From Theory to Practice: Forgiveness as a Mechanism to Repair Conflicts in CMC

Asimina Vasalou, Jeremy Pitt, and Guillaume Piolle

Intelligent Systems & Networks Group, Dept. of Electrical & Electronic Engineering, Imperial College London, London, SW7 2BT, UK

Abstract. In computer-mediated communication (CMC) online members often behave in undesirable ways, therefore creating a need for an active regulating force. Trust and reputation mechanisms have been adopted to address this problem and in doing so have eliminated the high costs of employing a human moderator. However, these systems have emphasized the need to 'punish' a given offender, while neglecting to account for alternative ways to repair the offence e.g. by forgiveness. In this paper, we define a theoretical model of forgiveness which is operationalized using a fuzzy logic inference system and then applied in a particular scenario. It is argued that forgiveness in CMC may work as a possible prosocial mechanism, which in the short-term can help resolve a given conflict and in the long-term can add to an increasingly prosocial and homeostatic environment.

1 Introduction

In human societies, when violating a norm, the offender is usually 'punished' both emotionally (e.g. experiencing embarrassment) and practically (e.g. by prosecution). The threat of these two sanctions is persistently evoked by physical markers, (e.g. people watching, the presence of law enforcement officials) and works preventively so that a sense of general social order is maintained within the community. Online societies differ from physical societies, in how both the emotional and practical implications are perceived. To begin with, anonymity and the absence of a physical self weaken the impact of the emotional consequences (e.g. shame or embarrassment) that an offender experiences as a result of his/her offence. To add to this, the presence of an active policing force is not visible until the member's behaviour has reached what is considered to be illegal according to law. Therefore, one of the problems identified through these two points is the need for an intermediate mechanism that will signal the offender early on and that will also inform the community about milder offences where punitive legal action against the offender is perhaps inappropriate.

Trust and reputation mechanisms have been widely adopted in addressing this issue [15]. These mechanisms have empowered members of online communities by allowing them to appraise and capture the granularity of their fellow members' actions (e.g. through ratings). However, in doing so, the designers of those

K. Stølen et al. (Eds.): iTrust 2006, LNCS 3986, pp. 397–411, 2006.

systems have placed emphasis and value on the quantitative appraisal that usually follows an offence while neglecting to account for the qualitative appraisal that often makes repair between two members possible [19]. In human-human interactions, a violation of norms is unavoidable but not necessarily unforgivable.

In this article, we address this issue by proposing forgiveness as a repair mechanism that is instantiated during a given conflict, possibly facilitating a resolution between online members. In previous work, we have described the conceptual framework of the forgiveness proposal [19]. We now reify this proposal by developing a stand-alone operational model of forgiveness that is straightforward to automate, and can be integrated into any platform or configured for any application domain.

This article is divided into 5 main sections. Section 2 gives an overview of the motivations for considering forgiveness in CMC. Section 3 presents a theoretical model of human forgiveness collectively investigated in the field of psychology. In Section 4, we describe the forgiveness model implemented as a fuzzy inference system driven by the theory described in Section 3. Section 5 integrates the model into a collaborative distance learning scenario. Finally, this paper concludes in Section 6 with a summary and a discussion of further work.

2 Motivation

There are strong incentives for considering forgiveness as a possible reparative mechanism in online communities. For example, issuing forgiveness is known to stimulate the offender into voluntary actions of repair [7]. Moreover, punishing the offender for an action they did not intentionally perform (e.g. bad ratings for accidentally delivering the wrong product) often results in emotions of anger and low-compliancy behaviors [7]. This could possibly motivate a member to withdraw from the online community due to the unjust treatment. Even more, one's judgment can be sometimes misguided and construed on false information. In this situation, a system that supports irreversible judgments is both unfair and unethical. Finally, although forgiveness does not necessarily mean that trust is automatically regained [3], it often provides closure, which may alleviate the aggression created from a disrupted interaction. This point is further demonstrated by the physical well-being of those who tend to issue forgiveness more frequently [21].

To summarize, forgiveness promises short-term benefits in CMC such as giving the offender an outlet through which to apologize or pacifying the victim of the offence, so that both victim-offender can resolve the conflict. At the same time, the short term benefits ultimately have the potential to increase the overall equilibrium of the online community.

