Self-disclosure Decision Making based on Intimacy and Privacy

Jose M. Such[a], Agustín Espinosa[a], Ana García-Fornes[a], Carles Sierra[b]

[a]*Departament de Sistemes Informàtics i Computació*
Universitat Politècnica de València, Camí de Vera s/n, València, Spain
{jsuch,aespinos,agarcia}@dsic.upv.es
[b]*Institut d'Investigació en Intel.ligència Artificial*
Spanish Scientific Research Council, UAB, 08193 Bellaterra, Catalonia, Spain
sierra@iiia.csic.es

Abstract

Autonomous agents may encapsulate their principals' personal data attributes. These attributes may be disclosed to other agents during agent interactions, producing a loss of privacy. Thus, agents need self-disclosure decision-making mechanisms to autonomously decide whether disclosing personal data attributes to other agents is acceptable or not. Current self-disclosure decision-making mechanisms consider the direct benefit and the privacy loss of disclosing an attribute. However, there are many situations in which the direct benefit of disclosing an attribute is a priori unknown. This is the case in human relationships, where the disclosure of personal data attributes plays a crucial role in their development. In this paper, we present self-disclosure decision-making mechanisms based on psychological findings regarding how humans disclose personal information in the building of their relationships. We experimentally demonstrate that, in most situations, agents following these decision-making mechanisms lose less privacy than agents that do not use them. [1]

Keywords: Multi-agent Systems, Privacy, Intimacy, Information Theory

1. Introduction

An autonomous agent usually encapsulates personal data attributes (PDAs) describing its principal [3, 25]. PDAs can describe a great range of topics [20]. For instance, names (real names, pseudonyms), physical characteristics, preferences, roles in organizations and institutions, social characteristics (affiliation to groups, friends), location (permanent address, geo-location at a given time), reputation, competences, personality, psychological state, behaviors, and other private information. When agents carry out interactions on behalf of

[1]NOTICE: this is the author's version of a work that was accepted for publication in Information Sciences. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published: Jose M. Such, Agustin Espinosa, Ana Garcia-Fornes and Carles Sierra. Self-disclosure Decision Making based on Intimacy and Privacy. Information Sciences, Vol. 211 pp. 93-111 (2012). `http://www.sciencedirect.com/science/article/pii/S0020025512003301`

their principals, they usually exchange PDAs. Hence, they play a crucial role to safeguard and preserve their principals' privacy [3].

Westin [27] defined privacy as a "personal adjustment process" in which individuals balance "the desire for privacy with the desire for disclosure and communication". Humans have different general attitudes towards privacy that influence this adjustment process [18, 1, 27]: *privacy fundamentalists* are extremely concerned about privacy and reluctant to disclose PDAs; *privacy pragmatists* are concerned about privacy but less than fundamentalists and they are willing to disclose PDAs when some benefit is expected; and finally, *privacy unconcerned* do not consider privacy loss when disclosing PDAs. In online interactions, just 10% of users are unconcerned [28]. Therefore, privacy is of actual concern to most users in the digital world [13].

Westin proposed his definition for privacy long before the explosive growth of the Internet. As far as we are concerned, it also applies to autonomous agents that engage in online interactions that require the disclosure of their principals' PDAs. Agents, then, should be able to autonomously balance their desire for privacy and their desire for disclosure and communication. Thus, they need to incorporate self-disclosure[2] decision-making mechanisms allowing them to autonomously decide whether disclosing PDAs to other agents is acceptable or not.

Most of the current self-disclosure decision-making mechanisms are based on the privacy-utility tradeoff [10, 19, 14, 29]. This tradeoff considers the direct benefit of disclosing a PDA and the privacy loss it may cause; for instance, the tradeoff between the reduction in time to perform an online search when some PDAs (e.g. geographical location) are disclosed and the privacy loss due to such disclosure [10].

There are many cases where the direct benefit of disclosing PDAs is not known in advance. This is the case in human relationships, where the disclosure of PDAs in fact plays a crucial role in the building of these relationships [5]. These relationships may or may not eventually report a direct benefit for an individual. For instance, a close friend tells you what party he voted for. He may disclose this information without knowing (or expecting) the future gain in utility this may cause. Indeed, this disclosure may not report him any benefit.

Moreover, current self-disclosure decision making models do not consider repeated disclosures and their implications. These implications have been broadly studied in psychology, which has lead to findings regarding how humans disclose personal information in the building of their relationships, such as the well-known *disclosure reciprocity* phenomenon [5]. This phenomenon is based on the observation that one person's disclosure encourages the disclosure of the other person in the interaction, which in turn, encourages more disclosures from the first person.

In this paper, we propose self-disclosure decision-making mechanisms that consider the disclosure reciprocity phenomenon and that relationships may not report any benefit (or this benefit may not be known in advance). An example of the application of these self-disclosure decision-making mechanisms in the long term is computer-mediated communication tech-

[2]We consider self-disclosure as the process by which individuals *disclose* PDAs about themselves to others [5].

nologies such as Internet-based social networking sites (e.g., Facebook, which has more than 800 million active users[3]), in which users disclose personal information and they establish (or develop) relationships to others [8]. These decision-making mechanisms could be used to aid, mediate, or even (partially) automate disclosures in these environments, in which privacy is indeed of great concern [32]. Moreover, Tim Berners-Lee, who is one of the fathers of the WWW as well as the Semantic Web, claims in [30] that the future of social networking is decentralized social networks such as Diaspora[4] and technologies such as the Friend of a Friend[5] (FOAF) [31] ontology for connecting decentralized social Web sites, and the people they describe. In decentralized social networks users have more control of their PDAs because PDAs can be stored locally in a device directly controlled by the user itself rather than in a centralized social network site. Thus, the user can control to whom she/he discloses personal information. Moreover, autonomous agents and Multi-agent Systems have the potential to fit well in this scenario because of their inherently distributed nature.

Our self-disclosure decision-making mechanisms are based on intimacy and privacy measures. We use these self-disclosure decision-making mechanisms to model privacy pragmatist and fundamentalist agents. Then, we compare the performance of pragmatists and fundamentalists to agents that are not equipped with these mechanisms, which we will refer to as unconcerned agents. We claim that, privacy pragmatist agents lose less privacy than unconcerned agents in order to achieve the same intimacy level. We also claim that privacy fundamentalist agents lose less privacy than both pragmatist and unconcerned agents but are unable to achieve the same intimacy. To prove these claims, we first present metrics grounded on information theory to measure the intimacy and the privacy loss between two agents; second, we present self-disclosure decision making mechanisms based on these metrics; and third, we present experiments performed comparing agents using these self-disclosure decision-making mechanisms with privacy unconcerned agents that do not use them.

The remainder of the paper is organized as follows. Section 2 introduces Uncertain Agent Identities (UAIs), which is a formalism for describing agent's believes based on PDAs. Section 3 presents a measure for the degree of intimacy between two agents based on UAIs. Section 4 presents a model for measuring the privacy loss of PDA disclosures based on UAIs. Section 5 proposes self-disclosure decision-making mechanisms for autonomous agents based on intimacy and privacy loss. Section 6 presents the experiments we carried out. Section 7 discusses related work. Finally, Section 8 presents some concluding remarks.

2. Uncertain Agent Identities

We assume a Multiagent System composed of a set of intelligent autonomous agents $Ag = \{\alpha_1, \ldots, \alpha_M\}$ that interact with each other through message exchanges. Agents in Ag are described using the same finite set of PDAs, $A = \{a_1, \ldots, a_N\}$. Each PDA $a \in A$ has a finite domain of possible values $V_a = \{v_1, \ldots, v_{K_a}\}$.

[3] http://www.facebook.com/press/info.php?statistics

[4] http://www.joindiaspora.com

[5] http://www.foaf-project.org/

Each agent $\alpha \in Ag$ has values for their PDAs that are not known by the other agents in Ag. Agents are able to disclose PDA values to others, but the values of the PDAs disclosed may not be true (or may be just an opinion). Thus, agents are uncertain about the PDA values of the other agents. Moreover, agents may not even be absolutely certain about the specific values for their own PDAs (e.g. an agent could be uncertain about whether it is competent in performing a given task). Therefore, agents maintain *uncertain agent identities* (UAIs) modeling their own PDAs and the PDAs of the rest of the agents in Ag.

Definition 1 (Uncertain Agent Identity). *Given a set of PDAs $A = \{a_1, \ldots, a_N\}$, each one with domain $V_a = \{v_1, \ldots, v_{K_a}\}$, an uncertain agent identity $I = \{P_1, \ldots, P_N\}$ is a set of discrete probability distributions P_i over the values V_{a_i} of each PDA a_i.*

We thus denote P_a as the probability distribution of a over V_a and $p_a(\cdot)$ as its probability mass function, so that $p_a(v)$ is the probability for the value of a being equal to $v \in V_a$.

An agent $\alpha \in Ag$ manages its own UAI and two UAIs associated to each agent $\beta \in Ag \setminus \{\alpha\}$. We will refer to the UAI of an agent α as I_α. We denote $I_{\alpha,\beta}$ as the UAI that α believes that β has, i.e., what α knows (or thinks it knows) about I_β. Moreover, it is crucial for an agent α to also have UAIs modeling what the other agents in Ag may know about its own UAI I_α for measuring privacy loss (as explained in section 4). We denote $I_{\alpha,\beta,\alpha}$ as the UAI that α believes that β believes that α has[6].

UAIs may be initialized regarding the actual knowledge that an agent has for the probability distributions of each of the PDAs. For instance, if the agent is completely uncertain about the distribution of a PDA a, then, its probability distribution $P_a \in I$ may be initialized to a uniform distribution, i.e., each $p_a(v)$ may be initialized to $\frac{1}{|V_a|}$ for each $v \in V_a$.

2.1. Uncertainty Measures

An agent may desire to measure how much uncertainty there is in the probability distribution of a PDA. Taking into account this uncertainty, the agent may decide, for instance, whether or not to take specific actions to reduce this uncertainty under a desired threshold.

A well-known measure of the uncertainty in a probability distribution is Shannon entropy [22]:

$$H(P_a) \quad = \quad -\sum_{v \in V_a} p_a(v) \log_2 p_a(v) \tag{1}$$

The entropy of each probability distribution in an UAI provides a measure of the uncertainty for each PDA. However, as an UAI can span over several PDAs, a method for

[6]Subindexes of an UAI should be read from left to right, starting with *the UAI* and adding an *that agent believes* for each agent that appears separated by a semicolon, except for the agent in the last position which is read as *that agent has*. For instance, I_α should be read as *the UAI that α has*, $I_{\alpha,\beta}$ should be read as *the UAI that α believes that β has* and $I_{\alpha,\beta,\alpha}$ should be read as *the UAI that α believes that β believes that α has.*

aggregating the uncertainties of all of the probability distributions in an UAI is needed. In this paper, we use a simple computational method that is the mean of the uncertainties in each of the probability distributions in an UAI:

$$H(I) = \frac{1}{|A|} \sum_{a \in A} H(P_a) \tag{2}$$

With this measure an agent is able to know how certain it is about an UAI. We assume that at initialization time the entropy of an UAI I is the highest possible, i.e., the uncertainty in I will decrease as the agent obtains more information related to the PDAs being modeled.

2.2. Updating UAIs

UAIs are supposed to be dynamic, i.e., they may change as time goes by. These changes will potentially reduce the uncertainty in an UAI. An agent α may update the UAIs that it manages as it gets more information about the probability distributions for the PDAs in these UAIs. In this section, we provide a method for updating the two UAIs that α has per each agent in Ag.

PDA values are private to each agent. We assume that α *discloses* its PDA values for a to β by sending a message[7] $\mu = \langle \alpha, \beta, \langle \alpha, a, P_a \rangle \rangle$, where α represents the sender, β represents the receiver, and $\langle \alpha, a, P_a \rangle$ represents the claim "the probability distribution for the PDA a of α is P_a".

UAIs are updated with the disclosures that agents carry out. The update process of an UAI has two steps: (i) updating the probability distribution of the PDA being disclosed; and (ii) inferring updates of probability distributions of other PDAs based on the PDA being disclosed and other information already known. We denote that an UAI I is updated with a message μ as I^{μ}. Moreover, we denote that an UAI I is updated sequentially and in order considering a tuple of messages $M = (\mu_1, \ldots, \mu_P)$ as I^M.

