 2007; Resnick et al., 2006). This body of work has been important as it has offered prescriptions for the design of trust-enabling social systems. Despite the value of this approach though, the current debate has failed to fully address the need for repair mechanisms, offered when trust breaks down.

This oversight can be partly attributed to a research focus on intentional and frequent acts performed by an offender with the aim to deceive others. For example, impersonators often become part of close-knit communities under a contrived identity. When discovered, identity deception can damage the trust cultivated within the wider community as users begin to question each others’ motives (Grady, 1998; Joinson and Dietz-Uhler, 2002). In addition, “trolls”, who are seemingly sincere users, seek simply to provoke a reaction by posting contentious comments which challenge a community’s commonly held beliefs (Donath, 1998). To eradicate such intentional and recurrent offences, social system designers have created tools...
that allow communities to self-regulate: in newsgroups, malignant members can be filtered out, their comments ceasing to affect others; in online markets, a fraudulent seller might be assigned a low reputation rating, thereby being driven out of the market. If the behaviour escalates, explicit measures can be taken by appointed individuals from the community stakeholders (e.g. in newsgroups intervention by a moderator; in Wiki, an administrator restoring original pages). In the last resort, there is also recourse to legal action.

As opposed to this dominant interest in intentional breakdowns, some researchers have alluded to certain situations during which one party may violate another’s trust unintentionally (e.g. Riegelsberger et al., 2005a). For instance, an expected action may not be executed because the trusted party is unable to perform or does not have the required ability to do that action: as an example, in a remote collaboration project, trust was endangered as senior members of one remote team made false assumptions about their junior counterparts’ skills (Rocco et al., 2000). Moreover, the fulfilment of obligations may be prevented by unforeseen events that are outside the trusted party’s control. A promised delivery might be delayed due to slow postal services (Riegelsberger et al., 2005a). Online markets recognise these possibilities, and buyers are advised to be patient, reverting to negative feedback only when necessary (Kollock, 1999); despite these words of caution, eBay sellers still report distressing experiences due to unwarranted or disproportionate negative feedback (Khopkar et al., 2005). Also, when placed in a social dilemma under conditions of anonymity, as it happens for instance in an e-commerce environment, members are more likely to “defect”, i.e. not cooperate, and by doing so to damage the well-being of others (Bos et al., 2002; Zheng et al., 2002; Vasalou et al., 2006a). However, although a member may “slip” once, it has been shown that when alerted, most will repair the damage by apologising and correcting their future behaviour (Vasalou et al., 2006a).

The above scenarios concern a different kind of offender: a benevolent member of the community who may have breached a norm unintentionally, or slipped once, regretting his/her behaviour thereafter. In face-to-face communication, proximity allows the offender to apologise, to elaborate on his/her intentions and to repair the breakdown, thus paving the way towards forgiveness (Boon and Sulsky, 1997; McCullough et al., 1998). Moreover, non-verbal expressions (e.g. blushing) given off by the offender may supersede and complement words of regret (Castelfranchi and Poggi, 1990; Miller, 1996; Keltner and Buswell, 1997). Anonymous, one-off encounters which are encouraged in many online settings add complexity in resolving such offences. This is partly due to the narrow time window of each interaction in combination with the impoverished communication channel constraining the cues of trustworthiness one can acquire about another member (e.g. integrity, willingness to comply to institutions, benevolence; Riegelsberger et al., 2005a). At the onset, this establishes interactions that are perceived as more risky, thus building emotional barriers that may stand in the way of resolution if trust breaks down. Within this high-risk interaction, reputation systems, which operate to maintain trust, are not effectively designed to repair trust. In our view, reputation systems have taken little provision to encourage the repair of trust breakdowns; we are aware of only two approaches that have been recently proposed or implemented in this context.1

A first approach has been introduced in eBay’s electronic marketplace: using a feature called the “mutual feedback withdrawal” users can retrospectively contest the reputation score they received. Only if both the victim and offender agree to engage in this process, then the resolution of the issue is taken offline. Later, the victim of the offence may retract and improve the offender’s online reputation score. At heart, eBay offers an online offender an outlet through which to apologize, elaborate on his/her intentions and repair, subsequently allowing the victim of the offence to restore the trust by removing the original low reputation rating. In a second approach, Vasalou and Pitt (2005) and Vasalou et al. (2006b) have proposed to facilitate resolution by embedding an intelligent “forgiveness” component within reputation systems. In their conceptualisation, upon a breakdown in trust, the system will detect the trustworthiness of the offender by taking into account a number of factors. Only if the offender is judged positively by the system, the victim is presented with those factors to consider before assigning a reputation score to the offender. Essentially, this intervention intends to alleviate the victim’s possibly negative attributions (e.g. the offender intended to perform the offence) while at the same time it aims to shield the unintentional/infrequent offender from receiving an unjust punishment, i.e. in the form of a low rating. However, it is as yet unclear whether either of these approaches, as compared to a basic reputation system, is useful for facilitating repair after an unintentional act of an infrequent offender, who happens to be a valued member of the community. This article sets out to answer this question by comparing an “apology” and a “forgiveness” approach against a reputation system.

Before we begin to discuss how to repair trust, Section 2 defines online trust. We go on to show how trust is sustained during one-off interactions with the use of reputation systems. An example is given to demonstrate the harmful consequences of reputation mechanisms that function without the operation of a reparative mechanism.

1Though repairing trust breakdowns in one-off mediated interactions have not received adequate attention, we acknowledge two related efforts in the wider field of human-computer interaction. In ambient environments agents pursue their users’ interests in an autonomous manner. Conflicts that inevitably emerge can be reduced by integrating forgiveness and regret into the trust framework that governs the agents’ actions (Briggs and Marsh, 2006). In using computer applications, users often encounter error messages which can temporarily obstruct their work. Apologetic messages have been tried out as a way to alleviate users’ frustration during these system errors (Tseng, 2004).
Next, forgiveness is presented as a process that facilitates repair in face-to-face interactions (thus also a possible mechanism for resolving online offences). The remainder of this article presents an empirical experiment that investigates which of three systems elicits forgiveness, and in turn repairs trust, during an unintentional offence of an otherwise benevolent member of the community.

2. Background

2.1. Online trust

Online trust has been defined as “an attitude of confident expectation in an online situation of risk that one’s vulnerabilities will not be exploited” (Corritore et al., 2003). Trust research promises to dampen users’ vulnerability towards potentially risky online technologies and computer-mediated exchanges with the provision of trust-enabling information. Thus, a range of models have been proposed and tested empirically, addressing context-specific risks and different sources of vulnerability for a number of scenarios (e.g. transactional or informational web sites: Corritore et al., 2003; e-commerce: McKnight et al., 2002; phishing attacks: Kumaraguru et al., 2006; computer-mediated communication: Riegelsberger et al., 2005a).