3 Theoretical Model

Forgiveness results from a number of prosocial motivational changes which reverse one's initial desire to adopt negative strategies towards the offender

(i.e. revenge, avoidance). In this sense, forgiveness replaces malevolent motivations towards the offender with constructive and positive behaviors which work to reverse the initial censure [10]. The forgiveness process, as described in psychology, is further depicted in Figure 1, where the offender, member x, violates a rule with action A. Following victim y's negative predisposition towards offending action A, four positive motivations collectively add up to possibly formulate forgiveness. The positive motivations we consider are empathy, actions of repair, the beneficial historical relationship of victim-offender and an appraisal-judgment of the offence.

The definition used here employs a degree of freedom in long-term relationships as the victim may forgive a single offence without explicitly reversing their attitude as a whole [10]. Likewise, while a certain violation may be forgiven, other past behaviors may still impede one's trust towards another. Despite popular definitions of forgiveness forgetting, condoning, trusting or removing accountability are not necessarily considered to be a part of forgiveness [3].

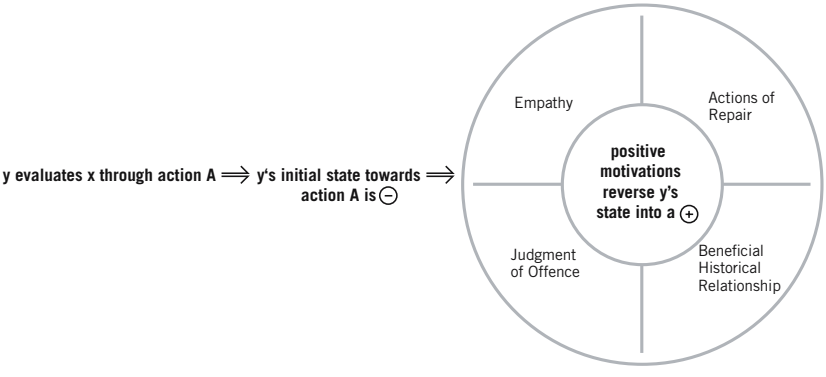

Fig. 1. A motivation-driven conceptualization of forgiveness where positive motivations add up to increase forgiveness

On the basis of the forgiveness definition given here, we propose the following:

> *Premise 1* – x violates rule A. Initially y, the observer/victim of x's offence is inclined negatively towards x. y assesses all the factors surrounding x's action-violation A and decides to issue forgiveness by applying a series of $(+)$ positive motivations to his initial $(-)$ negative state.

Next, we discuss the four central positive motivations of the theoretical forgiveness model which are the judgment of offence, actions of repair, beneficial historical relationship and empathy. The four positive motivations are described by eleven constituent parts which are the offence severity, offence frequency, intent, apology, reparative actions, prior interactions, utility of benefits, and frequency of benefits, visible acknowledgment, similarity and propensity to embarrassment.

Judgment of Offence. Observers/victims of one's offence make attributions by accounting for a number of factors surrounding the offence. First, the severity of the current act is assessed. More severe violations lead to harsher judgments [1, 2]. Furthermore, a historical trail of one's past behaviors is compared against the current violation. Together, frequency and severity of past acts impact one's inclination to forgive [2]. Additionally, apparent intent leads towards more negative attributions with low intent actions supporting more positive attributions [1, 8]. Given this, we state:

> *Premise 2* – y assesses x's action by (severity) AND (frequency/severity of x's historical actions) AND (x's intent)

Actions of Repair. A truthful apology or a good deed [2] that reverses the offence can pacify the observer or victim of the offence. In fact, apology and restitution, together constitute a strong partnership facilitating and even predicting forgiveness [20]. However, reversing one's violation with a reparative action brings up an important issue. Inevitably the weight of a good deed against a severe and frequently performed violation will have to be formulated or accounted for. As a result, we state:

> *Premise 3* – y issues forgiveness if x offers (an apology) AND (reparative action $B \geq$ action A)

Beneficial Historical Relationship. Prior familiarity and a relationship of commitment with the offender positively predispose the victim and increase the likelihood of forgiveness [11]. Good friends or successful business partners rely on a longer, richer and mutually-rewarding history fostering a propensity towards forgiveness. Therefore:

> *Premise 4* – y will issue forgiveness if (the utility of x's actions has been high) AND (x has been frequently beneficial to y)