We now detail how and which UAIs should be updated when receiving and when sending a message.

2.2.1. Receiving a Message

If α receives $\mu = \langle \beta, \alpha, \langle \beta, a, Q_a \rangle \rangle$ from β, then α can update $I_{\alpha,\beta}$ – the UAI that α believes that β has. The resulting UAI is denoted as $I^{\mu}_{\alpha,\beta}$.

Update. Given $\mu = \langle \beta, \alpha, \langle \beta, a, Q_a \rangle \rangle$, $P_a \in I_{\alpha,\beta}$, and $r_{\alpha,\beta}$ (which is the reliability that α attaches to β and is explained below), let $S^{\mu}_a = r_{\alpha,\beta} \cdot Q_a + (1 - r_{\alpha,\beta}) \cdot P_a$. Then, we update $P^{\mu}_a \in I^{\mu}_{\alpha,\beta}$ as:

$$P^{\mu}_a = \begin{cases} S^{\mu}_a & \text{if } H(S^{\mu}_a) < H(P_a) \\ P_a & otherwise \end{cases} \tag{3}$$

P_a is only updated if the message produces an information gain, i.e. resulting probability distribution S^{μ}_a is more certain than P_a.

[7] In this paper, we use the terms message and disclosure as equivalents because we only consider messages that involve a PDA disclosure.

Reliability. The model for reliability is based on the difference between the values that agents claim for their PDAs – the disclosures they send to other agents – and the values observed for these PDAs by other agents. We assume that α builds another UAI $\mathcal{O}_{\alpha,\beta}$ that is different from I_α, $I_{\alpha,\beta}$ and $I_{\alpha,\beta,\alpha}$ based on observations. $\mathcal{O}_{\alpha,\beta}$ contains probability distributions that α has inferred from the observation of β's behavior. An example of observation may be the following. Let $competentTaskA$ be a PDA with domain $\{true, false\}$. If β discloses $\langle \beta, \alpha, \langle \beta, competentTaskA, \{true \to 1, false \to 0\}\rangle\rangle$, α may request β to perform this task. Then, α can *observe* the result of the task to assess whether or not β is actually competent in carrying out the task and may infer the probability distribution for $competentTaskA$ as being $\{true \to 0.8, false \to 0.2\}$.

α measures the reliability of β as follows. Let a be a PDA β disclosed to α, let $P_a \in I_{\alpha,\beta}$ be the probability distribution that α believes that β has (from what β disclosed to α), and let $O_a \in \mathcal{O}_{\alpha,\beta}$ be the probability distribution that α has observed for the PDA a of β. Then, α's assessment of the reliability of β on the basis of observing that P_a should have been O_a is:

$$r_{\alpha,\beta} = \frac{1}{|A|} \sum_{a \in A} \frac{1}{1 + \mathrm{KL}(O_a \parallel P_a)} \tag{4}$$

Where $\mathrm{KL}(O_a \parallel P_a)$ is the Kullback-Leibler divergence [11] that measures the distance between two probability distributions:

$$\mathrm{KL}(O_a \parallel P_a) \;=\; \sum_{v \in V_a} o_a(v) \log_2 \frac{o_a(v)}{p_a(v)} \tag{5}$$

If all the probability distributions that α observed for all the disclosed PDAs from β are close to the probability distributions for these PDAs in $I_{\alpha,\beta}$, then KL values will be close to 0 and $r_{\alpha,\beta}$ will be close to 1. If all the probability distributions α observed for all the disclosed PDAs from β are far from the probability distributions for these PDAs in $I_{\alpha,\beta}$, then KL values will be high and $r_{\alpha,\beta}$ will be close to 0.

Inference. The rest of the probability distributions of PDAs not yet disclosed from β to α may be inferred considering the PDAs that have already been disclosed. The inference model that we consider in this paper is based on the existence of conditional probabilities $\Pr(b \mid a)$[8], considering a as a PDA β disclosed to α and b as the PDA to be inferred. Thus, if Q_b is a probability distribution defined as:

$$q_b(u) \;=\; \sum_{v \in V_a} \Pr(b = u \mid a = v) p_a(v) \tag{6}$$

[8]More sophisticated methods, e.g. based on bayesian networks [6], could be used. The important point is that inference should be considered when dealing with the disclosure of PDAs.

then,

$$P_b^{\mu} = \begin{cases} Q_b & \text{if } H(Q_b) < H(P_b) \\ P_b & otherwise \end{cases} \tag{7}$$

A simple method based on frequencies for estimating these conditional probabilities may be:

$$\Pr(b = u \mid a = v) \quad = \quad \frac{|\{\beta \mid \beta \in Ag \text{ and } P_a, P_b \in I_{\alpha,\beta} \text{ and } p_a(v) > \epsilon \text{ and } p_b(u) > \epsilon\}|}{|Ag| - 1} \tag{8}$$

This method averages the number of UAIs that α believes that other agents in Ag have in which the probabilities for a and b to be v and u are higher than a threshold ϵ. This is a simple method for estimating if the values v and u of PDAs a and b are commonly related to each other for agents in Ag. This method requires a minimum knowledge about the other agents in Ag.

2.2.2. Sending a Message

α discloses the probability distribution for its PDA a to β by sending a message $\mu = \langle \alpha, \beta, \langle \alpha, a, Q'_a \rangle \rangle$ to β. Then, α may update $I_{\alpha,\beta,\alpha}$ – the UAI that α believes that β believes that α has. The resulting UAI is denoted as $I^{\mu}_{\alpha,\beta,\alpha}$.

α updates $P_a \in I_{\alpha,\beta,\alpha}$ replacing it with Q'_a, i.e., α assumes that β believes the probability distribution for its PDA a is Q'_a from this moment on. α may also update the probability distributions of PDAs that α has not yet disclosed to β, which could be inferred from PDAs that α has already disclosed to β using the inference method explained in the above section.

We also consider that α may be not sincere when performing a disclosure. To this aim, we define what we call *the level of insincerity* as follows:

Definition 2 (Level of Insincerity). *Given a disclosure from α to β in the form of the message $\mu = \langle \alpha, \beta, \langle \alpha, a, Q'_a \rangle \rangle$, and the probability distribution $Q_a \in I_\alpha$, the level of insincerity of α in this disclosure is:*

$$S(\mu) \quad = \quad KL(Q'_a \parallel Q_a) \tag{9}$$

Informally speaking, we measure the distance between what α is disclosing and what is in its UAI (I_α). When the level of insincerity $S(\mu)$ is 0, this implies that Q'_a and Q_a are the same probability distributions — recall that KL() returns the distance between two probability distributions, and it returns 0 if the two probability distributions are equal. Thus, when $S(\mu) = 0$ we say that α is *sincere*. Otherwise, when $S(\mu) > 0$ we say that α is *insincere, with a level of insincerity of $S(\mu)$*.

3. Intimacy

According to [17], intimate human partners have extensive personal information about each other. They usually share information about their PDAs, including preferences, feelings, and desires that they do not reveal to most of the other people they know. Indeed,

self-disclosure and partner disclosure of PDAs play an important role in the development of intimacy [5].

An agent α could simply count the number of PDAs disclosed to β, count the number of PDAs that β disclosed to α to estimate its intimacy to β. However, as explained in section 2.2, when disclosing PDAs, it may be the case that more information is being disclosed without explicitly disclosing it. Therefore, PDAs not yet disclosed may be inferred from PDAs already disclosed so that α is actually giving β more information than just the PDAs explicitly disclosed to β.

Uncertainty and information are closely related to each other [9]. The amount of information obtained by an action can be measured by the reduction of uncertainty due to that action. Thus, information may be measured by the difference between the a priori uncertainty – uncertainty before the action – and the a posteriori uncertainty – uncertainty after the action. For instance, as stated in [23], if the action is the sending/reception of a message, the information gain that a message provides may be measured by the difference in uncertainty before sending/receiving the message and the uncertainty after sending/receiving the message.

Definition 3 (Information Gain of a Message). *Given an UAI I and a message μ, the information gain of message μ is:*

$$\mathcal{I}(I, \mu) = H(I) - H(I^{\mu}) \tag{10}$$

α may measure the amount of information it has about β by measuring the information gain of all the messages received from β. α may measure the amount of information β has about it by measuring the information gain of all the messages that α sent to β.

Definition 4 (Information Gain of a tuple of Messages). *Given an UAI I and a tuple of messages M, the information gain of M is:*

$$\mathcal{I}(I, M) = H(I) - H(I^{M}) \tag{11}$$

Sierra and Debenham [24] defined the intimacy between α and β considering the amount of information that α knows about β and vice versa. We adapt this definition for the case of UAIs. Thus, we define intimacy as follows.

Definition 5 (Intimacy). *Given the UAIs $I_{\alpha,\beta}$ and $I_{\alpha,\beta,\alpha}$, a tuple of messages M from β to α and a tuple of messages M' from α to β, the intimacy between α and β is:*

$$\mathcal{Y}_{\alpha,\beta} \;=\; \mathcal{I}(I_{\alpha,\beta}, M) \oplus \mathcal{I}(I_{\alpha,\beta,\alpha}, M')$$

Where $\oplus$ is an appropriate aggregation function. $\mathcal{Y}_{\alpha,\beta} = 0$ means that there is no intimacy between α and β from the point of view of α. The higher the $\mathcal{Y}_{\alpha,\beta}$, the more intimacy between α and β from the point of view of α. It is worth noting that the intimacy measure, as we define it, is not necessarily symmetric, i.e., $\mathcal{Y}_{\alpha,\beta}$ may be different from $\mathcal{Y}_{\beta,\alpha}$.

Intimacy is an amount of information resulting from the aggregation of the amount of information α has from β and α believes β has from α. In the experiments we performed (section 6) we used the arithmetic addition of these two amounts of information, i.e. $\oplus = +$.

4. Privacy Loss

Privacy loss is defined in previous works ([14, 15]) as the probability of being identified and the sensitivity of the PDAs — i.e., the importance of a PDA from a privacy perspective, e.g., a person may probably feel her/his credit card number as being more sensitive than her/his nationality. Thus, if an agent makes a disclosure, this may imply a privacy loss because the agent that receives the disclosure knows the agent that sends the disclosure and the value of the PDA disclosed. The specific amount of privacy loss is determined by: (i) the sensitivity of this PDA, i.e., a more sensitive PDA implies more privacy loss than a less sensitive PDA; (ii) the level of insincerity of the agents when it makes the disclosure, i.e., if the agent is insincere it may not experience privacy loss because the values of the PDA disclosed do not correspond to the values in its own UAI[9]; (iii) and finally, disclosing an attribute may also cause that other PDAs can be inferred from the PDA disclosed.

In order to consider the sensitivity of PDAs, we assume that agents in Ag can define the *subjective* sensitivity that they attach to their PDAs. Therefore, α has a function $w_\alpha : A \to [0, 1]$ such that $w_\alpha(a)$ is the *subjective* valuation that α attaches to the sensitivity of a.

In order to consider the level of insincerity and possible inferences, as explained in Section 2, each agent $\alpha \in Ag$ has its own UAI I_α that is not known by the other agents in Ag. Moreover, α has UAIs that it believes that other agents in Ag believe that α has, i.e., what other agents in Ag may know about I_α. In this sense, α could estimate (from its point of view) the extent to which β knows I_α by measuring the distance between I_α and $I_{\alpha,\beta,\alpha}$. α can calculate this distance by measuring the distance between each probability distribution for each PDA in these UAIs.

Given that a has the probability distributions $P_a \in I_\alpha$ and $Q_a \in I_{\alpha,\beta,\alpha}$, we use the Kullback-Leibler divergence [11] to measure the distance between P_a and Q_a. KL measures the amount of information needed to encode the differences between two probability distributions.

Based on the Kullback-Leibler divergence and the sensitivity of the PDAs, we define the privacy loss of disclosing a PDA.