To our current interest, trust has been shown to enhance collaboration, cooperation and information exchange in computer-mediated communication. In online communities, users are more likely to trust anonymous others who divulge personal information and who actively respond to their posts (Ridings et al., 2002; Joyce and Kraut, 2006). The risk carried in online social dilemmas, where one actor’s personal gain may harm another member of the community, is diminished when interactions take place over rich media such as video and audio compared to anonymous exchanges (Bos et al., 2002). Rich media increases members’ reported trust levels as well as collective rates of cooperation. Furthermore, trust in remote team collaborations can be increased with non-work related activities that create a shared group identity and cross-cultural understanding (Rocco et al., 2000; Zheng et al., 2002). Also, reputation systems are used to elicit feedback on participants’ past behaviour which is subsequently made available to other members of the community (Resnick et al., 2000). In e-commerce, for example, this allows a potential buyer to evaluate a seller’s trustworthiness by regarding the seller’s reputation.

Trust has been broadly positioned within a semantic dichotomy: cognitive and emotional (Rocco et al., 2000; Corritore et al., 2003; Riegelsberger et al., 2005a). While cognitive trust is mostly understood as a change in attitude due to rational factors such as the reliability and ability of another party, emotional trust comes as a result of affective bonds that may develop between people. We briefly illustrate how the two dimensions of trust may apply to an online market. Before engaging in a purchase decision, a buyer may choose to interact with a particular seller due to the low risk suggested by his high reputation score (cognitive trust). In the same setting, social information, such as a photograph or an off-topic discussion, exchanged prior to the transaction may also increase the buyer’s trust towards the seller, although the presence of this information may not be directly related to the task at hand (emotional trust) (Rocco et al., 2000; Zheng et al., 2002).

Although cognitive and emotional trust co-occurs, at times one facet of trust can prevail over the other (Corritore et al., 2003). This point is illustrated by one study showing that emotional trust may override cognitive trust, depending on the quality of interpersonal cues displayed by the trusted party, in this case an online advisor (Riegelsberger et al., 2005b). Riegelsberger et al. (2005b) confronted users with a series of risky transactions in advice uptake by giving them the choice of a “high ability” and a “low ability” advisor. Users were allocated to four different conditions each of which offered the choice of either a rich media advisor (avatar, video, audio, photo with text) or a text-only advisor, counterbalanced by ability. Overall, cognitive trust as increased by the advisor’s exhibited ability led participants to discriminate between the high- and low-quality advisors, therefore choosing high-quality advice independent of its format. However, when users were faced with a final high-risk choice, video and audio advice was preferred despite the advisor’s low ability, suggesting a potential bias of emotional trust.

In the next section we discuss the value of reputation systems in forecasting trust during mediated interactions. We then provide evidence for extending these systems in favour of repair.

2.2. Reputation systems

Reputation systems “collect, distribute and aggregate feedback about participants’ past behaviour” (Resnick et al., 2000). In an anonymous environment that encourages one-off exchanges, tracking members’ behaviour through time can become the prevailing indicator of a member’s trustworthiness. Indeed, in online auctioning, reputation history is used to signal a potential buyer on whether a particular seller should be trusted. As a result, low-quality sellers are forced to accept bids that reflect the real quality of their services (Lucking-Reiley et al., 2007), in some cases being driven out of the market.

The main focal point of reputation research has been to protect the “system” by distributing information that is accurate and thus “trustworthy”. For example, reputation scores are collective and equal aggregations of all scores given by any number of members. Several researchers assess evaluators’ previous quality of ratings by comparing them against the distribution of other members’ ratings. By using this method, members whose ratings converge with the common consensus are distinguished and given precedence, thus increasing the trustworthiness of the collective reputation score (Chen and Singh, 2001; Fujimura and Nishihara, 2003). Moreover, reputation systems
are effective only when users volunteer feedback. The elicitation of feedback is considered to be one of the prevailing challenges of these systems (Resnick et al., 2000). Dellarocas et al. (2003) have shown that the bidirectional feedback system of eBay successfully encourages reciprocal feedback exchanges: usually sellers will rate buyers first, with the hope of receiving feedback in return. Another persisting problem for reputation systems is a fraudulent member’s ability to rebuild his or her status in the community, for instance as is currently allowed in eBay. Low-quality sellers whose sales are declining due to a low reputation may start fresh by joining as new members; however, stable identities which users would bear for the lifetime of their participation could encourage members to sustain a good reputation (Friedman and Resnick, 2001).

To our present interest, Khopkar et al. (2005) examined the effect of one negative feedback on the subsequent behaviour of eBay sellers and buyers. Sellers who received one negative feedback following a transaction were 25% less likely to post new products than sellers who had no prior negative feedback, although this effect diminished over time. The impact of one or two negative feedback comments on sales is insignificant as compared to the financial loss one suffers when joining as a new member without a reputation (Resnick et al., 2006). Therefore, the authors of this work conclude that sellers were well-intended members who abandoned their identity due to feeling unappreciated and unfairly treated. In support of this, another study showed that participants, who were unfairly punished as a response to an unintentional offence, were less likely to further cooperate than participants who were forgiven (Kelln and Ellard, 1999). The presence of negative feedback as given to sellers also led to a phenomenon Khopkar et al. call stoning, whereby buyers become more suspicious of a seller’s intentions. Buyers viewing sellers’ negative feedback were less likely to volunteer positive feedback upon the completion of a successful transaction.

Khopkar et al.’s findings taken together suggest that reputation systems alone do not assist in preventing the unfair treatment of an infrequent and/or unintentional offender. Indeed, reputation systems seem to be based on the cognitive aspects of trust. The information they convey concerns members’ ability and performance, hence all the rational reasons for initiating a transaction in e-commerce or for accepting a member’s advice in an online forum. Yet, when trust breaks down, it is unclear whether the presence of a static reputation score can motivate forgiveness from the victim; the offender’s previously obtained high reputation may not be a salient cue for signalling his/her good intentions, regret and willingness to repair during a present offence.

The following section discusses the process of forgiveness by which reconciliation is reached in face-to-face interactions. When trust breaks down, the same aspects that highlight cognitive trust in reputation systems (e.g. frequency of prior offences and utilities) can also motivate forgiveness. However, forgiveness can result from apology and benevolence which are based on similar mechanisms as emotional trust.