Empathy. Empathy, one's emotional response towards another's affect [5] is regarded as a mediator, appeasing the victim and facilitating forgiveness. Empathy is evoked by offender's apologies among others, is a predictor of forgiveness and its intensity has been found to positively correlate to the extent of forgiveness the victim issues for the offender [12]. Empathy also manifests in embarrassment to form 'empathic embarrassment', a milder form of embarrassment 'incurred' by imagining oneself in another's place. Empathic embarrassment has four determinants. First, the salience of the offender's embarrassment controls the degree of felt empathic embarrassment. Visibly embarrassed offenders elicit more empathic embarrassment from others. Second, the emotion intensifies when the victim is somewhat familiar to the offender. Third, we foster stronger feelings of empathy towards those who are most similar to us in terms of personality or characteristics (e.g. a colleague or a cultural compatriot). Similarly, one will be more empathic towards an offender with whom s/he shares a similar history of offences. Finally, the observer's propensity to embarrassment determines to a great degree the empathic embarrassment s/he may experience. A highly

'embarrasable' observer will experience increased empathic embarrassment (see [9, 13] for a detailed account. On the basis of the previous discussion we propose the following:

> *Premise 5* – The extent of y's forgiveness will vary by the (degrees of empathy/empathic embarrassment y feels for x) which increases IF (x's embarrassment is visibly intense) AND (if y has some prior familiarity with x) AND (if y shares similar characteristics with x) AND (if y's propensity to embarrassment is high) AND (if x apologises for the offence)

This completes the theoretical basis for forgiveness as formulated in the field of psychology and specified in [10]. We have extracted five premises encapsulating the overall forgiveness decision and identified the four motivations for forgiveness, composed of eleven constituent parts. The objective of Section 4 is to propose a generic computational model built on the basis of this theory which can be then implemented and adapted into any domain.

4 Computational Model of Forgiveness

In this section, we develop a computational model that reifies the theoretical basis of forgiveness and is built using fuzzy inference systems (FIS). We first justify our reasoning for using FIS and outline FIS. We then describe the implementation of the decision maker. Finally, we give examples of the fuzzy rules which are used by the decision maker to make its inference.

4.1 Fuzzy Inference Systems as the Operational Basis

The theoretical work we have discussed so far, with the exception of a study conducted by Boon and Sulsky [1], has isolated and then measured the constituents (e.g. intent) of each motivation (e.g. judgment of offence) separately. Boon and Sulsky's study clearly demonstrates the independent rater 'disagreement' on how the different constituents weigh on the decision to forgive. Therefore, in operationalizing the theory, there is a need to define a more concrete model that describes the ranges, weights and interactions of all four motivations and their eleven constituent parts.

To address this issue we implemented the forgiveness decision maker by using the Takagi-Sugeno fuzzy inference system (FIS) [17], as fuzzy logic satisfied these three important aspects. (1) Ranges: FIS allowed for each motivation constituent to be stored in ranges, from high to low, which was particularly important as for example, an offence 1 may be considered 80% severe whereas an offence 2 is regarded as 20% severe. (2) Weights: The violation appraisal captured by the judgment of the offence motivation is the most powerful motivation of forgiveness. FIS allowed us to attribute more weight to the judgment of the offence over the remaining three motivations of actions of repair, beneficial historical interactions and empathy. (3) Interactions: The decision maker closely followed

the structure of the five premises so that each motivation was separately computed on the basis of its own constituents and then passed onto the final decision maker.

FIS Overview. Fuzzy Logic, as developed by Takagi-Sugeno [17] is a formalism which facilitates reasoning about imprecise facts, uncertainties, and value judgments – in other words, all the human factors that might inform a forgiveness decision. Fuzzy Logic is the basis of fuzzy inference systems, although there are different types of fuzzy systems as there are various different ways in which outputs can be determined.

In general, to build a fuzzy system, an engineer might start with a set of application-dependent fuzzy rules as specified by a domain expert. In our case, the fuzzy rules for the operational model are derived from the theoretical model described in Section 3. Fuzzy rules are expressed in the form "*if* ... *then* ..." that convert inputs to outputs, where both inputs and outputs are members of fuzzy sets (a fuzzy set is a set in which objects are members to some degree). So, for example, we might have a rule of the form:

> *If* the offender apologizes *and* does not repair the offence
> *then* the forgiveness value is increased by 10%

Similarly:

> *If* the offender apologizes *and* repairs the offence
> *then* the forgiveness value is increased by 30%

Given a set of such rules, it may be that a particular range of inputs fire (activate) any given subset of those rules. The rules which are fired then contribute proportionally to the fuzzy output: this is calculated by applying the implication method of fuzzy logic to the activated rules and aggregating all the results. The process of defuzzification converts the aggregated output into a 'crisp' value (the usual method is a centroid calculation, i.e. finding the centre of an area under a curve).

This entire process, called fuzzy inference, thus converts quantitative inputs into a precise output using qualitative statements: in our case, this precise output is a yes-no decision on whether to forgive or not.