Definition 6 (Privacy Loss). *Given two agents α and β, the message μ, and considering $Q_a \in I_{\alpha,\beta,\alpha}$, $Q_a^\mu \in I^\mu_{\alpha,\beta,\alpha}$ and $P_a \in I_\alpha$, the privacy loss for agent α if it sends μ to agent β is:*

$$\mathcal{L}(I_{\alpha,\beta,\alpha}, \mu) = \sum_{a \in A} w_\alpha(a) \cdot (\mathrm{KL}(Q_a \parallel P_a) - \mathrm{KL}(Q_a^\mu \parallel P_a)) \tag{12}$$

For each PDA, we measure the KL between its probability distribution in $I_{\alpha,\beta,\alpha}$ before being updated taking into account μ and its probability distribution in I_α and the KL

[9]Note that this insincere disclosure can still produce an information gain to the agent that receives the disclosure.

between its probability distribution in $I_{\alpha,\beta,\alpha}$ after being updated considering μ and its probability distribution in I_α. Then, we consider the difference between these two KLs stating the amount of information that $I_{\alpha,\beta,\alpha}$ would approach to I_α if the message μ is sent. This amount of information that would be *lost* due to the sending of the message is then weighted by the subjective sensitivity of the PDA. The final result of privacy loss is the addition of the results for all of the PDAs (recall that values for PDAs that are not disclosed could be inferred from PDAs that are disclosed as explained in Section 2). $\mathcal{L}(I_{\alpha,\beta,\alpha}, \mu) = 0$ means that sending μ to β causes no privacy loss to α. The higher the $\mathcal{L}(I_{\alpha,\beta,\alpha}, \mu)$, the more privacy loss sending μ to β causes to α.

As explained later on in section 5, it is also useful for agents to measure the total privacy that they have lost due to the messages that they sent to other agents.

Definition 7 (Total Privacy Loss). *Given two agents α and β, the tuple of all messages M sent from α to β and considering $Q_a \in I_{\alpha,\beta,\alpha}$, $Q_a^M \in I^M_{\alpha,\beta,\alpha}$ and $P_a \in I_\alpha$, the Total Privacy Loss from α to β is:*

$$\mathcal{L}(I_{\alpha,\beta,\alpha}, M) = \sum_{a \in A} w_\alpha(a) \cdot (\mathrm{KL}(Q_a \parallel P_a) - \mathrm{KL}(Q_a^M \parallel P_a)) \tag{13}$$

5. Self-disclosure Decision Making

In this section, we present two mechanisms for an agent α to decide which PDAs (if any) to disclose to another agent β. These mechanisms are based on general privacy attitudes and specific willingness to share a PDA. We model pragmatist and fundamentalist attitudes towards privacy. To this aim, we use the information metrics explained above.

5.1. Privacy Pragmatist Agents

Privacy pragmatists are concerned about privacy, but they are willing to disclose personal information when some benefit is expected ([18], [1] and [27]). In many situations, the actual benefit of disclosing personal information may not be known in advance. We present a self-disclosure decision-making mechanism modeling a pragmatic attitude towards privacy which is grounded on information measures. Specifically, we consider the estimation of intimacy gain between two agents (i.e., the amount of information two agents have about each other) and the privacy loss (the distance between what the agents believe that others believe about them and their actual UAI weighted by a subjective sensitivity).

We model a privacy pragmatist agent α as an agent that chooses to disclose a PDA that maximizes the estimation of the increase in intimacy (described in Section 3) while at the same time minimizing the privacy loss (described in Section 4). We call this tradeoff the *privacy-intimacy* tradeoff. The privacy-intimacy tradeoff is a multi-objective optimization problem. The most used approach to solve this kind of problems in the existing literature on multi-objective optimization is the transformation of the multi-objective optimization problem into a single-objective optimization problem [4]. Thus, we model a pragmatist

agent α as an agent that maximizes the difference between the increase in intimacy and the privacy loss[10].

Formally, let M be a tuple of messages that α sent to β, α will choose to disclose μ^* so that:

$$\mu^* = \arg\max_{\mu}(\mathcal{I}(I^M_{\alpha,\beta,\alpha}, \mu) - \mathcal{L}(I^M_{\alpha,\beta,\alpha}, \mu)) \tag{14}$$

One can easily note that the intimacy measure is not explicitly used in the privacy-intimacy tradeoff formula (Equation 14). This is because we implicitly estimate the increase of intimacy based on two main aspects: (i) we explicitly consider the information gain that a disclosure from α may cause to β, i.e., $\mathcal{I}(I^M_{\alpha,\beta,\alpha}, \mu)$; (ii) we then assume, based on the *disclosure reciprocity* phenomenon [5], that β will reciprocate this information gain with a disclosure to α in the future — later on in Section 5.1.1 we explain how α can check this and act consequently in the event β not reciprocating to α. Therefore, maximizing $\mathcal{I}(I^M_{\alpha,\beta,\alpha}, \mu)$ will also maximize the intimacy $\mathcal{Y}_{\alpha,\beta} = \mathcal{I}(I^M_{\alpha,\beta,\alpha}, \mu) \oplus \mathcal{I}(I^{M'}_{\alpha,\beta}, \nu)$, considering ν as a future message received by α from β as the reciprocation to μ. This is due to the cumulative nature of intimacy as we define it.

The privacy loss part of the privacy-intimacy tradeoff formula (Equation 14) is the only one that considers the sensitivity of what is to be disclosed (as can be seen in the definition of privacy loss in Equation 12). This is because, when an agent performs a disclosure, it knows how sensitive that disclosure is for itself. However, an agent may not be able to anticipate how sensitive this disclosure will be seen by the agent that receives the disclosure — as it would be required for considering sensitivity in the information gain part of the privacy-intimacy tradeoff formula, i.e., α would need to know the sensitivity function of β w_β. For instance, Huberman et al. [7] demonstrated that people whose weight was less than average value their weight as a less sensitive issue, while people whose weight was greater than average value that weight as a more sensitive issue. Thus, a person may be very reluctant to disclose her or his weight while another person may be willing to disclose her or his weight. In the event that the first person discloses her or his weight to the second person, the first person will feel this disclosure as very sensitive. However, the second person will receive this disclosure as little sensitive. Mechanisms for estimating the sensitivity that other agents have are out of the scope of this paper, but they represent a very challenging future line of research. One approach could be based on the semantics of the PDAs (e.g., considering an ontology such as in [23]).

Finally, it is worth noting that agents may not be sincere when they perform disclosures. Agent α will choose a message μ^* that maximizes the amount of information for the privacy-intimacy tradeoff. Considering that $\mu^* = \langle \alpha, \beta, \langle \alpha, a, Q_a \rangle \rangle$ and $P_a \in I_\alpha$, to model sincere agents when disclosing a PDA, μ^* must satisfy that $S(\mu^*) = \mathrm{KL}(Q_a \parallel P_a) = 0$. That is, Q_a, which is the probability distribution that α sends, and P_a, which is the probability distribution that is on α's own UAI I_α, must be the same probability distribution. To model

[10]Other approaches to solve this kind of problems can also be applied. To learn more approaches to solve multi-objective problems refer to, for instance, [2] and [4].

agents that are insincere when disclosing a PDA, μ^* must satisfy that $S(\mu^*) = \mathrm{KL}(Q_a \parallel P_a)$ matches the desired level of insincerity.

5.1.1. Balance

Agent α assumes, based on the reciprocity phenomenon, that β will reciprocate its disclosures, so that $\mathcal{I}(I_{\alpha,\beta,\alpha}, \mu)$ is an estimation for $\mathcal{I}(I_{\alpha,\beta}, \nu)$, considering ν as a future message received by α from β. However, this may be not the case for many reasons, such as if β is not reliable when it makes claims about itself. For instance, β may not be reliable if β is not sincere when it makes claims about itself or if β is unable to provide reliable information about itself. Moreover, there could be agents that do not reciprocate disclosures because they are not willing to increase their intimacy to α for whatever reason (e.g. agents that are only interested in surveilling information about α). This could lead to $\mathcal{I}(I_{\alpha,\beta,\alpha}, \mu) >> \mathcal{I}(I_{\alpha,\beta}, \nu)$ if ν is actually received.

α may assess to what extent β will reliably reciprocate future disclosures from α by considering the amount of information that β has sent to α and the amount of information that α has sent to β. To this aim, we use the concept of balance [24].

Definition 8 (Balance). *Given the UAIs $I_{\alpha,\beta}$ and $I_{\alpha,\beta,\alpha}$, a tuple of messages M from β to α and a tuple of messages M' from α to β, the balance between α and β from the point of view of α is:*

$$\mathcal{B}_{\alpha,\beta} = \mathcal{I}(I_{\alpha,\beta}, M) - \mathcal{I}(I_{\alpha,\beta,\alpha}, M') \tag{15}$$

Agent α may use the balance $\mathcal{B}_{\alpha,\beta}$ as the basis for a disclosure strategy. That is, α may use the balance to assess to what extent β will reliably reciprocate future disclosures from α. Then, α may decide not to perform a disclosure to β if $\mathcal{B}_{\alpha,\beta} < \zeta$. In this case, ζ is what we call the *reciprocity threshold* that acts as a threshold of the minimum balance that α expects from its interaction partners. Moreover, α may specify a different ζ_β for each agent $\beta \in Ag$. In this way, α may even consider an increasing ζ_β as the intimacy $\mathcal{Y}_{\alpha,\beta}$ increases so that $\zeta'_\beta = \zeta_\beta + \lambda \cdot \mathcal{Y}_{\alpha,\beta}$, where λ is a normalizing constant. Using this dynamic ζ_β, we can model, for instance, that intimate partners can trust each other more than simple acquaintances can.

An approach to obtain an appropriate ζ (or an initial ζ_β) can be based on the existing polls to obtain the privacy attitude of humans, such as [26] or any of the other surveys that Alan Westin conducted between 1978 and 2004 [12]. In the experiments section (specifically in Section 6.3) we present an interval of the values of ζ that can make pragmatic agents to behave different. Therefore, based on the responses to a poll to obtain the privacy attitude of humans, we can obtain a degree of privacy attitude that can be directly matched to a reciprocity threshold in that interval. For instance, a person that has a degree of privacy attitude that is considered pragmatist but is very close to unconcerned may be modeled with a $\zeta = -5$ (as shown in Section 6.3).

5.2. Privacy Fundamentalist Agents

Privacy fundamentalists are extremely concerned about privacy and very reluctant to disclose PDAs ([18], [1] and [27]). They feel like they have already lost much privacy and are not willing to lose privacy any more.

We model fundamentalist agents as pragmatist agents that establish a maximum total privacy loss ξ. In this way, a fundamentalist agent α considers the privacy-intimacy tradeoff to decide what PDA (if any) to disclose to β. α also considers the balance $\mathcal{B}_{\alpha,\beta}$ to assess to what extent β will reliably reciprocate future disclosures from α. Then, α may decide not to perform a disclosure to β if $\mathcal{B}_{\alpha,\beta} < \zeta$. The difference between pragmatists and fundamentalists is the following. If α is a fundamentalist agent, when the total privacy loss of α to β reaches ξ, α will not disclose PDAs to β any more.

Suppose that α has sent a sequence of messages $M = \{\mu_1, \ldots, \mu_P\}$ to β. Then, let $\rho = min_\mu \mathcal{L}(I^M_{\alpha,\beta,\alpha}, \mu)$, i.e., ρ is the minimum privacy loss for α if she decides to disclose any PDA not yet disclosed to β. α will not disclose any other PDA to β if $\rho + \mathcal{L}(I_{\alpha,\beta,\alpha}, M) > \xi$.

Moreover, α may specify a different ξ_β for each agent $\beta \in Ag$. In this way, α may consider an increasing ξ_β as the intimacy $\mathcal{Y}_{\alpha,\beta}$ increases so that $\xi'_\beta = \xi_\beta + \lambda \cdot \mathcal{Y}_{\alpha,\beta}$, where λ is a normalizing constant.

An approach to obtain an appropriate fundamentalist threshold (ξ) can be based on the existing polls to obtain the privacy attitude of humans, in a similar way than as explained in the previous section for the reciprocity threshold ζ. In this case, the higher the degree of fundamentalism, the lower the fundamentalist threshold.

6. Implementation and Experimental Results

We implemented unconcerned, pragmatist and fundamentalist agents in Java. We implemented pragmatist and fundamentalist agents as agents that use the self-disclosure decision-making mechanisms explained in Section 5. On the contrary, we implemented unconcerned agents as agents that do not use the mechanisms presented in this paper. Specifically, we implemented unconcerned agents as agents that do not take into account privacy loss when disclosing PDAs to other agents. We considered unconcerned, pragmatist and fundamentalists to be sincere when disclosing a PDA.