2.3. Forgiveness

When an offence occurs, the victim of the offence might seek to avoid further contact with the offender and to withdraw from the relationship by adding distance (avoidance). The victim might also take direct actions that will harm the offender (revenge). When forgiveness is granted, the victim replaces these two tendencies with constructive behaviours that benefit the offender. Thus, forgiveness is a prosocial process during which negative motivations towards the offender are reduced and replaced with positive motivations (McCullough et al., 1997, 1998; McCullough, 2001). The positive motivations that override this initial negative disposition can be influenced by (1) the severity of the current offence, (2) the frequency/severity of previous offences, (3) the offender’s intent, (4) the offender’s apology and efforts to repair and (5) previous interactions with the offender during which he/she has demonstrated benevolence.

Specifically, following an offence, the severity of the action is first considered; more serious violations lead to harsher judgments (Boon and Sulsky, 1997). Moreover, the offender’s past actions are compared to the current event. Frequency and severity of past actions impact one’s inclination to forgive (Buss, 1980). Furthermore, actions of apparent intent lead to more negative judgments while low intent actions support more positive judgments (Manstead and Semin, 1981; Boon and Sulsky, 1997). In addition, after a truthful apology, or a good deed that reverses the offence, the victim is more likely to perceive the offender as trustworthy and honest (Weiner et al., 1991); at the same time, the victim’s emotions of anger are reduced and forgiveness is increased (Witvliet et al., 2002). Apology additionally evokes empathy towards the offender which has been shown to predict forgiveness (McCullough et al., 1997). Likewise, non-verbal signs can substitute spoken words; visibly embarrassed offenders (e.g. blushing) elicit more empathy and are in turn more likely to be forgiven (Castelfranchi and Poggi, 1990; Miller, 1996; Keltner and Buswell, 1997). Finally, prior interactions with the offender during which s/he may have shown benevolence increase the likelihood of forgiving the current offence (McCullough et al., 1998).

It should be noted that forgiving a single offence does not necessarily entail revising one’s attitude as a whole (Exline et al., 2003). While a certain violation may be forgiven, other past behaviours may still impede one’s trust towards another. Therefore, despite popular definitions, forgetting, condoning, trusting or removing accountability is not always considered to be a part of forgiveness.

There are manifold reasons for seeking to resolve disrupted interactions by way of forgiveness. Forgiveness mediates and resolves conflicts to sustain healthy long-term
relationships (McCullough et al., 1997). A forgiving attitude promotes cooperation when compared to a vengeful outlook2 (Axelrod, 1984). In the long term, forgiveness has been shown to positively affect health; withholding forgiveness and thus ruminating about an offence increases stress levels and can impede on the healthy functioning of the immune system (for a review, see Worthington and Scherer, 2004). For current purposes, forgiveness issued during an unintentional offence stimulates the offender into voluntary actions of repair (Kelln and Ellard, 1999).

The five factors that motivate forgiveness as outlined above can only partly apply to anonymous, one-off interactions in CMC. To begin with, the narrow time window of each interaction does not allow members to build common experiences during which benevolence may develop. Also, non-verbal signals, such as the blush, which conveys regret in face-to-face communication, are not communicated as a result of the sometimes impoverished communication channel (Parkinson, in press). Therefore, only the four factors severity of current offence (1), frequency and severity of previous offences (2), intent (3), apology and willingness to repair (4), apply to CMC. The following section discusses how these relevant motivating factors are supported by the design of three different systems. Several hypotheses are drawn about each system’s ability to repair a trust breakdown.

3. Objectives and hypotheses

The aim of this study was to investigate which of three systems repairs a breakdown of trust by encouraging the victim to forgive and to (re-)trust an unintentional and infrequent offender; we focused on trust breakdowns occurring in one-off online interactions. The three systems contrasted were:

(1) A reputation system that demonstrated the offender’s previous good standing with other members. This kind of system allowed the victim of a trust breakdown to review the offender’s current action and the severity/utility of previous actions; therefore, the cognitive aspects of forgiveness were communicated.

(2) An apology channel that displayed the offender’s apology. The apology channel implemented in this study was a scaled down application of the eBay “mutual feedback withdrawal” feature. Firstly, the apology was sent from the offender to the victim before the victim assigned a reputation score to the offender. Secondly, the apology was offered inside the online forum rather than by other possible means, e.g. telephone, email exchanges. An apology forum integrated inside a reputation system additionally conveyed the offender’s low intent, willingness to repair and regret. This approach displayed all the possible, cognitive and affective, factors that may motivate forgiveness.

(3) A “forgiveness” component integrated into the previously mentioned apology forum which on the one hand displayed the four factors of forgiveness through the communication channel, but also communicated an objective positive forgiveness judgement by the system itself. Specifically, in implementing this third approach, we drew from the work of Vasalou et al. (2006b) who have argued for proactively promoting forgiveness in social systems. In this view, the system should collect the forgiveness motivations (e.g. apology, previous history; see Section 2.3) relevant to each member and use them to compute a measure of trustworthiness. In the case of an infrequent and unintentional offender with high computed trustworthiness, the system would intervene and present the victim with a compilation of the forgiveness factors depicting the offender’s trustworthiness. This intervention is pre-emptive to the victim’s assignment of a reputation rating. In this instance, forgiveness is also endorsed by the system.

Exline et al. (2003) warn not to equate forgiveness with the reinstatement of trust. However, in one-off interactions a victim has no prior negative experience with the offender, such as previous offences which may stand in the way of trust restoration. Thus, in the context of this study, a victim’s forgiveness for an offender was also expected to restore his/her trust in the offender.

As noted above, a reputation system is limited to supporting the cognitive aspects that promote forgiveness, such as the offender’s frequency of previous offences and benefits to others. Therefore, given the definition of forgiveness, after a violation of trust, a benevolent offender with an impeccable reputation score is more likely to be forgiven by the victim (Boon and Sulsky, 1997). However, a reputation system working in isolation is not designed to convey the remaining important motivators of forgiveness, e.g. the offender’s intent, regret or willingness to repair an offence (Weiner et al., 1991; Boon and Sulsky, 1997; McCullough et al., 1998). These latter aspects were communicated in both the apology and forgiveness conditions. Because of this, these two treatments were expected to result in higher repair rates, i.e. an increase in both forgiveness and trust, than the reputation condition.