FIS of Forgiveness. The FIS decision maker that we implemented (see Fig. 2) receives numerical values of the eleven constituent motivations as its input in order to make a yes-no forgiveness decision (d) as its output.

The forgiveness decision maker goes into effect only when an offence occurs i.e. a user has violated a norm. At that time, the eleven constituent motivation signals of forgiveness are computed. They are then input into FIS2 through FIS5. The outputs of FIS2 through FIS5 represent the weights of the four forgiveness motivations which are input to FIS1 to compute a final output value d. FIS1 is the operationalization of premise 1 in Section 3 and the final value d constitutes the forgiveness recommendation (*if* $d > 0.5$, *then* forgiveness = true). We note that the weight of the judgment of offence motivation on the overall forgiveness decision (d) is 0.5 while actions of repair, beneficial historical relationship and empathy are each weighted 0.166.

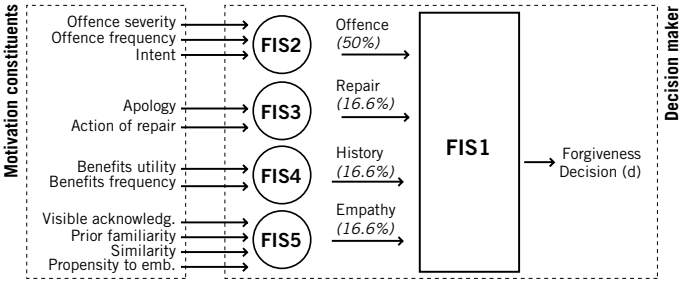

Fig. 2. The Forgiveness Decision Maker

4.2 Examples of Fuzzy Rules of Forgiveness

The fuzzy inference systems FIS1-5 are based on a set of rules that follow the structure, form and theory of premises 1 through 5 of Section 3. The full set of rules can be found at [18]. Here, we give representative examples of two fuzzy sets, the fuzzy set for judgment of offence and the fuzzy set for the overall forgiveness decision.

The judgment of offence as expressed in premise 2 is reliant on the three constituents of severity, frequency and intent. As each increases so does the probability of a low forgiveness rating. An example of two rules follows that demonstrates this difference in granularity:

> *If* severity is low *and* frequency is low *and* intent is high
> *then* the judgment of offence motivation is 0.4

In contrast:

> *If* severity is low *and* frequency is high *and* intent is high
> *then* the judgment of offence motivation is 0.2

Following the calculation of each individual motivation, its value is input in the fifth FIS and a crisp value of forgiveness is computed on the basis of its own rules. For example:

> *If* judgment of offence is high *and* actions of repair is high
> *and* beneficial historical interactions is low *and* empathy is high
> *then* forgiveness is 0.83

5 Application Domain and Integration of the Model

In the previous section, our aim was to create a generic forgiveness model whose integration and input values are ultimately determined by the domain it is fit into. The objective of this section is to integrate the forgiveness model into a specific CMC scenario.

Arguably, the mechanism could be integrated into an e-commerce platform such as eBay where a seller may be unjustly rated due to unforeseen factors

(e.g. slow post resulting in late delivery). Forgiveness could also be appropriate in an e-health scenario in which the discussion of sensitive topics may often lead to misinterpretations. Here we demonstrate how to fit the forgiveness mechanism into an e-learning scenario. The computational model of forgiveness developed in this paper has many quantitative as well as qualitative components. For example, a user can reverse the offence with a quantitative action which may be most appropriate in e-commerce or apologize with a qualitative statement, relevant to an e-health forum. Collaborative distance learning relies on transactions (e.g. assignments) but it also has a social capacity, i.e. students may use the tools available to communicate before transacting. Therefore, the transactional and social elements of e-learning permit us to test both the quantitative and qualitative aspects of the forgiveness model.

In this section, we first outline a collaborative distance learning domain into which we customized the forgiveness model. Next, we describe the collection and then the computation of the eleven constituent parts of the four motivations which as illustrated in the previous section, are used by the FIS as inputs (see Figure 2). Finally, we demonstrate how the constituent motivations and decision maker integrate into a complete comprehensive architecture.

5.1 Overview

The forgiveness tool is integrated in a collaborative distance-learning environment. The workflow of this environment supports two-party interactions at a given time, where team tasks are broken down into segments and executed sequentially. The term 'collaboration' in this domain, constitutes the successful delivery of an assignment and can be contingent on a number of factors such as timeliness, good communication skills, quality of work etc.