We performed experiments in which unconcerned, pragmatist, and fundamentalist agents interact with other *target* agents. For each experiment, we calculated the intimacy that each agent achieved with each target agent and the total privacy that each agent lost with each target agent. The results for intimacy and privacy loss are given in bits because the unit of measure for information is the bit when base 2 logarithms are used to calculate entropies and KLs. Moreover, the results are the average of the results obtained when repeating each experiment 100 times.

The experiments that we performed were composed of a number of disclosure rounds (DRs). In each DR, agents were given the chance to choose an interaction partner and then decide whether or not to perform a new disclosure to that particular interaction partner. Agents may or may not disclose one attribute to that particular interaction partner. Thus, agents perform at most one disclosure for each DR.

The UAI I_α of each agent in each experiment was created as randomly generated distributions P_a for each PDA $a \in A$ over the domain V (both A and V are specified in Table 1). The probability distributions in the UAIs that each agent has modeling other agents and what other agents might know about it (i.e., in case of agent α all the UAIs $I_{\alpha,\beta}$ and $I_{\alpha,\beta,\alpha}$ for each $\beta \in A$) are initialized to uniforms over V, i.e., agents are completely uncertain about the UAIs that other agents have at initialization time.

We performed several experiments varying all of the possible parameters. Specifically, we performed experiments considering: only sincere and reciprocating targets (Section 6.1); a varying number of malicious targets (Section 6.2); different reciprocity and fundamentalist thresholds (Sections 6.3 and 6.4); a unique shared sensitivity array and one sensitivity array per agent (Section 6.5); and finally, different number of agents, attributes and values (Section 6.6).

We implemented all agents as capable of making observations for the attributes of other agents. In this way, agents can call to an `observe()` method to obtain observed distributions for attributes of other agents. Then, agents estimate the reliability of other agents using these observed values and the values other agents claimed for themselves (i.e. the disclosures they made) as inputs for the reliability model presented in section 2. We assumed that when an agent calls the `observe()` method, it always returns the correct probability distribution for an attribute of another agent. However, we also conducted an experiment (Section 6.7) in which we consider that the `observe()` method introduces a random normally-distributed noise when it returns the probability distribution for an attribute of another agent.

6.1. Sincere and Reciprocating Targets

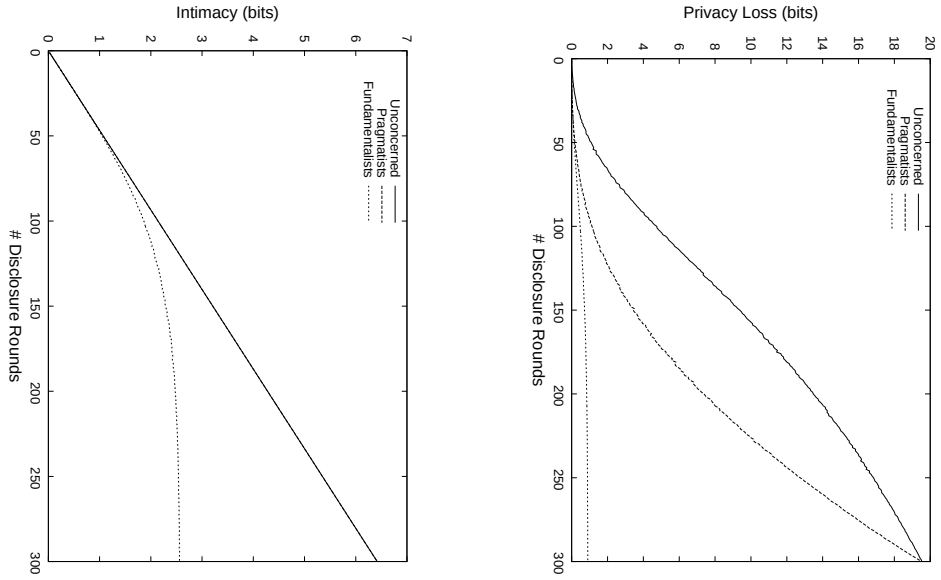
In this section, we present the experiments that we performed comparing unconcerned, pragmatist, and fundamentalist agents when interacting with other target agents. These target agents reciprocate all of the disclosures they receive. Moreover, they perform such reciprocations in a sincere way, i.e., a target agent α reciprocates with a level of insincerity of $\mathrm{KL}(Q'_a \parallel Q_a) = 0$ that means that Q'_a and Q_a are the same distribution, considering Q'_a as the distribution disclosed and Q_a as the distribution in I_α for the a attribute.

The parameters used for this experiment are summarized in Table 1. We considered 10 unconcerned agents, 10 pragmatist agents, 10 fundamentalist agents, and 30 target agents. We also considered 10 PDAs with a domain of 10 values each of them. Agents have a shared randomly distributed array of sensitivities for each of the attributes. Moreover, the reciprocity threshold and the fundamentalist threshold are set to -1 and 2 respectively. All of these parameters are varied in the subsequent sections. Finally, we performed experiments varying the number of DRs ranging from 1 DR to 300 DRs. The maximum number of DRs is 300 DRs because it is equal to the number of target agents (30) multiplied by the number of attributes (10). Thus, this number of DRs is enough for any agent (regardless its attitude) to have the chance to disclose all of their PDAs to all of the target agents.

Figure 1(a) shows the average intimacy achieved by the agents for each number of DRs considered. Both unconcerned and pragmatist agents achieve the same intimacy for all of the experiments. Moreover, both unconcerned and pragmatist agents achieve more intimacy with target agents than fundamentalists. This is because when fundamentalists reach their

Parameter	Description	Value
Nun	# Unconcerned	10
Npr	# Pragmatists	10
Nfu	# Fundamentalists	10
Nta	# Target Agents	30
A	Personal Data Attributes	$\{a_1, \ldots, a_{10}\}$
V	PDAs' Domain	$\{v_1, \ldots, v_{10}\}$
w	Subjective Sensitivity	Random $[0, 1]^{10}$
ζ	Reciprocity threshold	-1
ξ	Fundamentalist Threshold	2
NDRs	Max. # of Disclosure Rounds	300
NRep	# of Repetitions of the Experiment	100

Table 1: Parameters used in the experiments.



(a) Intimacy

(b) Privacy Loss

Figure 1: Results considering sincere and reciprocating target agents.

maximum privacy loss ξ, they will no longer disclose PDAs so that intimacy is no longer increased.

Figure 1(b) shows the averaged privacy loss of the agents for each number of DRs. As expected, pragmatist agents lost less privacy than unconcerned agents for most of the experiments. For instance, for 16 DRs unconcerned agents lost 10 times more privacy than pragmatists; for 60 DRs unconcerned agents lost 5 times more privacy than pragmatists; for 130 DRs unconcerned lost 3 times more privacy than pragmatists; for 180 DRs uncon-

cerned agents lost twice the privacy that pragmatists lost; and for 220 DRs unconcerned agents lost 1.5 times more privacy than pragmatists. Therefore, for most of the experiments performed, pragmatist agents lost less privacy than unconcerned agents while achieving the same intimacy.

The privacy loss was similar for both pragmatist and unconcerned agents in the experiments with a high number of DRs (from 270 up to 300 DRs). This is because, in these experiments, the agents disclosed almost all of their PDAs to all of the agents, so that they ended up losing all their privacy regardless their privacy attitude.

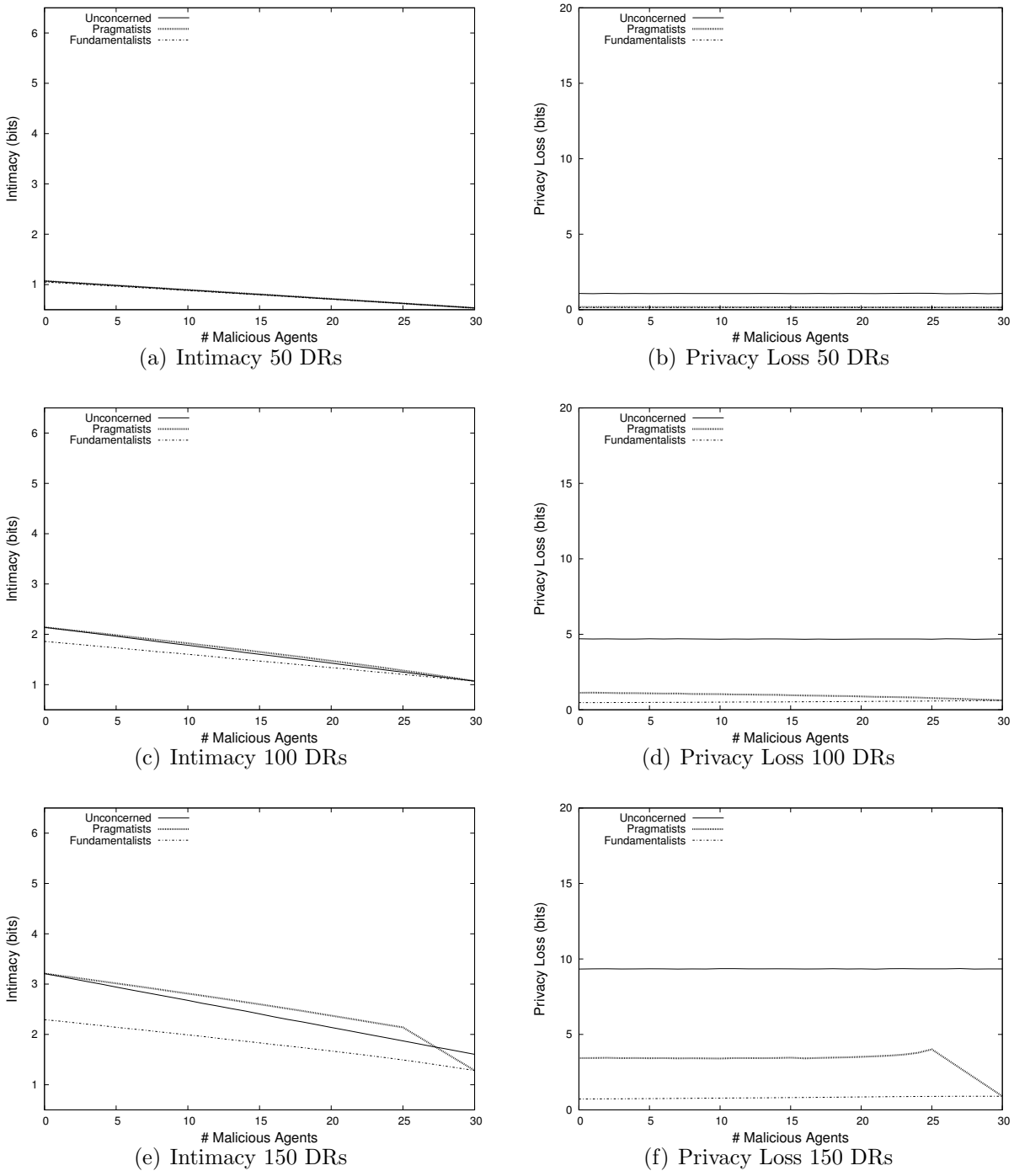
As expected, pragmatist and fundamentalist agents lost less privacy than unconcerned agents. Moreover, fundamentalists lost less privacy than pragmatist agents. This is due to the fact that fundamentalists do not lose privacy beyond the threshold they define ξ.

6.2. Malicious Targets

In these experiments, we consider unconcerned, pragmatist, and fundamentalist agents interacting with target agents. For each experiment, we establish a number of malicious target agents (MTs) among the target agents. We consider malicious agents to be agents that are only interested in obtaining information from other agents without increasing intimacy. We model malicious agents as agents that either do not reciprocate or lie (are not sincere) about themselves. We implemented malicious agents such that when they receive a disclosure they do not reciprocate with a probability of 0.5. Moreover, when they reciprocate (the other 0.5 times) they are not sincere. We implemented malicious agents with a level of insincerity of 5 bits (recall that a level of insincerity of 0 means that an agent is completely sincere when disclosing her attributes). Thus, a malicious target agent α reciprocate with a level of insincerity of $\mathrm{KL}(Q'_a \parallel Q_a) = 5$ considering Q'_a as the distribution disclosed and Q_a as the distribution in its UAI I_α for the a attribute. Therefore, there are 5 bits of difference between the distribution disclosed and the distribution in her UAI.