2 Axelrod (1984) coordinated a computer tournament during which participants from 16 countries each submitted a game strategy for the prisoner’s dilemma game. After countering each rule against the other, Tit-for-Tat emerged as the most effective strategy of all. Tit-for-Tat first cooperated and then echoed the opponent’s strategy thereafter. If the opponent defected and then cooperated, then Tit-for-Tat responded with a punishment followed by cooperation. This behaviour was termed as forgiving by Axelrod. If in addition noise was introduced in the competing system, thus leading to unintentional defections, a forgiving version of Tit-for-Tat proved optimal.

3 Trust was indirectly evaluated with two measures: reputation and cooperation. In e-commerce, users generally trust those whose reputations are high (e.g. Lucking-Reiley et al., 2007). This finding is inverse and
These expected outcomes are formalised in the following three hypotheses:

**Hypothesis 1.** Participants of the apology and forgiveness treatments will report increased forgiveness towards the offender as compared to participants viewing the offender’s reputation score only.

**Hypothesis 2a.** Participants of the apology and forgiveness treatments will give the offender a higher reputation score directly after the offence and also in a future interaction than participants in the reputation treatment.

**Hypothesis 2b.** Participants of the apology and forgiveness treatments will cooperate with the offender more often, if given the chance to interact further than participants in the reputation treatment.

A second inquiry was conducted at a collective level to examine the impact of forgiveness on the victim’s emotions. Forgiveness occurs as the victim’s negative motivations towards the offender are reduced (McCullough, 2001). Fewer negative motivations, as encouraged by forgiveness, should also reduce the negative emotions (e.g. anger and irritation) the victim feels for the offender (Witvliet et al., 2002). This prediction is captured in a fourth hypothesis:

**Hypothesis 3.** Negative emotions reported by the victim for the offender will be positively correlated with unforgiveness.

4. Methods

4.1. Overview

The experiment conducted was designed on the basis of a trust framework proposed by Riegelsberger et al. (2003, 2005a) who have argued for researching trust as an asymmetric and asynchronous relationship between two parties, the trustor and the trustee. Asymmetry exists as the trustor only assumes risk in anticipation of an asynchronous fulfillment from the trustee. This relationship can be delineated by two sequential moves. The trustor makes the first move on the basis of the trustee’s perceived trustworthiness. By doing so, the trustor assumes a certain degree of risk over an uncertain outcome. The trustee, the second mover, may either fulfil or violate the first mover’s trust.

This high-level definition lends itself to a variety of online interactions each of which may pose different types of risk. For instance, a member of an online emotional support forum who posts private information expects empathy in return. In e-commerce, a buyer sends money in anticipation of the goods. Although the present research is concerned with trust developing (or breaking down) in a remote human–human interaction, it is noted that Riegelsberger’s framework also applies to other types of exchanges. For instance, the trustee may be a web site entrusted with private information (Corritore et al., 2003), or a software agent consulted for advice (Riegelsberger et al., 2005b).

The study conducted here captured the above definition with a social dilemma, the trust game (Kollock, 1998; Camerer, 2003; Riegelsberger et al., 2003), which has been used as a research paradigm in several previous CMC studies (e.g. Bolton et al., 2003; Keser, 2003; Vasalou et al., 2006a). The trust game designates one player to the role of the first mover (trustor) while a second player is the second mover (trustee). The payoffs of this game were adapted from Hopfensitz and Reuben (2005). The first mover begins with 150 points, 50 points more than the second mover. If the first mover chooses to transfer his surplus of 50 points to the second mover, the points are multiplied by 6 and become 300. The second mover has the option to confirm the first mover’s trust by sharing the fair half of the gains (cooperate), or to keep the full amount (defect). Fig. 1 displays the players’ moves and the possible payoffs.

As stated above, the aim of this research was to investigate which of three systems (reputation, apology or forgiveness) is most effective for repairing a trust breakdown (defection in the trust game). The offender (trustee) was a member with an impeccable historical track, who breached the norm (defected in the game) as a result of an error, experiencing and expressing regret thereafter. In exploring this particular scenario, it was important to retain control over the offender’s profile across the three treatments. Thus, trustees were simulated in this study and participants were always assigned to the trustor’s role.

4.2. Participants

Sixty-three students of a UK university were recruited by email invitations promising monetary reward for participation in this study. Proficiency in English was a requirement for attendance. Participants were undergraduate students of
a science and technology discipline. They were between the ages of 18 and 24; the average age was 21. All participants were experienced web users. Forty-one participants used reputation systems regularly, 12 participants had not previously used a reputation system and 9 participants chose not to disclose their prior usage of reputation systems.

4.3. Procedure

Participants arrived in three groups of 21, upon which they were led to a quiet room and seated by a computer. The room was set up in such a way that none of the participants faced one another nor could they see each others’ screens. The trust game was played using a web-based application. The instructions of the game as given to participants can be found in Appendix A.

For believability purposes, participants thought they were randomly assigned to either first movers (trustor) or second movers (trustee). However, participants were always first movers and hence required to decide between trusting the second mover by giving an initial endowment of 50 points or to withdraw from the risk by keeping the 50 points and ending the game. The second mover was always simulated in the application, each round with a new non-gender indicative pseudonym. Participants were instructed that in each round they were playing with a new remote player from the University of Geneva. To avoid end-game effects (i.e. decreased trust in the final rounds of the game), usually observed in these kinds of games (e.g. Bos et al., 2002), participants were told they would randomly play up to 10 rounds of the game. After each round, an onscreen notification informed them whether they would proceed to the next round. In the experiment, the rounds played were set to six rounds in total.

The trust game application included a “starred” reputation system; one star signalling the lowest reputation score obtainable and five stars signalling the highest. This approach is similar to the Amazon auctioning reputation system where the stars summarise the mean of reputation scores received in previous transactions (Resnick et al., 2000). Although second movers’ reputation scores in each round were simulated and held constant across participants, participants were instructed that reputation scores were calculated on the basis of other players’ ratings and were constantly updated. Whenever participants made the initial trust move, the same reputation system was made available to them for rating the second mover.

In the first four rounds of the game, participants played with a different simulated second mover. To make the second movers believable, each one displayed a different reputation score ranging from two, the lowest, to five, the highest; the second mover of round one had no reputation score displayed since he had not been previously rated. During the first four rounds, if trusted, the second movers always reciprocated the initial trust move by returning the fair half amount.