When signing up to participate in the community, a member is requested to fill out a short survey reporting two successful and two unsuccessful past teamwork experiences. The first signifies a benefit gained as a result of the team collaboration, while the second represents an offence executed during the collaboration. These reports are in turn processed by a human moderator, who checks them in terms of quality (e.g. grammar, clarity of articulation). The moderator then posts the reports online so that the distance-learning community can collectively rate them. The final output of this process is a list of successful and unsuccessful collaboration incidents, each of which has a corresponding 'utility' or 'severity' rating derived from the mean of all ratings. These ratings represent an objective measure of severity or utility. The collaboration reports and their corresponding ratings are updated annually when new users sign up. This way the knowledge base is constantly updated.

Upon collaborating with another member, a user selects the benefit or offence which most closely characterizes his/her experience from the knowledge base. This report is stored, and over time builds up a member's history. In the event of an offence report, two sequential events happen. At first, the offender is offered reparative tools e.g. enabling him/her to reverse or apologize for the action-violation. The intelligent component executed with fuzzy inference

systems is instantiated and assesses whether the particular offender should be granted forgiveness. The victim of the offence is then informed of this decision and is presented via the interface with relevant-to-the offence information (e.g. the offender's past history and all the other factors used by the FIS to compute value d). The act of forgiveness is ultimately the user's decision as his/her personal judgment may differ from the one inferred by the FIS.

5.2 Collection and Computation of the Constituent Motivation Signals

Earlier, eleven forgiveness constituent motivations were mentioned, each of which impact on one's decision to forgive. In face-to-face interactions these constituents may be collected by memory, perception or interaction. For example, the offender's visible acknowledgment is immediately perceived through his/her face. Similarly, the offender has immediacy of contact, therefore making it possible to apologize for an offence. In this forgiveness application, we constructed new ways for collecting this kind of data. We now detail the computation and/or collection method for each motivation individually.

j_0: The **severity** of an offence is a value that is assigned to each type of offence automatically and is measured from 0 to 0.5. Rating values higher than 0.5 are classified as beneficial collaborations. As described in the previous section, the severity value is the mean of ratings for each offence as given by users of the community upon signing up.

j_1: The **frequency** of a particular offence is computed by:

$$j_1 = \frac{\left(\frac{n_{offencekind}}{n_{offences}} + \frac{n_{offences}}{n_{collaborations}}\right)}{2} \tag{1}$$

where $n_{offencekind}$ denotes the number of the offender's offences of the current kind, $n_{offences}$ is the offender's total number of offences across time and $n_{collaborations}$ is the offender's total collaborations within the community. Two aspects of frequency are encapsulated in this formula: the frequency of the current offence is computed with the first division and the frequency of the offender's total past offences is computed with the second division. Among other possibilities, this equation intends to capture the instances where a user has infrequently violated a particular norm but at the same time frequently violates many others.

j_2: A judgment on the offender's **intent** is reported both by the offender and the victim via a user interface which activates upon the offence. The report values range from 0 to 1 (i.e. [0, 1]). Each user's intent-report is given a different weight depending on his/her credibility which is computed on the basis of past offence frequency and severity. Specifically, a user's credibility C_u is:

$$C_u = \frac{\sum_{i=1}^{n} R_i}{n} \tag{2}$$

where n is the number of the total collaborations that the user has had within the community and R_i $(1 \leq i \leq n)$ is the rating of each collaboration. As previously

mentioned, ratings between 0 and 0.5 are considered offences, whereas ratings greater than 0.5 and less than 1 are categorized as benefits. The following formula then encapsulates intent:

$$j_2 = \frac{(I_o \times C_o) + (I_v \times C_v)}{2} \tag{3}$$

In this formula C_o denotes the offender's credibility, C_v is the victim's credibility, and I_o and I_v signify the offender's and victim's intent report rating.

j_3: **Apology** from an offender is reported via a user interface and is then offered to the victim of the offence. A_o is a binary value where 1 indicates that the offender has apologised to the victim and 0 indicates the absence of an apology offer. The credibility and honesty of the offender's report is then given a rating A_v by the victim, ranging from 0 to 1. This rating is weighed into the overall apology value. Similar to computing intent, the offender's and victim's credibility C_o and C_v are taken into account and weighted into the overall apology value. Therefore, apology is given by:

$$j_3 = \frac{(A_o \times C_o) + (A_v \times C_v)}{2} \tag{4}$$

j_4: The offender may offer a **reparative action** RA_o to the victim by either reversing the offence or by completing a new task. This process is facilitated by a user interface. The value for RA_o is binary. When the action of repair has been completed, the victim rates it with RA_v, ranging from 0 to 1. RA_v is then weighed into the total reparative action value. The offender's and victim's credibility is also computed into the final reparative action value. The formula for reparative action is:

$$j_4 = \frac{(RA_o \times C_o) + (RA_v \times C_v)}{2} \tag{5}$$

j_5: The **utility of benefits** is a value that is assigned to each type of benefit automatically and is more than 0.5 and less than 1. As described in Section 5.1, this value is the mean of ratings for each benefit as given by users of the community upon signing up.

