Value-driven Collaborative Privacy Decision Making

Francesca Mosca and **Jose M. Such** and **Peter McBurney**
Department of Informatics
King's College London

Abstract

Multiparty privacy conflicts (MPCs) occur when the privacy of a group of individuals is affected by the same piece of information, yet they have different (possibly conflicting) individual privacy preferences. One of the domains in which MPCs manifest strongly is online social networks, where the majority of users reported having suffered MPCs when sharing photos in which multiple users were depicted. Previous work on supporting users to make collaborative decisions to decide on the optimal sharing policy to prevent MPCs share one critical limitation: they lack transparency in terms of how the optimal sharing policy recommended was arrived at, which has the problem that users may not be able to comprehend why a particular sharing policy might be the best to prevent a MPC, potentially hindering adoption and decreasing the chance for users to accept or influence the recommendations. In this paper, we report our work in progress towards an AI-based model for collaborative privacy decision making that can justify its choices and allows users to influence them based on human values. In particular, the model considers both the individual privacy preferences of the users involved as well as their values to drive the negotiation process to arrive at an agreed sharing policy. We formally prove that the model we propose is correct, complete and that it terminates in finite time. We also provide an overview of the future directions in this line of research.

Introduction

Collaborative platforms like online social networks (OSNs), cloud-based file storage and sharing, cloud-based collaborative documents, and so on, proved particularly challenging for users to manage privacy and access to their data, especially when that data affects multiple users at the same time (Such and Criado 2018; Paci, Squicciarini, and Zannone 2018). Whenever documents regard multiple users, e.g. internal files of a company, pictures on social networks, or working sheets on online sharing platforms, the privacy settings and the sharing rights should be understood and agreed by all the users involved. If this does not happen, issues deriving from access control are likely to arise and no system has yet been implemented in order to detect them in advance or solve them efficiently once they occur (Such and Criado 2018; Paci, Squicciarini, and Zannone 2018).

A *multi-party privacy conflict* (MPC) is defined as a situation where more than one person is involved in some content (people appearing in a picture together, co-authors of a document, etc.) and they disagree on the level of sharing/privacy to be assigned to that content before sharing it online. In particular, MPCs on online social newtorks (OSNs) have received some attention in recent literature (Besmer and Richter Lipford 2010; Lampinen et al. 2011; Wisniewski, Lipford, and Wilson 2012; Acquisti, Brandimarte, and Loewenstein 2015; Liang et al. 2015; Acquisti et al. 2017; Such et al. 2017; Such and Criado 2018). Nowadays on most platforms it is possible only for the uploaders to define the privacy settings of photographic contents; if another person who is depicted in the picture (a co-owner of the picture) disagrees with such setting, there is no efficient support from the platform side (Wisniewski, Lipford, and Wilson 2012): the co-owners can flag the picture or contact the uploader to ask for its removal, but the damage may have already been caused, as the item is potentially visualised by many before any counteraction is made. So, reparative solutions are in general not enough (Such et al. 2017).

Pushed by the need of improving the support to the user provided by the collaborative platforms and taking advantage of the knowledge gathered by previous studies on usable privacy towards helping users making privacy choices online (Acquisti et al. 2017) and on empirical evidence about MPCs in practice (Wisniewski, Lipford, and Wilson 2012; Such et al. 2017), our main contribution is the definition of a model that (i) invites the users to interact between themselves in accordance with their values, in order to select the best collective sharing policy to prevent and/or resolve a MPC, and (ii) is able to justify its suggestion based on the values of those users involved, helping users understand the properties of the suggested action and the limitations of the discarded ones. In particular, in this paper we propose our theoretical model, which follows knowledge-based AI techniques and the theory of basic values (Schwartz 2010) to support users while collaboratively deciding on a common sharing policy to resolve an MPC. According to the typology of dialogues of (Walton and Krabbe 1995), our model is a negotiation dialogue, because the participants have potentially conflicting goals, which may not be satisfied si-

multaneously (McBurney and Parsons 2009). Therefore, we particularly based on AI-techniques like negotiation frameworks (Fatima, Kraus, and Wooldridge 2014), which have already been applied to other privacy domains (Such 2017).

Related Work

We refer the reader to (Such and Criado 2018) for a comprehensive review of literature on MPCs in social media and to (Paci, Squicciarini, and Zannone 2018) for a more general survey on access control in collaborative systems. As highlighted in these works, scholars recently suggested solutions to MPCs mainly following four different approaches.

First of all, game-theoretic approaches model users as rational entities and suggest the best option according to some utility function users aim to maximise (Squicciarini, Shehab, and Paci 2009; Such and Rovatsos 2016; Rajtmajer et al. 2017). While these proposals provided elegant frameworks from a formal point of view and build upon well-studied analytic tools, they may not work well when used in practice (Such and Criado 2018). This is because users' behaviour does not seem perfectly rational in practice, as assumed in these approaches (Lampinen et al. 2011; Wisniewski, Lipford, and Wilson 2012), and very few other factors are considered (Rajtmajer et al. 2017).

The second type of approaches to support users when solving MPCs base on recommendation engines, aiming to suggest the best common sharing policy based on the history of each user or of the platform (Fogues et al. 2017a; 2017b). However, it is not trivial to build a recommendation model for MPCs with high accuracy and high flexibility, in order to satisfy the user in a very dynamic environment; also these systems are not transparent and self-explaining, making it difficult for the user to understand the reason for which they should take a particular decision.