In round 5, the second mover for the first time betrayed the participant’s trust by returning nothing. At this stage, participants received one of three treatments: reputation, apology or forgiveness (see Section 4.4). At the end of round 5, participants were asked to assign a reputation score to the second mover. Immediately after, they filled out a questionnaire measuring forgiveness and negative emotions towards the offender. At the beginning of round 6, participants were informed they were randomly assigned to the same player again, i.e. to the offender of round 5. The reputation score given to the offender in round 5 by the participant, was averaged into the second mover’s displayed reputation in round 6. If trusted with the initial 50 points, the second mover repaired the trust breakdown by returning the fair half of the gains. At the end of round 6, those participants who had made the initial trust move were asked to assign a reputation score to the second mover. All participants once more responded to the emotion questionnaire as given in round 5.

The game ended after round 6. Participants received their earnings calculated from the points earned, one point equalling 0.5 pence, with an average of £7.50 per participant. They also filled out a questionnaire reporting their opinion about the purpose of the experiment. The aim of this was to ensure that participants were blind to the real purposes of the experiment and ignorant to the simulated interaction with the trustee. Table 1 summarises the six rounds of the game. To investigate treatment effects on forgiveness and trust restoration, we will discuss measures from the offence (round 5) and the subsequent interaction (round 6).

4.4. Treatments

Following the defection of round 5, 60 participants received one of the following three treatments (participants were distributed evenly by gender across the three treatments):

- **Reputation (N = 20):** In round 5, a reputation score of 5 out of 5 stars demonstrated the offender’s good standing in previous interactions with other players.
- **Apology (N = 20):** This treatment was identical to the reputation treatment. In addition, a written apology from the simulated offender was displayed. This apology communicated the offender’s regret, responsibility over the event and willingness to repair, similar to what has been done in previous related research (e.g. Weiner et al., 1991). The apology text read: “Hey [first mover ID], I am sorry for transferring nothing back. I accidentally clicked on zero absentmindedly-realised it when it was too late. Sorry! I promise I will be more careful next time.” The implementation of this approach is in line with the eBay “mutual feedback withdrawal” feature.

Three participants did not make the initial trust move in round 5 and were thus omitted from the data.
Forgiveness \((N = 20)\): The forgiveness treatment was identical to the apology treatment. However, a “forgiveness” component was also included and explicitly reminded the victim of all the factors comprising the offender's positive standing. The forgiveness text read: “Before going forward, have you considered the following? [Second mover ID] has a 5 star rating which means that other members have rated his/her contributions as being fair. The system records detect that this is the 1st time [Second mover ID] has given 0 back to another player. [Second mover ID] sent you an explanatory message which is an optional feature.” This approach was adapted from Vasalou and Pitt (2005) and Vasalou et al. (2006b).

4.5. Dependent variables

Four measures were collected:

- **Cooperation:** In the round following the defection (round 6) participants had the choice to trust the offender with the initial 50 or to keep the points and withdraw from the interaction. Cooperation was assigned a “1” if participants chose to trust and “0” if they did not.
- **Reputation score:** Directly after the offence (round 5) and after the subsequent interaction (round 6) participants rated the offender with a reputation score ranging from 1 (low) to 9 (high). If participants withheld the 50 points in round 6, we assumed that their judgment, and thus reputation score stayed constant and used the value from round 5 in the analysis.
- **Forgiveness:** To our knowledge there was no forgiveness measure available for one-off interactions such as those concerned in this research. Therefore, forgiveness was measured with an adapted version of the 12-item questionnaire, “transgression-related interpersonal motivations inventory” (TRIM). TRIM was originally designed to evaluate forgiveness in interpersonal relationships (McCullough et al., 1998; see Appendix B.1). The first five questions measure revenge towards the offender and the remaining seven questions measure avoidance. Higher scores of revenge and avoidance indicate lower levels of forgiveness. The questions were extended to reflect the tone of one-off interactions (see Appendix B.2). For example, the question “I keep as much distance between us as possible” was reformulated to “if I were to play with him/her again, I would keep as much distance between us as possible”. Each item was rated on a scale of 1 (strongly disagree) to 5 (strongly agree).
- **Negative emotions:** Following round 5 and also round 6, participants were asked to report their emotions towards the second mover. As done in previous studies measuring emotions (e.g. Hopfensitz and Reuben, 2005), the real purpose of the investigation was concealed by presenting eight different emotions, namely: anger, joy, shame, guilt, surprise, disappointment, irritation and amusement. The intensity of emotions was reported on a five-point scale anchored at 1 (low) and 5 (high). The emotion variables of interest to this study were the negative emotions of anger and irritation. Anger and irritation were summed up into one score, forming two emotion variables for round 5 and round 6.

5. Results

5.1. Validation of forgiveness measure

The first stage of the analysis focused on the reliability of the forgiveness measure, as manipulated to reflect one-off interactions. A principle component analysis of the modified 12-point forgiveness questionnaire showed two factors with Eigenvalues greater than 1.0, presumably one for revenge and one for avoidance. An examination of the communalities table (Varimax rotation) suggested that four items be omitted as they loaded on both factors \((>0.4)\). The remaining items all met the threshold criteria of 0.4 on one of the two factors. The final eight-point questionnaire consisted of four questions measuring revenge and four questions measuring avoidance (see Appendix B.3).

5.2. Analysis

A multivariate analysis was performed with condition as a factor (reputation, apology and forgiveness) on the following measures: reputation score after round 5, reputation score after round 6, cooperation and forgiveness. Table 2 summarises the results.

---

Table 1
Summary of the game

<table>
<thead>
<tr>
<th>Round</th>
<th>Second move</th>
<th>Second mover’s reputation</th>
<th>Measures collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cooperate</td>
<td>none</td>
<td>Reputation score (choice)</td>
</tr>
<tr>
<td>2</td>
<td>Cooperate</td>
<td>2</td>
<td>Reputation score (self-reports)</td>
</tr>
<tr>
<td>3</td>
<td>Cooperate</td>
<td>5</td>
<td>Emotions (self-reports)</td>
</tr>
<tr>
<td>4</td>
<td>Cooperate</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Defect</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Cooperate</td>
<td>Reputation score given in round 5 averaged into score displayed in round 6</td>
<td>Willingness to trust (choice)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reputation score (choice)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Emotions (self-reports)</td>
</tr>
</tbody>
</table>
The main effect of condition (reputation, apology and forgiveness) was significant \((F(54,110)=2.69, p=0.009)\). There was a main effect of condition on cooperation \((F(2,57)=8.14, p=0.001)\), reputation score after round 5 \((F(2,57)=5.26, p=0.008)\) and reputation score after round 6 \((F(2,57)=7.17, p=0.002)\). Condition on forgiveness was non-significant. Post hoc tests using the Tukey criterion contrasted reputation to the apology and forgiveness conditions.