The parameters used for the experiments are the same as the ones in Table 1. We performed experiments varying the number of MTs from 0 up to 30. Thus, we model environments in which agents interact with a varying % of MTs among the target agents from 0% up to 100% (recall that the number of target agents Nta is set to 30). Moreover, we also considered a varying number of DRs from 50 up to 300. This allows us to assess the properties of the intimacy and privacy loss metrics when agents have few chances to disclose any of their PDAs to the target agents (50 DRs), and when agents have the chance to disclose all of their PDAs to all of the target agents (300 DRs) — but recall that agents will only disclose if they find this to be appropriate.

Figures 2 and 3 show, for each number of DRs, the average intimacy achieved by the agents and the average privacy loss of the agents for each number of MTs considered. We can observe that, in general, all of the agents, regardless their privacy attitude, achieved less intimacy as the number of MTs increased. This is because as the number of MTs increases there are more target agents that do not reciprocate or do so with very unreliable information. Moreover, for a moderate number of MTs (that varies depending on the number of DRs), pragmatists achieved more intimacy than unconcerned. This is because pragmatists choose to interact with the most reliable and reciprocating agents, while unconcerned agents



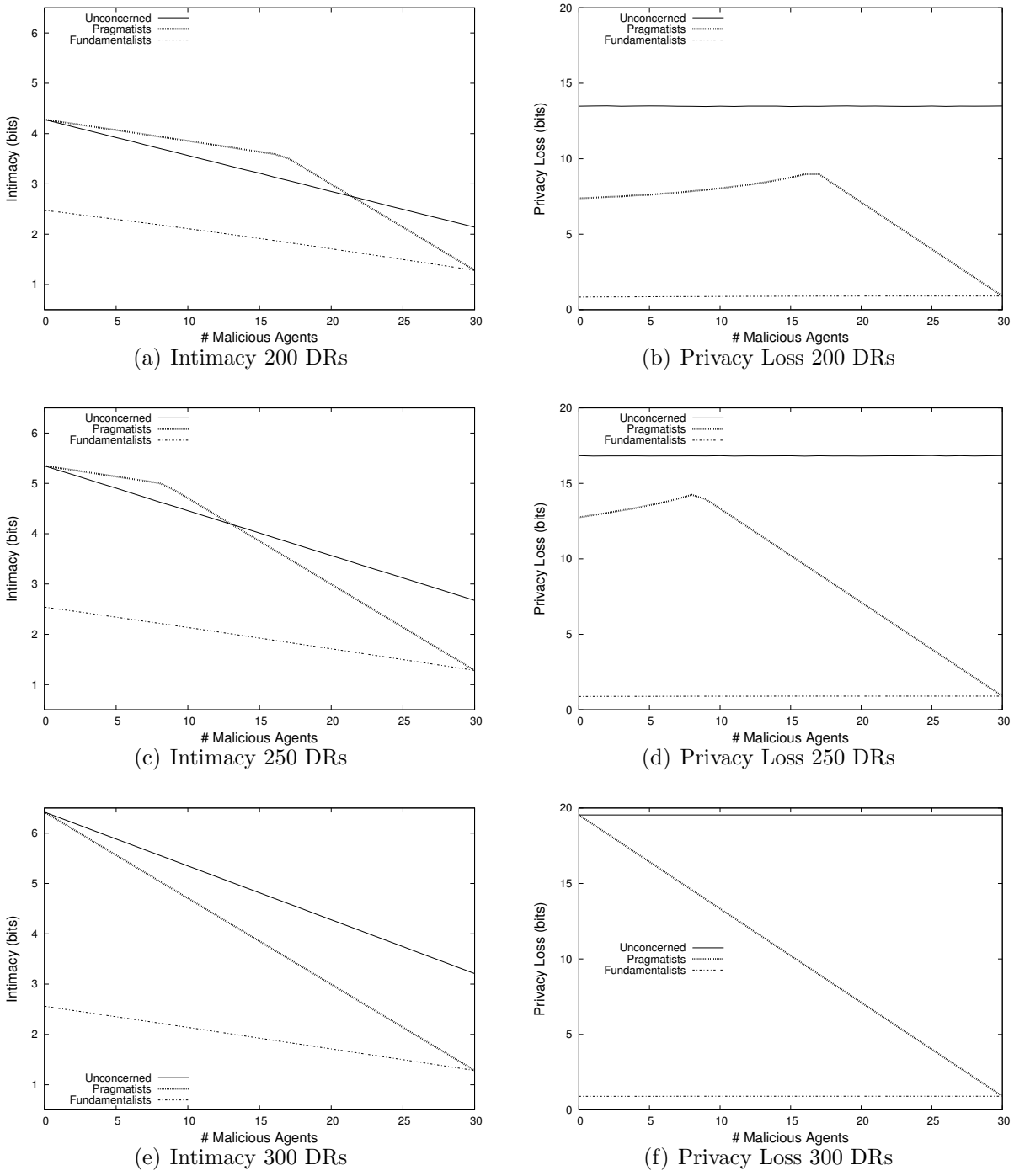
(a) Intimacy 50 DRs (b) Privacy Loss 50 DRs

(c) Intimacy 100 DRs (d) Privacy Loss 100 DRs

(e) Intimacy 150 DRs (f) Privacy Loss 150 DRs

Figure 2: Results considering malicious target agents (50, 100, and 150 DRs).

are not concerned about privacy and do not expect their disclosures to be reciprocated. Moreover, as in the previous experiments (Section 6.1), both unconcerned and pragmatist agents achieved more intimacy with other target agents than fundamentalists (except for 50



(a) Intimacy 200 DRs (b) Privacy Loss 200 DRs

(c) Intimacy 250 DRs (d) Privacy Loss 250 DRs

(e) Intimacy 300 DRs (f) Privacy Loss 300 DRs

Figure 3: Results considering malicious target agents (200, 250, and 300 DRs).

DRs). This is because when fundamentalists reach their maximum privacy loss ξ, they will not disclose PDAs so that intimacy is no longer increased.

As expected, pragmatist and fundamentalist agents always lost much less privacy than

unconcerned agents for all numbers of MTs and DRs. Moreover, unconcerned agents lost the same privacy for the same number of DRs regardless the number of MTs, because they always disclose one PDA to one agent for each DR without considering the privacy loss this may cause. Fundamentalists lost less privacy than pragmatists and unconcerned agents. Moreover, fundamentalists lost the same privacy, regardless the number of MTs. This is because when fundamentalists achieve their maximum privacy loss ξ, they will no longer disclose PDAs, regardless whether or not targets are being malicious or reliable and reciprocating.

We now detail the results that we obtained for each number of DRs considered. Figure 2(a) shows that for few DRs (50 DRs), all of the agents achieved the same intimacy regardless their privacy attitude. This is because at this stage agents may have performed very few disclosures to the same target agent (recall that there are 30 target agents). However, pragmatic and fundamentalist agents lost less privacy than unconcerned, as shown in Figure 2(b).

For 100 DRs, we obtained that fundamentalists start being unable to achieve the same intimacy as unconcerned and pragmatists (Figure 2(c)). This is because they will no longer disclose when they reach their fundamentalist threshold ξ. However, due to this threshold, we can see in Figure 2(d) that fundamentalists lose less privacy than pragmatists and unconcerned, except for a high number of MTs in which fundamentalists and pragmatist lose the same privacy. Pragmatists are aware that most of the agents to which they are interacting are malicious, so that they end up not disclosing to them, and thus, losing less privacy. Moreover, pragmatists are able to achieve the same intimacy as unconcerned but losing less privacy.

Figure 2(e) shows the average intimacy achieved by the agents for each number of MTs considered for 150 DRs. As can be observed, pragmatists are able to achieve greater intimacy than unconcerned agents for 1 up to 28 MTs (from 3.3% up to 93.3% MTs). This is due to the fact that pragmatists choose to interact with the most reliable and reciprocating agents, while unconcerned agents are not concerned about privacy and do not expect their disclosures to be reciprocated. From 28 MTs on, pragmatists achieved less intimacy than unconcerned agents because pragmatists will not disclose PDAs to MTs and there are not enough reliable and reciprocating agents in the system to achieve more intimacy. We can also see the same pattern for 200 DRs (Figure 3(a)) and 250 DRs (Figure 3(c)). In this case, the higher the number of DRs, the sooner pragmatists start achieving less intimacy than unconcerned. Moreover, when the number of DRs is the maximum (300 DRs) pragmatists achieved less intimacy than unconcerned from 1 MT, as shown in Figure 3(e). This is because in 300 DRs unconcerned always disclosed all of their attributes to all of the target agents while pragmatists may have not disclosed all their attributes if they realized that they are interacting with MTs.

Figure 2(f) shows the average privacy loss of the agents for each number of MTs considered for 150 DRs. For 0 up to 25 MTs, pragmatists lost an slightly increasing amount of privacy. This is due to the fact that as the number of MTs increases, pragmatists have more difficulties in finding agents that reciprocate their disclosures. Therefore, they lose a little (and slightly increasing) amount of privacy while seeking reliable and reciprocating

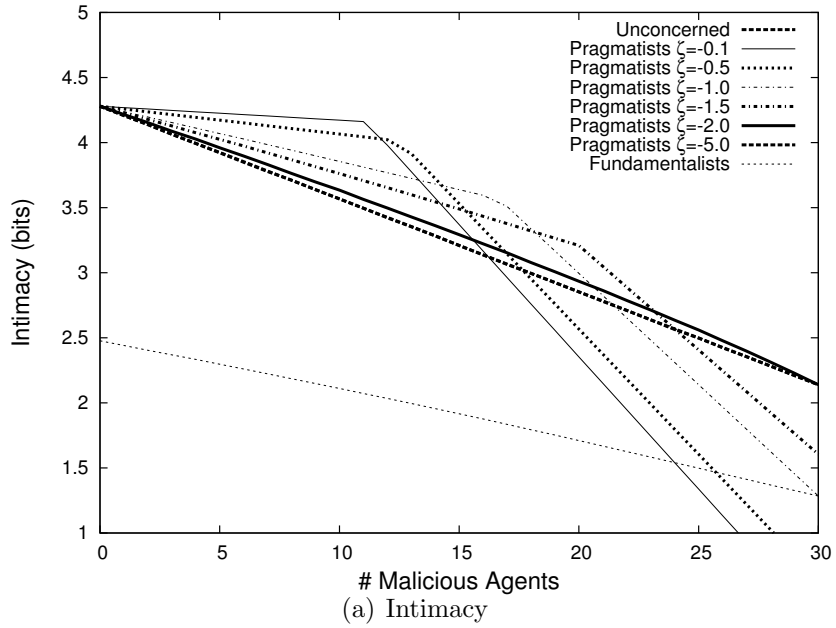
agents to whom concentrate their disclosures. For MTs from 25 up to 30, pragmatists lost less privacy as the number of MTs increased. This is because once pragmatists discover that an agent is malicious, they no longer disclose PDAs. As the number of MTs increases the number of total PDAs disclosed decreases so that privacy loss also decreases. We can also observe the same pattern for 200 DRs (Figure 3(b)) and 250 DRs (Figure 3(d)). The main difference is that privacy loss starts decreasing earlier as the number of DRs increases. This is because, as there are more DRs pragmatists need more target agents that reciprocate their disclosures. Therefore, pragmatists have more chances to interact with a MT. Moreover for 300 DRs (Figure 3(f)), pragmatists have a privacy loss that decreases linearly with the number of MTs. For 300 DRs, pragmatists disclose all of their attributes to all of the target agents when all the target agents are sincere and reciprocating (MTs=0). Thus, they end up losing all of their privacy. However, as the number of MTs increases they start losing less privacy because they can detect malicious agents, and thus, they do not perform disclosures to them.