j_6: The value of **benefits frequency** between two members is calculated by dividing the number of benefits $n_{benefits}$ that the victim has experienced while collaborating with the offender, by the total number of collaborations between the victim and the offender $n_{collaborations}$. As such, beneficial historical relationship is:

$$j_6 = \frac{n_{benefits}}{n_{collaborations}} \tag{6}$$

j_7: The offender's **visible acknowledgement** (e.g. the blush) value is controlled by the degrees of the offence frequency formula j_1. That is, if the offender has rarely performed the action in question, the visible acknowledgment value will be high and the victim of the offence will be signaled of the offender's emotional display.

j_8: **Prior familiarity** between two members is defined by the formula:

$$j_8 = min\left(1, \frac{|n_{collaborations}|_t}{f}\right) \tag{7}$$

where $n_{collaborations}$ denotes the number of collaborations between victim and offender in time interval t. Both t and f values are application specific. In our scenario, is set to the 3-month academic quarter during which the student-users will be using this system. We intuitively consider familiarity to be gained after collaborating for at least 3 times, so that f equals 3. This formula then tracks the number of total collaborations between the victim and offender during the 3-month time interval and considers familiarity to be achieved following three or more collaborations.

j_9: **Similarity** between two members is given by:

$$j_9 = match(\mathbf{d}_o, \mathbf{d}_v) \tag{8}$$

where $\mathbf{d}_v$ is the victim's set of all past forgiveness decisions (d), each containing the eleven constituent motivations $j_0 - j_{10}$. For each element of $\mathbf{d}_v$, the *match* function finds the closest set of constituent signals to those of the offender's in the set of $\mathbf{d}_o$. It then goes on to compare the final forgiveness decision (d) of those sets. Similarity is the sum of all identical decisions divided by the victim's total number of forgiveness decisions.

j_{10}: Finally, **propensity to embarrassment** is collected with a short self-report questionnaire [14] that all members fill out when first signing up. The propensity value is registered and stored in the input conversion layer hereafter.

5.3 System Architecture

The overall framework integrating both the eleven constituent motivations and the forgiveness decision maker is depicted in Figure 3. It consists of two main modules:

- An *input conversion layer* which stores and computes the values of the eleven signals $j_0 - j_{10}$.
- The *decision maker* that outputs the final forgiveness decision (d).

The input conversion layer of the system, stores a member's successful (beneficial) or unsuccessful (offensive) collaborations as two separate objects. Those objects are labeled the *Collaboration Report* object and the *Offence Appraisal* object respectively. Following a collaboration with another member, a user reports on his/her experience. If the experience was positive, then the user's report is stored in a Collaboration Report object. The Collaboration Report object captures the identity of the user, a timestamp, and a measure of the benefit of the collaboration. In contrast, if the collaboration experience was negative, the user's report is stored into an Offence Appraisal object. The Offence Appraisal object captures the type of offence, the identity of the offender, a timestamp, measure of the offence severity and parameter values which are used to compute some of

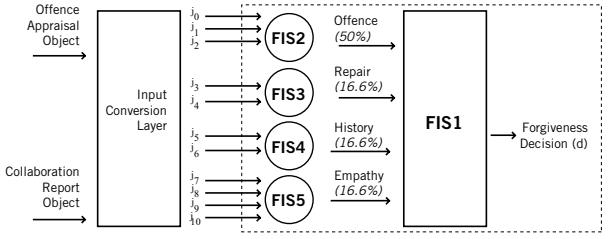

Fig. 3. System architecture integrating the eleven constituent motivations and the FIS decision makers

the constituent motivation values. These include the intent, apology and reparative actions offered by the offender. The embarrassment propensity constituent motivation is a constant value that has been stored in the input conversion layer. The motivations that rely on historical data, such as offence frequency, historical relationship, similarity and prior familiarity between the offender-victim and visible acknowledgement are computed separately in the input conversion layer to be later passed as signals to the decision maker. Upon completing the interaction, both the Collaboration Report and the Offence Appraisal objects are stored so that each user builds up a history over time.