After round 5, participants completed a measure of forgiveness. The Tukey post hoc test revealed that participants in the apology condition did not report significantly more forgiveness than those in the reputation condition. By contrast, the forgiveness measure was significant in the forgiveness condition \((p=0.05)\) as compared to the reputation condition. Therefore, Hypothesis 1 was only partly confirmed. Apology encouraged forgiveness but only in the presence of the objective judgment as implemented in the forgiveness condition.

The second set of outcomes focused on whether the apology and forgiveness treatments encouraged the restoration of trust as compared to the reputation condition. The Tukey post hoc tests revealed that after the offender’s defection in round 5, participants in the apology condition gave reputation scores to the offender that were not significantly higher than the scores given by participants in the reputation condition. However, the trend was in the predicted direction \((p=0.1)\). Conversely, participants in the forgiveness condition assigned higher reputation scores to the offender as compared to participants in the reputation condition \((p=0.03)\). In round 6, once the offender had shared the fair half of the gains, participants of the apology \((p=0.003)\) and forgiveness \((p=0.002)\) treatments assigned significantly higher reputation scores to the offender as compared to participants of the reputation condition. Therefore, Hypothesis 2a was confirmed for the forgiveness condition, but only partly supported in the apology condition; participants in the apology condition did not assign higher reputation scores immediately after the offence (round 5) but only after the offender’s cooperation in round 6. Finally, in round 6, participants in the apology \((p=0.001)\) and forgiveness conditions \((p=0.001)\) made the initial trust move by transferring the 50 points more frequently than participants in the reputation condition. This result corroborated Hypothesis 2b. Fig. 2 displays the post hoc comparisons on the measures of cooperation, reputation and forgiveness.

Hypothesis 3 expected that victims who report more forgiveness for the offender will also report reduced negative emotions towards him/her. To investigate this hypothesis, two linear regressions were performed. If the model of forgiveness, positing the reduction of negative motivations and presumably the reduction of negative emotions, fits the data of this study, higher negative emotions after round 5 and 6 should be predicted by higher self-reported forgiveness scores (which indicate higher revenge and avoidance, thus increased un-forgiveness). In the first regression, the negative emotion measure as collected after round 5 was the dependent variable and forgiveness reported after round 5 was the independent variable. The regression model was significant \((F(1,58)=23.56, p=0.001)\). In the second regression, the negative emotion measure after round 6 was the dependent variable and forgiveness reported after round 5 was the independent variable. Once again, the regression model was significant \((F(1,58)=35.59, p=0.001)\). These results were in support of Hypothesis 3. Table 3 reports the results from the regression.

The post-experiment questionnaire revealed that all participants were naïve to the manipulations made in this study. Additionally, none of the participants suspected playing with a simulated partner.

6. Discussion

6.1. Results summary and implications

In previous analysis of an eBay panel of users, it was found that buyers, who viewed a seller’s negative reputation rating, in the absence of other information, withheld giving a positive reputation rating even though their transaction with the seller had been carried out successfully (Khopkar et al., 2005). Therefore, when viewing the seller’s previous negative reputation rating, buyers made negative inferences about the seller’s intentions. For example, whereas a delay of a promised product is usually blamed on slow postal services, a low rating seems to suggest that the seller is responsible. The present findings extend Khopkar et al.’s conclusions by revealing a similar phenomenon regarding the impact of reputation feedback on the victim’s (as opposed to a new partner’s) judgment. Participants, who had viewed the offender’s reputation score only, demonstrated lower trusting attitudes towards
the offender compared to those who also received the two reparative treatments (apology and forgiveness). The offender’s displayed reputation score before the offence was the highest possible, thus signalling the offender’s high trustworthiness. However, an unblemished reputation was insufficient to convey the unintentional nature of the offender’s present action. Lack of information about the offender’s intention appears to have led to negative attributions: though the offender was unintentional and previously benevolent, the victim was more likely to perceive the offender’s actions as being intentional, thus leading him/her to rate the interaction with lower reputation feedback. Due to this initial negative impression, the victim’s trust levels (as measured by the reputation variable) remained low even though the offender acted in good will during a new transaction.

Some participants in the reputation treatment also chose to withdraw from round 6, although the monetary risk attached to this trust move was only 25 pence. At a collective level, we found that participants’ negative emotions as reported for the offender after round 5 were correlated with un-forgiveness, i.e. revenge and avoidance. Hence, it appears plausible that alongside the negative attributions initially working to reduce trust, increased negative emotions towards the offender escalated the victim’s responses to actions of revenge and retaliation (as evident by the low-risk move to withdraw). Given that the offender had been rated only four times before the defection, this interpretation should be taken with some caution. The reputation system was sufficient for motivating 60 out of 63 first movers to make the initial trust move. Nonetheless, when trust breaks down, a reputation score acquired over more rounds may yield different results. Even so, our findings taken together confirm that a victim’s trust in an unintentional offender cannot be restored with a reputation system that operates in isolation.

Indeed, other emotional mechanisms work to motivate the victim’s forgiveness, for instance the offender’s apology. Therefore, we contrasted the reputation system to a system that also displayed the offender’s apology to the victim. We anticipated that the latter would elicit higher degrees of forgiveness and an increase in trust. Our results only partly confirmed this prediction. After viewing the offender’s apology, participants did not immediately report increased forgiveness. Similarly, their trust in the offender was not restored, as was evident by the low reputation feedback they gave. However, when provided with the opportunity to interact once again with the offender, his/her prior apology restored victims’ trust. Participants always choose the initial trust move. Additionally, the offender’s apology combined with his/her reparative move, led participants to choose higher reputation scores, as compared to participants who had not viewed the apology. Given the apparent restoration of trust in round 6, we postulate that participants also forgave the offender; retrospectively, the forgiveness measure should have been also included after round 6.

There are several possibilities as to why the offender’s apology did not repair the offence immediately after it occurred. Usually, the impact of apology on forgiveness is

---

Table 3
Regression results

<table>
<thead>
<tr>
<th></th>
<th>Negative emotions after round 5</th>
<th>Negative emotions after round 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>S.E.</td>
</tr>
<tr>
<td>Constant</td>
<td>1.428</td>
<td>0.887</td>
</tr>
<tr>
<td>Forgiveness</td>
<td>0.177</td>
<td>0.036</td>
</tr>
</tbody>
</table>

---

Fig. 2. Summary of post hoc comparisons.
investigated in pre-existing relationships between the victim and the offender (e.g. romantic relationships: Boon and Sulsky, 1997; McCullough et al., 1997, 1998). In those cases, prior knowledge about the offender’s motivations, intent and benevolence may be sufficient to motivate the victim to forgive the offender at the onset. It is possible that one-off interactions taking place between strangers are generally not forgiven unless the offender proves his/her intent with an action. This point maintains what Exline et al. (2003) argue: forgiveness does not always equate to an immediate restoration of trust as it is sometimes postulated (e.g. Axelrod, 1984).