6.3. Reciprocity Threshold

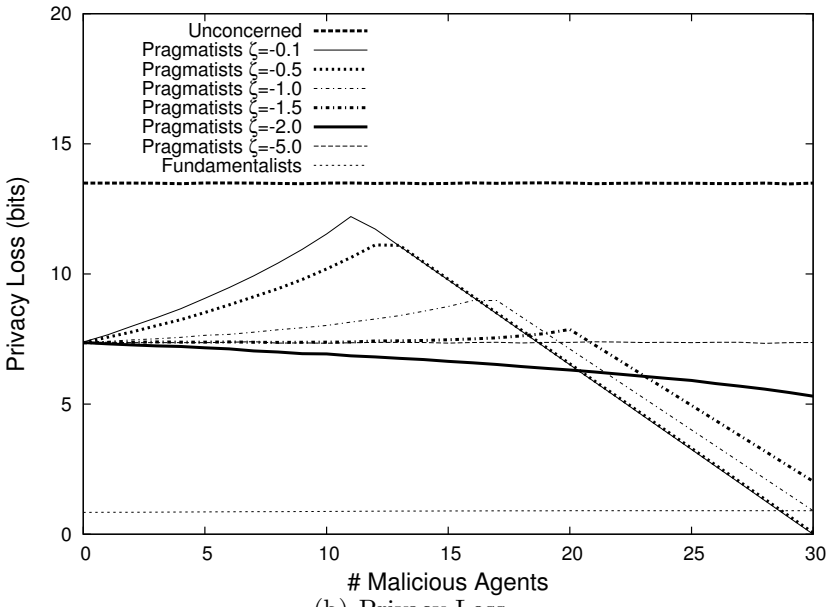
In this section, we detail the experiments that we performed to ascertain to what extent the reciprocity threshold ζ influences the behavior of pragmatic agents. To this aim, we repeated the experiment detailed in the previous section so that each time the reciprocity threshold has the value of -0.1, -0.5, -1, -1.5, -2, and -5 bits. Intuitively, we sought to consider pragmatic agents that expect almost the same information gain that they provide to others ($\zeta = -0.1$), pragmatic agents that expect reciprocations but are more permissive than the first ones ($\zeta = \{-0.5, -1.0, -1.5, -2.0\}$), and pragmatic agents that seek to be reciprocated but are less concerned with the possibility of being interacting with malicious agents that may not reciprocate them ($\zeta = -5.0$). The rest of parameters are as in Table 1.

For the sake of the clarity and appropriate visibility of the figures depicted in this section, we only show the results obtained for 200 DRs but similar results were also obtained for other number of DRs. Figure 4(a) shows the intimacy achieved. As one can observe, the reciprocity threshold ζ clearly influences the intimacy achieved by agents. The lower the reciprocity threshold, the higher the maximum intimacy that pragmatic agents achieved. For instance, the maximum intimacy that pragmatic agents achieved is when $\zeta = -0.1$, since there are few MTs and pragmatic agents only interact with reliable and reciprocating agents. However, the lower the reciprocity threshold, the sooner the change in tendency starts, i.e., intimacy starts decreasing as the number of MTs increases. This is because it is more difficult for pragmatic agents to find enough target agents that reliably reciprocate them as they expect (i.e., the reciprocity threshold). For instance, when $\zeta = -0.1$ pragmatic agents start being unable to achieve the same intimacy from ≈ 12 MTs on.

We can also observe in Figure 4(a) that the higher the reciprocity threshold, the less the difference between pragmatic and unconcerned agents. Moreover, when $\zeta = -5$ we can see that pragmatic agents achieve exactly the same intimacy as unconcerned agents. This is due to the fact that pragmatic agents do not see MTs as actually malicious until reaching the reciprocity threshold. Thus, if this threshold is high enough (e.g., $\zeta = -5$),



(a) Intimacy



(b) Privacy Loss

Figure 4: Results considering a varying reciprocity threshold ζ per number of malicious target agents.

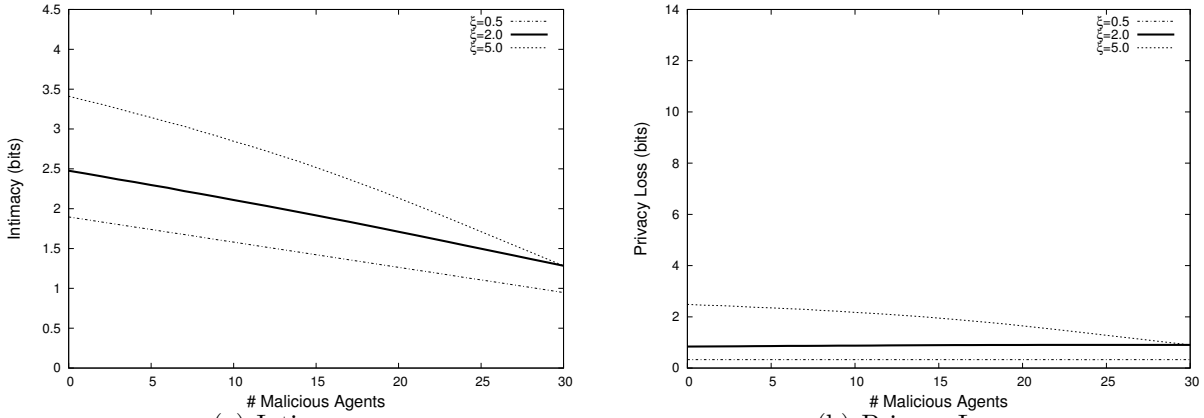
pragmatist agents disclose because they are more permissive with the amount and reliability of reciprocations.

Figure 4(b) shows the average privacy loss of agents when considering a varying reciprocity threshold. The higher the reciprocity threshold the less variable is the privacy loss of pragmatic agents. Moreover, for $\zeta = -5$ we can see that the privacy loss is constant. This is due to the fact that pragmatists disclose expecting very few reciprocations. Thus,

they end up performing the same disclosures regardless the number of MTs. However, we can observe that this privacy loss is still lower than the privacy loss of unconcerned agents because pragmatists choose to disclose attributes that minimize privacy loss.

6.4. Fundamentalist Threshold

We also sought to ascertain how the fundamentalist threshold ξ influences the behavior of fundamentalist agents. To this aim, we repeated the experiment detailed in Section 6.2, but varying the fundamentalist threshold. Specifically, we considered $\xi = \{0.5, 2.0, 5.0\}$.



(a) Intimacy (b) Privacy Loss

Figure 5: Results for fundamentalists per fundamentalist threshold ξ and per number of malicious target agents.

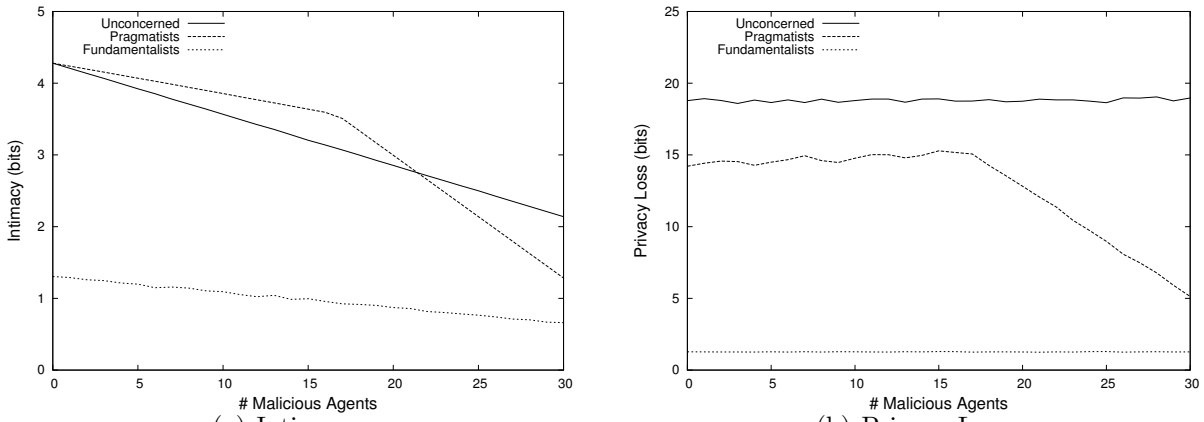
Figure 5(a) shows the average intimacy achieved per fundamentalist threshold with 200 DRs (similar results were obtained for other number of DRs). As one could expect, the higher the fundamentalist threshold ξ, the higher the intimacy that fundamentalist agents achieved. Moreover, as depicted in Figure 5(b) — that shows the average privacy loss per fundamentalist threshold — the higher the fundamentalist threshold ξ, the higher the privacy loss of fundamentalist agents.

We can also observe in figures Figure 5(a) and 5(b) that when $\xi = 5.0$ and the number of MTs is higher than 10, the intimacy achieved by and the privacy loss of fundamentalist agents starts decreasing at a faster rate. This is because if the fundamentalist threshold is high enough (e.g., $\xi = 5.0$), fundamentalist agents start behaving like pragmatic agents, i.e., they disclose attributes as long as they are reciprocated. Thus, for a high number of MTs, they had less intimacy and privacy loss because they chose not to disclose attributes when they realized that they were interacting with a MT.

6.5. Sensitivity Array per Agent

In our previous experiments we considered a unique randomly-generated sensitivity array for all of the agents. In this section, we illustrate the behavior of the agents when the sensitivity array is randomly generated for each agent, i.e., each agent has its own randomly

generated sensitivity array. This corresponds to a more realistic scenario, because humans usually have different sensitivities for the same PDA, as stated in the related literature on privacy such as in [7]. The other parameters used in this experiment are as in Table 1, and the number of DRs is 200.



(a) Intimacy (b) Privacy Loss

Figure 6: Results considering a random sensitivity array per agent.

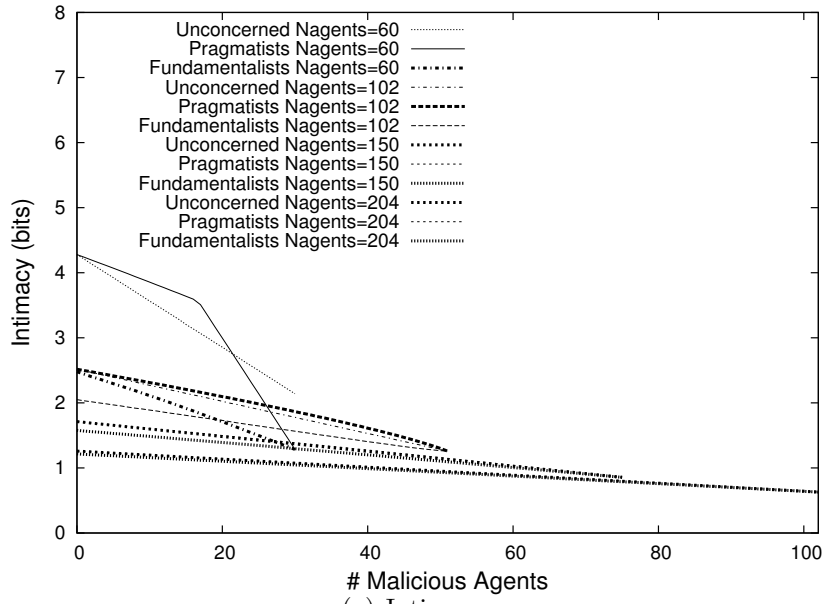
Figures 6(a) and 6(b) show the intimacy achieved by and the privacy loss of unconcerned, pragmatists, and fundamentalists when considering a random sensitivity array per agent. The results obtained are very similar to the ones obtained for the same parameters in Section 6.2 (i.e., Figures 3(a) and 3(b)) but with a unique and shared sensitivity array for all of the agents. This seems to mean that the behavior of our self-disclosure decision-making mechanisms is barely affected when each agent has a different sensitivity array — recall that the experiment was repeated 100 times as described in Table 1 so that each time a sensitivity array is randomly generated for each agent.