6 Conclusions

This article presented forgiveness in light of the prosocial and healing benefits it brings to human societies. We proposed the inclusion of forgiveness online as a way to encourage prosocial behaviors both in the victim and offender. The motivation behind our work is the reparative nature of forgiveness in some cases, while the destructive consequences of its absence in others. We went on to discuss the formation of forgiveness by the collective 'accumulation' of four positive motivations. Resulting from this definition, we designed an operational model additively shaped by the motivations' interactions, implemented with fuzzy inference systems. In doing this, our guiding principle was to create a model that is straightforward to automate, and can be integrated into any platform e.g. multi-agent systems or configured for any application domain e.g. e-commerce. The fuzzy sets that FIS uses as a basis to make an inference are written in a natural processing language which is both comprehensive and replicable by a wider audience ranging from social scientists to computer scientists. Even more, fuzzy rules offer flexibility in changing the weights of the motivations to reflect any expert's judgment.

6.1 Raised Issues

The objective of this article was to bring forward the neglected but yet significant topic of forgiveness while at the same time creating an operational model that can be easily adapted in a number of domains. Although psychology

offers positive prospects for forgiveness applications, we cannot neglect the possible challenges we may face when integrating and evaluating such a model in a computer-mediated environment:

Vulnerability: Forgiveness may encourage harmful behaviors by withdrawing well-deserved punishment [6]. As in many applications, users may 'hijack' the system and find ways to manipulate it to their advantage. Therefore, a responsible and careful facilitation is vital.

Semantics: Human actors' expectations, perception and understanding of forgiveness often exceed the actual function of forgiveness as formally given in psychology [3]. For example, despite colloquial beliefs, forgetting or trusting is not part of forgiveness. We intend to address this point with the design of clear and communicative language during the forgiveness facilitation. It should be emphasised that forgiveness does not automatically repair trust. Even more, given the disparity between lay understanding and formal definitions of forgiveness, it is argued that the word 'forgiveness' should not be displayed directly during the users interaction.

Training and Incentives: A well-known problem in reputation mechanisms is that users are not inclined to report their experiences unless they are negative. In order for the forgiveness model to work properly, this issue has to be resolved. It is therefore vital that users are trained on why this mechanism is important, what information it requires to work efficiently, and given incentives to report equally on their positive and negative experiences. This issue should be also considered when designing the reporting interfaces so that the information input required is minimal.

Promotion of inhibition: The 'collection' and presentation of judgment factors may enhance prosocial decisions during offences that warrant forgiveness but they may have the opposite effect during severe offences that are well-deserving of punishment. Often, online users are more uninhibited (e.g. [16]) compared to their offline conduct. One could clearly argue that due to this online disposition, higher severity offences emphasized in the interface, may support unjustifiably severe punishments. It is therefore proposed that the forgiveness facilitation takes place only in the event of positive forgiveness decisions, while users can rely on 'traditional' trust and reputation mechanisms during negative forgiveness decisions.

The offender's privacy: In presenting the relevant constituent motivations to the victim of the offence, it could be argued that the offender's privacy is compromised. Although we do not address this issue directly, it is recommended that users are first trained on the purpose of the tools and also given the choice to turn off the forgiveness component if desired.

Objective Ratings: The severity and utility ratings for each offence and collaboration are provided by the overall community. Therefore, the ratings used by the model to make its inference are objective and representative of the collective opinion. To that effect, studies have shown weak correlations between subjective and objective judgments [4]. In arguing for personalization rather than objectivity, we choose objectivity as we believe it is important to promote a collective view rather than to allow for individuals skewed

judgments. While we believe that objective assessments of this kind are important, users autonomy should be respected. In that sense, the interface will output the eleven constituent motivations so that users' decisions can be informed both by their own judgment but also by the FIS inference.