Another plausible explanation lies in the inherent design of online systems which sometimes impose an articulated form of communication (Parkinson, in press). In face-to-face interactions, offenders, who have regretted their actions, will often experience embarrassment, guilt and shame. As such, verbal apologies can contain involuntary non-verbal expressions (e.g. blush, shrunken posture; Keltner and Buswell, 1997) that taken together confirm or disprove the offender’s sincerity. In online communication, our results summon up a different possibility. In the absence of non-verbal signals that confirm the genuineness of an apology, words may be regarded as contrived and of strategic nature. Therefore, though compared to reputation feedback an apology contains overt information about the offender’s intention words may still remain an untrustworthy cue until they are confirmed by actions.

The above finding suggests that eBay’s “mutual feedback withdrawal” can be effective given that it prompts both parties to sustain the interaction; the offender is able to explain, and whenever possible to correct, the offensive action, a combination of events that successfully restores the victim’s trust. However, the “mutual feedback withdrawal” approach operates retrospectively to the breakdown and as such it presents two important limitations. Firstly, it does not shield the victim from the initial perception of being deceived. Because of this, the victim may experience increased anger towards the offender. This study showed that the victim’s reported emotions of anger and irritation were correlated with un-forgiveness. Importantly, victims’ low levels of forgiveness after round 5 kept negative emotions at high levels after round 6 even though the offender had cooperated. Thus, systems whose designs do not enable forgiveness immediately, expose the victimised member to negative emotions that may also obstruct the prosocial outcomes of forgiveness, e.g. willingness to cooperate and to resolve the breakdown. Secondly, a retrospective reparative system does not protect the offender from the unjust low rating, a form of punishment that is sufficient to drive the offender to abandon the online community (Khokpar et al., 2005).

To address this problem, a third treatment was investigated in which an online forgiveness intervention was made by the system. In contrast to the findings from the apology treatment, those in the forgiveness treatment reported more forgiveness immediately and also demonstrated higher trusting attitudes both after the offence in round 5 and after the subsequent interaction in round 6. We believe that the feedback provided by the intelligent “forgiveness” component, worked to alleviate participants’ negative attributions towards the offender. As a consequence, the victim became more forgiving and trusting directly after the offence.

This third set of results, at a first stage, invites designers of social systems to consider how to support forgiveness implicitly. For example, a reputation system should be designed to emphasise a member’s benevolence to others as well as previous offences. At a second stage, as reputation systems evolve, eventually an explicit implementation of the “forgiveness” component might become feasible. eBay has recently added a “Detailed Seller Ratings” feature in which buyers can be rated on five different dimensions such as their communication ability or timeliness in shipping. A “forgiveness” system component might intervene when a buyer is about to give a low rating to a seller who has been previously rated highly on that measure. Similarly, if the offender’s trustworthiness is judged as high, the system may recommend that a reasonable grace period be granted: when submitting negative feedback against the offender during this time window, members can be advised to delay their submission. However, as with reputation systems, this approach opens up possibilities for abuse and inevitably social system designers would have to consider the possible dangers of misuse (Vasalou et al., 2006b discuss these implications).

6.2. Limitations

The experimental paradigm chosen in this study was the trust game played in the lab. While there are arguments opposing the artificial environment of controlled experiments (e.g. Boehner et al., 2007), in our view, the trust game was an appropriate methodology for testing our hypotheses. One of the systems pursued was the novel “forgiveness” component, which does not yet exist in the real world. Additionally, controlled experiments can reveal more information about users’ motivations. Most naturalistic studies on reputation systems are often confined to an analysis of behavioural patterns while users’ motivations are usually a topic of speculation (Resnick et al., 2006).

Despite its benefits, there are several limitations that derive from the context-free format of this methodology which we must acknowledge when generalising these results. The trust game in its pure form does not allow the researcher to incorporate aspects such as the ability of the trusted party which is certainly relevant in most online transactions (Riegliesberger et al., 2003). Therefore, similar to previous studies (e.g. Keser, 2003) this experiment modelled cognitive trust by frequency and severity/utility of previous actions. Moreover, victims’ loss of trust was manipulated by introducing a monetary loss. There are two points of concern with this type of risk modelling. Firstly,
Riegelsberger et al. (2003) have voiced concerns about experimentally manipulating loss: though participants lost money as a consequence of the trust violation, this loss was a promised gain (compensation for participation) that was not in their possession previously. Secondly, a trust violation in real life may involve financial loss but also other emotional penalties imposed by the social nature of the interaction (Riegelsberger et al., 2003). Because of these reasons, participants’ engagement in the lab may have been lower than the equivalent real world offence involving a significant personal loss.

We also note a possible concern with the procedure of this experiment. Participants played the game in the presence of others co-located in the room, a format that may have introduced further psychological processes. For example, norms consistent to one’s social identity (Postmes et al., 2001) may have been evoked, thus motivating participants to work more collaboratively. To control for this possibility, steps were taken to physically separate participants from each other, while during the experiment silence was maintained. We also believe that participants’ high level of expertise with web applications and reputation systems limits this possible confound, i.e. previous experience ensured that participants were engaged in the interaction rather than focusing on the presence of other participants.

Finally, trust was indirectly measured by participants’ willingness to further interact with the offender and the reputation score assigned to the offender. These choices reflect aspects that are particularly pertinent to the design of e-commerce systems, a point to be considered when generalising these findings to other kinds of one-off interactions. Future extensions of this work focusing on other online environments may choose to measure different facets of trust. For instance, when researching trust during mediated communication is not always an act of intentionality. An offence can be accidental and performed by a member who has been benevolent in previous interactions. We found that a reputation system does not alleviate such online offences. An apology from the offender does not repair the trust breakdown directly after the offence. However, if the offender is given the opportunity to repair the action, the victim’s trust can be restored. An intelligent “forgiveness” component, built into the apology forum, can relate to the victim an objective positive judgment about the offender’s trustworthiness. This method restores the victim’s trust and increases forgiveness directly after the offence, irrespective of the offender’s motion to reverse the offense with a counter action.