6.6. Varying the Number of Agents, Attributes and Values

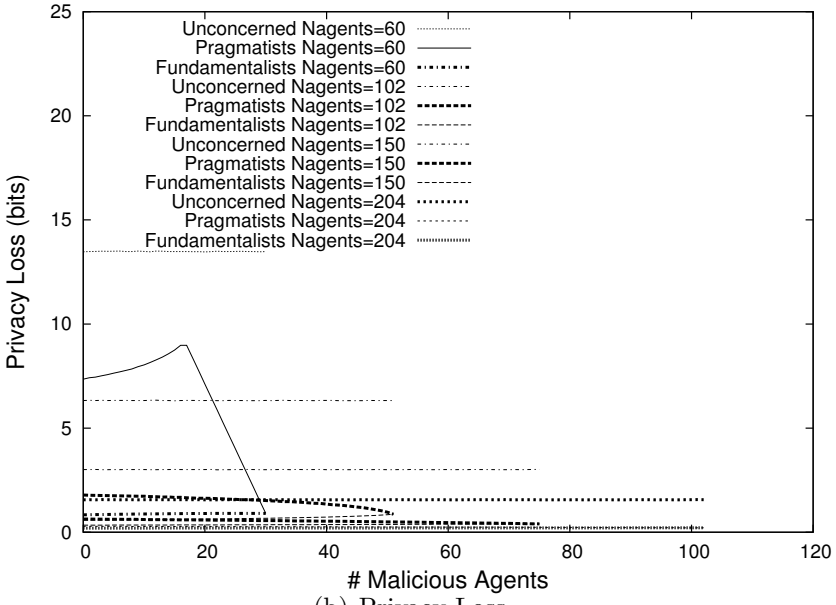
In this section, we consider a varying number of agents, attributes and values. The number of agents, attributes and values that we consider is finite and it is based on current social networking technologies. The number of agents is based on the number of people to which one could establish a particular degree of intimacy. For instance, average user in Facebook has 130 friends[11]. The number of attributes and values is based on the Friend of a Friend (FOAF) [31] ontology for connecting social Web sites and the people they describe. FOAF ontology considers user profiles involving less than 100 different terms to describe users. These profiles can also be seen as sets of attribute-value pairs. Moreover, the range of the terms in FOAF is usually finite and statically defined. Again, the rest of parameters used in this experiment are as in Table 1, and the number of DRs is 200.

Figures 7(a) and 7(b) show the results that we obtained when increasing the number of agents. When there are 60 agents, there are the same number of unconcerned, pragmatic,

[11] `http://www.facebook.com/press/info.php?statistics`



(a) Intimacy

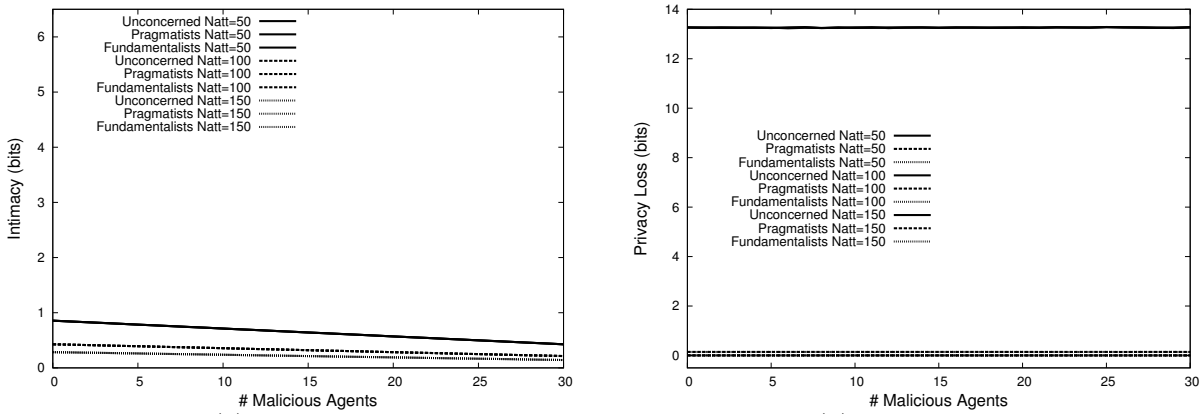


(b) Privacy Loss

Figure 7: Results obtained per number of agents.

fundamentalist, and target agents as in Table 1 (i.e., 10 unconcerned, 10 pragmatists, 10 fundamentalists, and 30 target agents). Moreover, we maintain the same ratio of unconcerned, pragmatic, fundamentalist, and target agents when increasing the number of agents. Note that for each number of agents, we have that half of this number of agents is the number of target agents in that experiment, which in turn is the maximum number of MTs.

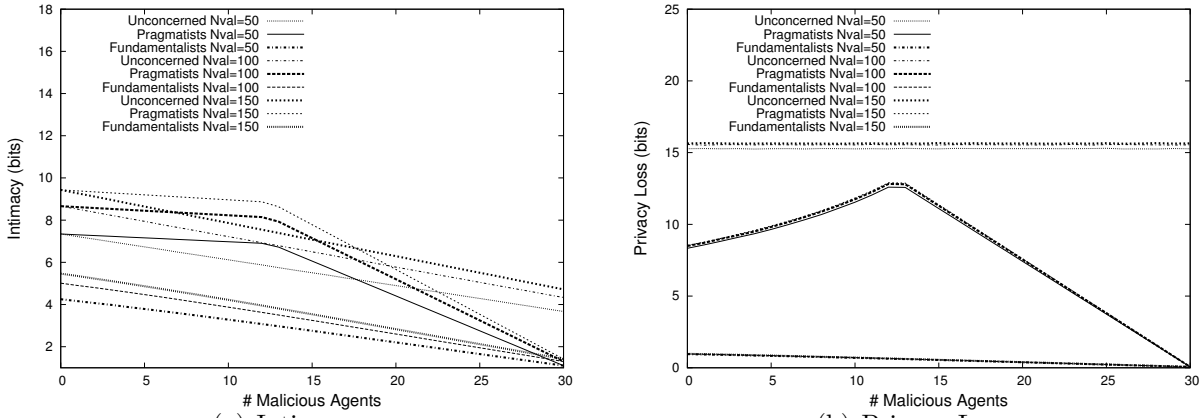
In both figures 7(a) and 7(b), the results for 102, 150, and 204 agents are very similar



(a) Intimacy (b) Privacy Loss

Figure 8: Results obtained per number of attributes.

to the results obtained for 60 agents and 150, 100, and 50 DRs respectively in Section 6.2. Thus, we can conclude that the behavior of the presented self-disclosure decision-making mechanisms when increasing the number of agents and leaving the number of DRs unchanged is very similar to when decreasing the number of DRs and leaving the number of agents unchanged.



(a) Intimacy (b) Privacy Loss

Figure 9: Results obtained per number of values per attribute.

Figures 8(a) and 8(b) show the results that we obtained when increasing the number of attributes up to 150 attributes. One can observe, that the behavior of our self-disclosure decision-making mechanisms is very similar to the behavior shown in Figures 2(a) and 2(b), i.e., the behavior is the same as when there are 10 attributes but very few DRs (i.e, 50 DRs). Thus, this behavior is also very similar to when increasing the number of agents as shown above.

Finally, figures 9(a) and 9(b) show the results that we obtained when increasing the number of values per attribute up to 150 possible values. We can observe that the number

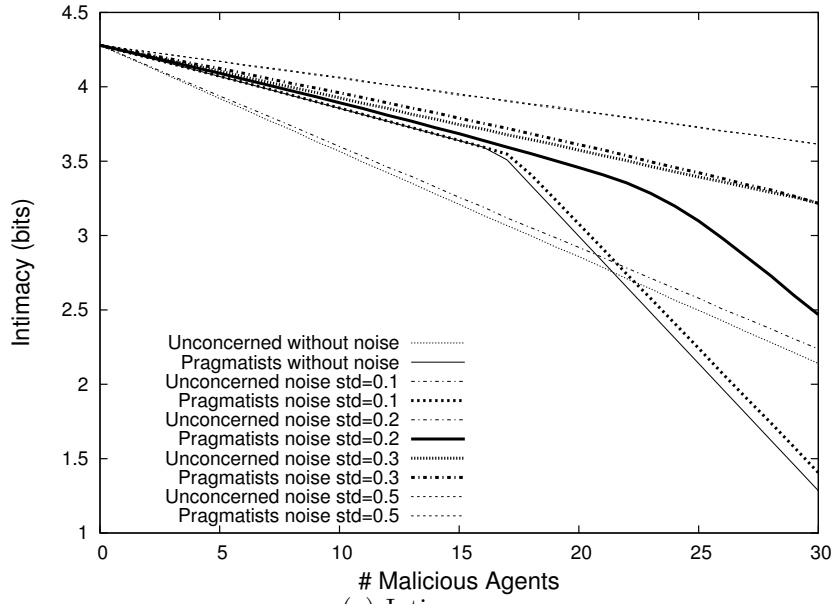
of values does not affect the overall behavior of agents regardless their privacy attitude. The only difference is that as the number of attributes increases, the absolute value for the intimacy achieved and the privacy lost also increases.

6.7. Uncertainty in Observations

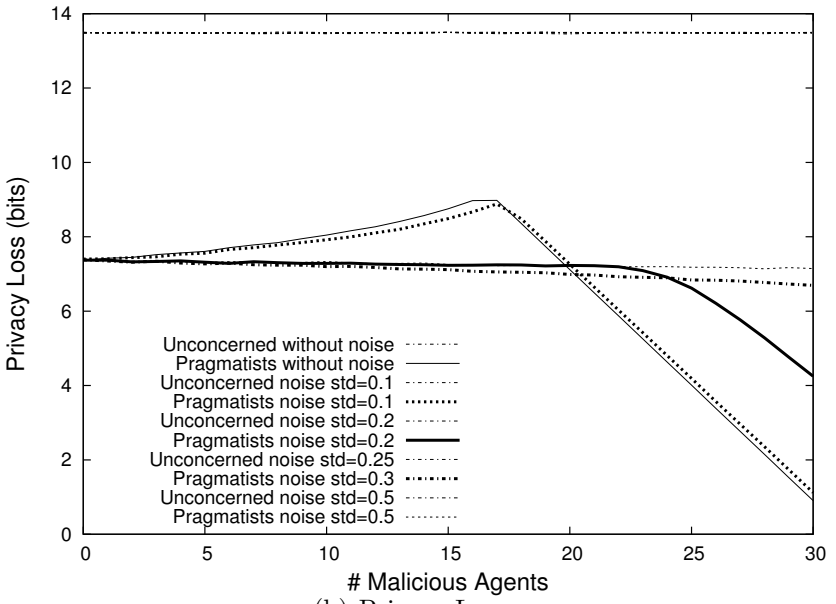
Over the course of this section, we have assumed that when an agent calls the `observe()` method, this method always returns the correct probability distribution for an attribute of another agent. However, this may not be the case in environments in which agents may not have the ability to always observe the correct probability distribution. We now assume that the `observe()` method introduces a random normally-distributed noise when it returns the probability distribution for an attribute of another agent. We considered a normally-distributed noise with mean 0.0 and a varying standard deviation. Specifically, we considered the following standard deviations: 0.1,0.2,0.3, and 0.5. The `observe()` method sums the generated noise to all of the components of the probability distribution to be returned, normalizes all of the components to still have a probability distribution (i.e., the addition of all of the components is equal to 1), and then returns the resulting probability distribution. The rest of the parameters are the ones detailed in Table 1, and the number of DRs is 200.

Figures 10(a) and 11(a) show the intimacy achieved by unconcerned, pragmatists, and fundamentalists per number of MTs. As one can observe, the higher the standard deviation of the random noise introduced, the higher the intimacy achieved by all of the agents regardless of their privacy attitude. This is due to the fact that agents think that their disclosures are reciprocated based on the observations that they perform. Thus, as these observations become less accurate, they are less able to realize that they are getting less reliable reciprocations. We can also observe, that for a standard deviation of 0.1 — this means that 99.7% of the random noise generated is distributed as follows: 68.2% of the random noise generated is in the interval $[-0.1, 0.1]$, 27.2% of the random noise generated is in the interval $[-0.2, -0.1[\bigcup]0.1, 0.2]$, and 4.2% of the random noise generated is in the interval $[-0.3, -0.2[\bigcup]0.2, 0.3]$ — the differences with respect to when the `observe()` method introduces no noise are very few. We can also see that for a standard deviation of 0.2, the overall behavior is very similar regardless of the privacy attitude of agents. However, from a standard deviation of 0.3 — this means that 99.7% of the random noise generated is distributed as follows: 68.2% of the random noise generated is in the interval $[-0.3, 0.3]$, 27.2% of the random noise generated is in the interval $[-0.6, -0.3[\bigcup]0.3, 0.6]$, and 4.2% of the random noise generated is in the interval $[-0.9, -0.6[\bigcup]0.6, 0.9]$ — we start observing differences in the behavior of pragmatic agents. Specifically, they hardly show the pattern of a faster intimacy decrease when there is a large amount of MTs, and their behavior is more similar to unconcerned agents. Moreover, for a standard deviation of 0.5, the behavior of pragmatists is the same as the behavior of unconcerned.