6.2 Further Work

Further work on forgiveness will focus on four separate lines of investigation: refining the constituent motivation formulas, evaluating the fuzzy rules, designing the presentation of the facilitation tool and exploring the impact of the forgiveness mechanism on human behaviour. Specifically, we intend to address the first point with the design of more sophisticated formulas for the eleven constituents. For instance, the victim's beneficial history with the offender can be seen in light of the utility and frequency of benefits, while it is also possible to measure utility in terms of relative utility. An offender of medium utility may be considered a good partner in a community of overall low beneficial transactions. Secondly, although the fuzzy rules were tailored around the theory of forgiveness, there is still a need to evaluate the forgiveness mechanism to determine whether the inference is accurate. This will be done through a series of questionnaires correlating users' judgments to the ones generated by the system. Thirdly, as discussed earlier, the word forgiveness is loaded with different meaning, depending on who the speaker is. In designing an intelligent interface which will facilitate forgiveness, it is important to convey the constituent motivations and the final forgiveness decision in the appropriate language. Finally, the most important point of interest is whether the forgiveness mechanism offers the benefits hypothesized. Some open research questions on this topic are whether people will follow the forgiveness recommendation and if the act of forgiving via this form of facilitation will alleviate anger resulting from a disrupted interaction.

Acknowledgements

Special thanks to Paolo Petta, Maria Chli and Jens Riegelsberger for their valuable input. Theodore Georgas is gratefully acknowledged for his assistance. The reviewers' comments were especially helpful. This research was funded by the HUMAINE IST Framework VI Network of Excellence.

References

1. S. Boon and L. Sulsky. Attributions of blame and forgiveness in romantic relationships: A policy-capturing study. *Journal of Social Behavior and Personality*, 12:19–26, 1997.
2. A. Buss. *Self-consciousness and social anxiety.* W. H. Freeman: San Francisco, CA, 1980.
3. J. Exline, E. Worthington, P. Hill, and M. McCullough. Forgiveness and justice: A research agenda for social and personality psychology. *Personality and Social Psychology Review*, 7:337–348, 2003.

4. F. Fincham, H. Jackson, and S. Beach. Transgression severity and forgiveness: Different moderators for objective and subjective severity. *Journal of Social and Clinical Psychology*, (in press).
5. R. Gruen and G. Mendelsohn. Emotional responses to affective displays in others: The distinction between empathy and sympathy. *Journal of Personality and Social Psychology*, 51:609–614, 1986.
6. M. Holmgren. Forgiveness and the intrinsic value of persons. *American Philosophical Quarterly*, 30(4):341–451, 1993.
7. B. Kelln and J. Ellard. An equity theory analysis of the impact of forgiveness and retribution on transgressor compliance. *Personality and Social Psychology Bulletin*, 25:864–872, 1999.
8. A. Manstead and G. Semin. Social transgression, social perspectives, and social emotionality. *Motivation and Emotion*, 5:249–261, 1981.
9. D. Marcus, J. Wilson, and R. Miller. Are perceptions of emotion in the eye of the beholder? a social relations analysis of judgments of embarrassment. *Personality and Social Psychology Bulletin*, 22:1220–1228, 1996.
10. M. McCullough. Forgiveness who does it and how do they do it. *Psychological Science*, 10(6):194–197, 2001.
11. M. McCullough, K. Rachal, S. Sandage, E. Worthington, S. Brown, and T. Hight. Interpersonal forgiving in close relationships: Ii. theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, 75:1586–1603, 1998.
12. M. McCullough, E. Worthington, and K. Rachal. Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, 73(2):321–336, 1997.
13. R. Miller. Empathic embarrassment: Situational and personal determinants of reactions to the embarrassment of another. *Journal of Personality and Social Psychology*, 53:1061–1069, 1987.
14. A. Modigliani. Embarrassability scale. In *Measures of personality and social psychological attitudes*, pages 173–176. Academic Press: San Diego, CA, 1991.
15. P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. *Communications of the ACM*, 43(12):45–48, 2000.
16. W. Stritzke, A. Nguyen, and K. Durkin. Shyness and computer-mediated communication: A self-presentational theory perspective. *Media Psychology*, 6(1):1–22, 2004.
17. M. Sugeno. *Industrial Application of Fuzzy Control*. Elsevier, 1985.
18. A. Vasalou and G. Piolle. Fuzzy rules of forgiveness. http://www.luminainteractive.com/research/frules.pdf, 2005.
19. A. Vasalou and J. Pitt. Reinventing forgiveness: A formal investigation of moral facilitation. In *Trust Management, Third International Conference, iTrust 2005*, volume LNCS3477, pages 146–160. Springer, 2005.
20. C. Witvliet, E. Worthington, N. Wade, and J. Berry. Justice and forgiveness: Three experimental studies. Presentation at the Christian Association of Psychological Studies, Arlington Heights, IL, 2002.
21. E. Worthington and M. Scherer. Forgiveness is an emotion-focused coping strategy that can reduce health risks and promote health resilience: Theory, review, and hypotheses. *Psychology and Health*, 19:385–405, 2004.