6.3. Future work

Throughout this article we distinguished between the cognitive and emotional aspects that promote forgiveness. We argued and showed that cognitive aspects motivating forgiveness, e.g. as made evident via high-quality reputation feedback, were not sufficient to restore the victim’s violated trust. This two-faceted distinction was used to illustrate how effectively forgiveness was supported in the three systems under investigation. However, explicit measures for these two dimensions were not included. At the same time, our scope was limited to investigating a particular theoretical view that posits how forgiveness impacts on the victim’s levels of trust and emotions. We did not consider the impact of forgiveness on the full spectrum of the trustee’s perceived trust dimensions as delineated in trust models. An enquiry revealing which particular facets of trust (e.g. credibility vs. predictability; cognitive vs. affective) are damaged during an unintentional trust breakdown may suggest further avenues for designing reparative systems.

In a final remark, we note that this work is restricted to dyadic online interactions. Nonetheless, it is important to recognise the key role of apologies on the collective as a single offender can apologise to a body of users.5 To give an example, Joinson and Dietz-Uhler (2002) discussed the community’s reaction following the admission and apology of a member who had participated under a fabricated and thus deceptive identity. His apology stimulated an emotional dialogue, with many members arguing what norms should prevail within their community; some forgave the offender because of his previous benevolence, while others felt personally betrayed and condemned his actions. Hence, collective apologies can contribute to an awareness and agreement on social norms. However, public admission can also irreparably damage the core trust that has developed within the community over time. Future work might focus more closely on the benefits and drawbacks of collective apologies.

7. Conclusion

This article argued that offensive behaviour in computer-mediated communication is not always an act of intentionality. An offence can be accidental and performed by a member who has been benevolent in previous interactions. We found that a reputation system does not alleviate such online offences. An apology from the offender does not repair the trust breakdown directly after the offence. However, if the offender is given the opportunity to repair the action, the victim’s trust can be restored. An intelligent “forgiveness” component, built into the apology forum, can relate to the victim an objective positive judgment about the offender’s trustworthiness. This method restores the victim’s trust and increases forgiveness directly after the offence, irrespective of the offender’s motion to reverse the offense with a counter action.

Acknowledgments

Jens Riegelsberger (Google), Toby Regner (University of Jena), Etienne Roesch (University of Geneva), Lidia Oshlyansky (Swansea University) and two anonymous reviewers are acknowledged for their comments. We thank Amjad Hanif (eBay) for his advice on the eBay system. We would also like to thank Arvind Bhusate, Tom Carlson and Daniel Ramirez who provided assistance during the experiment. This research was funded by the HUMAINE IST Network of Excellence.

In a recent HUMAINE Workshop (2005), Peter Goldie (University of Manchester) argued that, when feasible, social system designers should encourage offenders to apologise to the community as this method brings awareness to the community’s norms. He paralleled the notion of collective apology to a form of “public” apology, given in concession of a public officer’s wrongdoing.
Appendix A

You will be playing an online game. Before starting the game, you are asked to carefully read through the next few screens to understand the rules and regulations of the game. In this game there are two players who each begin with a designated number of points. One player will start off with 150 points and the other with 100 points. There are two moves in the game listed below:

1. The first move: The player who has 150 points (first mover) has the option to transfer 50 points to the other player (second mover) or to keep the full amount of 150 points and cash in. If the first mover decides not to transfer the 50 points, the round of the game is complete. Both players leave the game with the amount they started: the first mover leaves the game with 150 points and the second mover leaves with 100 points.

2. The second move: If the first mover transfers 50 points to the second mover, this amount is automatically multiplied by 6 so that the points the second mover receives are 300. The second mover now has his initial 100 points plus 300 points thanks to the first mover's transfer. The second mover has the choice to transfer either 0 (nothing) or 150 (half) points back to the first mover. This amount will be added to the 100 points that are left in the first mover's wallet.

What to expect during the game:

- You will be randomly selected to be either first or second mover. Once your role is decided you are assigned to it for the rest of the game.
- You may randomly play up to 10 rounds of the game. You will know whether you proceed to the next round at the end of each round.
- You will be randomly assigned to play with a new person in every round. Although this happens rarely, you may happen to play with the same player twice.
- The points in the game are real money. The exchange rate is 50p for 100 points. You will receive this money at the end of the experiment at the discretion of the facilitator. Other participants of the game will not be informed on the amount you made while playing the game.
- After each round you may get some emotion questions to answer. These will be given to you at random rounds.
- You will be notified onscreen when the game ends.

Appendix B

B.1. Transgression-related interpersonal motivations inventory questionnaire (TRIM)

1. I will make him/her pay (revenge).
2. I wish that something bad would happen to him/her (revenge).
3. I want him/her to get what he/she deserves (revenge).
4. I'm going to get even (revenge).
5. I want to see him/her hurt and miserable (revenge).
6. I keep as much distance between us as possible (avoidance).
7. I live as if he/she doesn't exist, isn't around (avoidance).
8. I don't trust him/her (avoidance).
9. I find it difficult to act warmly toward him/her (avoidance).
10. I avoid him/her (avoidance).
11. I cut off the relationship (avoidance).
12. I withdraw from him/her (avoidance).

B.2. Modified TRIM questionnaire for one-off online interactions

1. If I get the opportunity in the game, I'll make him/her pay.
2. I wish that something bad will happen to him/her in the game.
3. I want him/her to get what he/she deserves in the game.
4. If I get the opportunity in the game, I’m going to get even.
5. I would like to see him/her hurt and miserable in the game.
6. If I were to play with him/her again, I would keep as much distance between us as possible.
7. If I play with her/him again, I will pretend like he/she doesn’t exist, isn’t around.
8. I don’t trust him/her in the game.
9. If I play with her/him again, I will find it difficult to act warmly toward him/her.
10. If I play with her/him again, I will avoid him/her.
11. If I play with her/him again, I will cut off the relationship.
12. If I play with her/him again, I will withdraw from him/her.

B.3. Final TRIM questionnaire for one-off online interactions

1. I wish that something bad will happen to him/her in the game.
2. I want him/her to get what he/she deserves in the game.
3. If I get the opportunity in the game, I’m going to get even.
4. I would like to see him/her hurt and miserable in the game.
5. If I play with her/him again, I will pretend like he/she doesn’t exist, isn’t around.
6. If I play with her/him again, I will avoid him/her.
7. If I play with her/him again, I will cut off the relationship.
8. If I play with her/him again, I will withdraw from him/her.
References

Parkinson, B. Emotions in direct and remote social interaction: getting through the spaces between us. Computers in Human Behavior, in press.