Figures 10(b) and 11(b) show the privacy loss of unconcerned, pragmatists, and fundamentalists per number of MTs. As one can easily observe, unconcerned and fundamentalists had the same privacy loss for all of the standard deviations tested. Unconcerned do not consider if they are being reciprocated or not, so they are not interested on performing observations about other agents. Fundamentalists had the same constant privacy loss. They



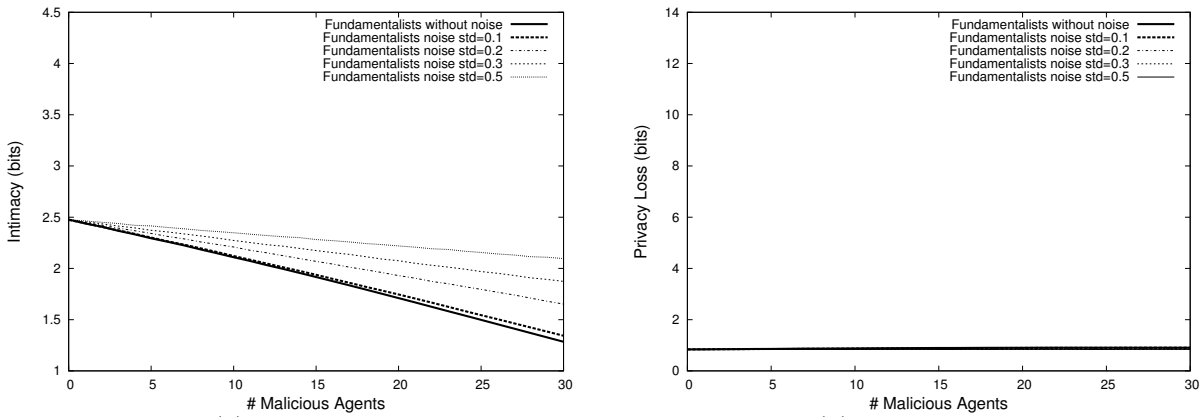
(a) Intimacy



(b) Privacy Loss

Figure 10: Results considering a random normally-distributed noise in observations (unconcerned and pragmatists).

used the same fundamentalist threshold as in Table 1, that in this case is low enough so that they did not disclose any more when they reached this threshold regardless the observations that they performed. We can also see that for standard deviations of 0.1 and 0.2, pragmatists have a very similar behavior to when the `observe()` method introduces no noise. However, the higher the standard deviation, the more constant the privacy loss



(a) Intimacy (b) Privacy Loss

Figure 11: Results considering a random normally-distributed noise in observations (fundamentalists).

becomes. This is because as pragmatists have more uncertain observations they have more difficulties to realize that they are interacting with MTs. However, we can see that even when the standard deviation is 0.5 (which is a very adverse scenario), pragmatists lost much less privacy than unconcerned.

Finally, to sum up, we conclude that for a moderated noise (std 0.1 and 0.2), our self-disclosure decision-making mechanisms behaved very similar to when there is no added noise. If this noise is very adverse (std 0.3 and 0.5) our self-disclosure decision-making mechanisms did not behave so similar, specially for the case of pragmatists. However, even in these very adverse conditions, pragmatists lost much less privacy than unconcerned.

7. Related Work

The privacy-utility tradeoff in online interactions has been studied in the last few years [10, 19, 14, 29]. This tradeoff considers the direct benefit of disclosing a PDA and the privacy loss it may cause. Agents disclose a PDA if the particular privacy-utility tradeoff for disclosing this PDA is acceptable. Given a set of PDAs A, the utility of disclosing these PDAs $U(A)$ and the privacy cost of disclosing these PDAs $C(A)$, the privacy-utility tradeoff is usually modeled as $A^* = \arg\max_A U(A) - C(A)$ [10]. For instance, in [10], the authors use the reduction in time for performing an online search if some PDAs such as geographical location are given as a utility function.

This tradeoff is often called privacy-trust tradeoff as well [21], [16]. An entity then is willing to disclose PDAs for increasing the trust others have in it. However, the authors of these works associate different levels of payoff for an entity for different trust levels others have in an entity. Then, a trust level is matched with a direct benefit so that an increase in trust results in an increase in utility. As a result, this can be finally modeled as a privacy-utility tradeoff.

Our proposed self-disclosure decision-making mechanisms differ from the works based on the privacy-utility tradeoff in two main aspects: (i) they consider repeated disclosures and

their implications, such as the *disclosure reciprocity* phenomenon [5]; and (ii) they are based on the privacy-intimacy tradeoff that can deal with situations where the direct benefit for disclosing a PDA is unknown.

The LOGIC negotiation model [24] describes relationships between a pair of negotiating agents using intimacy and balance measures based on information theory. In this paper, we adapt these two measures to deal with PDAs and present a privacy loss metric (privacy loss is not directly considered in LOGIC). Then, the two self-disclosure decision-making mechanisms that we propose are based on intimacy and balance on the one hand, and privacy loss on the other hand.

8. Conclusions

In this paper, we presented self-disclosure decision-making mechanisms based on information measures. These self-disclosure decision-making mechanisms model pragmatic and fundamentalist attitudes towards privacy by considering the increase in intimacy and the loss of privacy a disclosure may cause. Both intimacy and privacy loss are based on uncertain agent identities, a formalism that we presented to describe agents based on personal data attributes.

These self-disclosure decision-making mechanisms aim to be used in environments in which there can be repeated disclosures and in which disclosures may no report any benefit (or this benefit may not be known in advance). Thus, other already existing self-disclosure decision-making mechanisms, which do not consider repeated disclosures and that need that disclosures have an associated utility, cannot be used.

We experimentally showed that pragmatists lose less privacy than unconcerned agents for the same intimacy. We also showed that fundamentalists lose less privacy than both pragmatic and unconcerned agents but are unable to achieve the same intimacy. Moreover, in environments in which agents must interact with a moderate percent of malicious agents, pragmatists achieve even greater intimacy than unconcerned agents while losing less privacy. In environments in which agents must interact with a high percent of malicious agents, both pragmatists and fundamentalists lose much less privacy than unconcerned agents.

We also showed experimentally the properties of the self-disclosure decision-making mechanisms presented in this paper with respect to their main parameters and possible environmental conditions: the reciprocity and fundamentalist thresholds; the sensitivity array; the uncertainty in observations; and the number of agents, PDAs, and values.

As future work, we are exploring strategies for pragmatists and fundamentalists not to be sincere when disclosing a PDA. This could be useful once these agents detect that they are interacting with malicious agents. They could choose to keep on disclosing PDAs while being insincere instead of not disclosing any other PDA to such malicious agents. Thus, using such strategies agents would be able to lie to liars.

9. Acknowledgments

This work has been partially supported by CONSOLIDER-INGENIO 2010 under grant CSD2007-00022, projects TIN2011-27652-C03-00, TIN2010-16306, and ACE (CHIST-ERA

2011) of the Spanish Government, and the Generalitat of Catalunya grant 2009-SGR-1434. We would also like to acknowledge the anonymous reviewers for their very useful comments that have helped us to improve this paper.

References

[1] M.S. Ackerman, L.F. Cranor, J. Reagle, Privacy in e-commerce: examining user scenarios and privacy preferences, in: Proceedings of the 1st ACM conference on Electronic commerce, ACM, 1999, pp. 1–8.

[2] K. Deb, Multi-objective optimization, in: E.K. Burke, G. Kendall (Eds.), Search Methodologies, Springer US, 2005, pp. 273–316.

[3] M. Fasli, On agent technology for e-commerce: trust, security and legal issues, Knowledge Eng. Review 22 (2007) 3–35.

[4] A. Freitas, A critical review of multi-objective optimization in data mining: a position paper, ACM SIGKDD Explorations Newsletter 6 (2004) 77–86.

[5] K. Green, V.J. Derlega, A. Mathews, Self-disclosure in personal relationships, in: The Cambridge Handbook of Personal Relationships, Cambridge University Press, 2006, pp. 409–427.

[6] J. He, W. Chu, Z. Liu, Inferring privacy information from social networks, in: Intelligence and Security Informatics, volume 3975 of *Lecture Notes in Computer Science*, Springer-Verlag, 2006, pp. 154–165.

[7] B.A. Huberman, E. Adar, L.R. Fine, Valuating privacy, IEEE Security and Privacy 3 (2005) 22–25.

[8] A. Joinson, C. Paine, Self-disclosure, privacy and the internet, in: Oxford handbook of Internet psychology, Oxford University Press, 2007, pp. 237–252.

[9] G.J. Klir, Uncertainty and Information: Foundations of Generalized Information Theory, Wiley, 2006.

[10] A. Krause, E. Horvitz, A utility-theoretic approach to privacy and personalization, in: Proceedings of the 23rd national conference on Artificial intelligence, AAAI Press, 2008, pp. 1181–1188.

[11] S. Kullback, R.A. Leibler, On information and sufficiency, Annals of Mathematical Statistics 22 (1951) 49–86.

[12] P. Kumaraguru, L. Cranor, Privacy indexes: A survey of westin's studies, Technical Report CMU-ISRI-5-138, Carnegie Mellon University, School of Computer Science, Institute for Software Research International, 2005.

[13] T. Kwon, Privacy preservation with x.509 standard certificates, Information Sciences 181 (2011) 2906 – 2921.

[14] G. Lebanon, M. Scannapieco, M.R. Fouad, E. Bertino, Beyond k-anonymity: A decision theoretic framework for assessing privacy risk, in: In Privacy in Statistical Databases, Springer, 2006, pp. 217–232.

[15] N. Li, T. Li, t-closeness: Privacy beyond k-anonymity and l-diversity, in: In Proceedings of IEEE International Conference on Data Engineering, IEEE, 2007.

[16] L. Lilien, B. Bhargava, Privacy and trust in online interactions, in: Online Consumer Protection: Theories of Human Relativism, Information Science Reference, 2008, pp. 85–122.

[17] R. Miller, D. Perlman, S. Brehm, Intimate relationships, McGraw-Hill Higher Education, 2007.

[18] J.S. Olson, J. Grudin, E. Horvitz, A study of preferences for sharing and privacy, in: CHI '05 extended abstracts on Human factors in computing systems, ACM, 2005, pp. 1985–1988.

[19] S. van Otterloo, The value of privacy: optimal strategies for privacy minded agents, in: AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, ACM, 2005, pp. 1015–1022.

[20] K. Rannenberg, D. Royer, A. Deuker (Eds.), The Future of Identity in the Information Society: Challenges and Opportunities, Springer Publishing Company, Incorporated, 2009.

[21] J.M. Seigneur, C.D. Jensen, Trading privacy for trust, in: iTrust, Springer, 2004, pp. 93–107.

[22] C.E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (1948).

[23] C. Sierra, J. Debenham, Information-based agency, in: IJCAI'07: Proceedings of the 20th international joint conference on Artifical intelligence, Morgan Kaufmann Publishers Inc., 2007, pp. 1513–1518.

[24] C. Sierra, J. Debenham, The LOGIC negotiation model, in: AAMAS '07: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, ACM, 2007, pp. 1–8.
[25] J.M. Such, A. Espinosa, A. García-Fornes, A Survey of Privacy in Multi-agent Systems, Knowledge Engineering Review (2012) In Press.
[26] H. Taylor, Most People Are "Privacy Pragmatists" Who, While Concerned about Privacy, Will Sometimes Trade It Off for Other Benefits, Harris Interactive, 2003.
[27] A. Westin, Privacy and Freedom, New York Atheneum, 1967.
[28] A. Westin, Social and political dimensions of privacy, Journal of Social Issues 59 (2003) 431–453.
[29] A. Yassine, S. Shirmohammadi, Measuring users' privacy payoff using intelligent agents, in: Proc. of the IEEE Int. Conf. on Computational Intelligence for Measurement Systems and Applications (CIMSA), IEEE, 2009, pp. 169 –174.
[30] C.A. Yeung, I. Liccardi, K. Lu, O. Seneviratne, T. Berners-Lee, Decentralization: The future of online social networking, in: Workshop on the Future of Social Networking, W3C, 2009.
[31] L. Yu, L. Yu, Foaf: Friend of a friend, in: A Developers Guide to the Semantic Web, Springer Berlin Heidelberg, 2011, pp. 291–314.
[32] E. Zheleva, L. Getoor, To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles, in: WWW '09: Proceedings of the 18th international conference on World wide web, ACM, New York, NY, USA, 2009, pp. 531–540.