Another front of research proposes the use of AI techniques developed from the multi-agent systems community, from sets of norms (Criado and Such 2015), to argumentation approaches (Kökciyan, Yaglikci, and Yolum 2017) or to negotiation frameworks (Such and Criado 2016; Such and Rovatsos 2016). These approaches aim to support users while deciding the most socially preferred sharing action. Yet, the mechanisms may be difficult to comprehend and/or too rigid, often demanding too much effort from the user (Such and Criado 2018; Acquisti et al. 2017).

The fourth and final type of approach uses data processing (such as image processing) (Ilia et al. 2015; Vishwamitra et al. 2017) to tackle MPCs in social networks. User can manipulate the images before sharing them online by, for example, blurring the faces of the depicted people who do not concede access to that item to third parties. In this case, the main drawback is that, even without guaranteeing the protection of the user's privacy (someone can still be identified by other details than the face), there is a significant loss of sharing utility, as blurred pictures may not be as enjoyable.

The first three types of approaches focus on supporting the users involved in finding and agreement on a common sharing policy for the data in different ways, while the fourth type focuses on enforcing individual privacy policies. Both approaches are complementary. For instance, if there is an agreement possible between the users involved, then this has the potential to be more satisfying both in terms of privacy but also in terms of loss of sharing utility, and empirical evidence tells us that many of the MPCs could indeed be solved in practice (Such et al. 2017; Such and Criado 2018). If, however, an agreement is not found or is not possible, then something like blurring pictures may provide a last resort baseline privacy.

In this paper, we focus on supporting users to find an agreement. The model we propose differs from the three approaches mentioned above that also focus on supporting users reaching an agreement, as it focuses on transparency and on suggesting actions closer to users' motivations and values, both lacking in previous literature as detailed in the paragraphs above. In particular, the model follows previous works on usable privacy, which show the importance of including in the models information to guide users towards safer and better choices, without imposing a particular decision (Acquisti et al. 2017). We hypothesise that a valid contribution to the solution of MPCs is to include in the model a self-explaining component, that makes it easier for the user to understand and endorse the model's recommendation. The model produces suggestions along the negotiation that promote or demote users' values, like a Jiminy Cricket from *Pinocchio*, to help users pick the actions that are most aligned with their values.

The Model

Here we describe in detail the model to support the users while reaching a collaborative decision on a common sharing policy for a given piece of data that is co-owned by them or that affects their privacy (e.g. a photo in which multiple users are depicted). The model defines a negotiation framework where each user is driven through the dialogue by the values she considers important. The model takes as input from every user involved in the decision (e.g. those depicted in a photo) what would be their preferred sharing policy for the item to be shared and their value order, defining the relative importance each basic value has for them (both - policy and value ordering - can be elicited with minimal user intervention as detailed below). The model supports all sides of the negotiation suggesting the action which best suits each user's preferences and value order, and it provides feedback on how other possible options impact on the promotion/demotion of those same values, leaving the user the last word of the decision.

Individual Sharing Policies

We represent a collaborative platform (e.g. an OSN) where agents are supposed to interact with each other and share contents as a social graph $G = (U, E, I)$, where the nodes $u \in U$ are the users, the edges $e_{ij} = (u_i, u_j) \in E$ are the links between the users, and each edge has a weight representing the closeness or intimacy of the relationship between the two connected users; such intimacy i takes values in $\{1, ..., N_i\} \subset \mathbb{N}$, with N_i being the maximum, and can be elicited by using predictive techniques as presented in (Fogués et al. 2014).

Content is shared in the platform according to sharing policies, which define the criteria users must satisfy in order to access such content.

Definition 1. A *sharing policy* is a tuple $p = \langle d, i \rangle \in \mathcal{P}$, where $d \in [0, N_d]$ represents the maximum allowed *distance* a user must be from the owner of the content, meant as the length of the shortest path of the social graph connecting the two users, and $i \in [0, N_i]$ represents the minimum required *intimacy* over each edge of the path connecting the two users.

Note that the policy definition used in this paper could be translated to and back from the usual group-based access control policies of social media sites like Facebook (Such and Criado 2016). Also, note that we assume that individual privacy preferences for each item, i.e. the sharing policy p each of the users involved would prefer if they were to decide about the item alone, are provided directly by the user or they are elicited, to minimise user effort, following data-driven AI techniques like machine learning shown to work very well to elicit individual privacy preferences in social media while minimising user effort such as (Squicciarini et al. 2011; Misra and Such 2017), or derived from suitable defaults based on approaches like (Watson, Lipford, and Besmer 2015).

Schwartz Basic Values

We base on the *theory of basic values* by Schwartz (Schwartz 2010) to model human values, as this is the most well-known and established theory of values with strong empirical evidence backing the theory, validated in many studies and over different countries (Bilsky, Janik, and Schwartz 2011), and which has been successfully applied in various contexts, including environmentalism (Stern, Dietz, and Guagnano 1998), recruitment (Patterson et al. 2016), consumers habits (Thøgersen, Zhou, and Huang 2016), healthcare (Moyo et al. 2016), and many others.

Values are defined by Schwartz as socially desirable concepts that allow humans to interact between themselves, representing mental goals and the way used to describe and communicate such goals (Schwartz 2012). People take daily decisions according to the values they believe into. Values compete with each other, but the individual realises the dissonance and decides what to do by giving priority to some values over the others.

In our context, we identify five relevant categories of basic values from Schwartz' theory:

- *self-direction (sd):* the user is open-minded and ready to change the negotiating strategy during the decision making process to suggest new solutions;
- *power (po):* the user holds her initial idea, giving no space to accommodate the others' preferences;
- *security (se):* the user prefers the safer option, in this case the most restrictive one in terms of publicity;
- *conformity, tradition (tr):* the user's choice is highly influenced by the society's expectations;

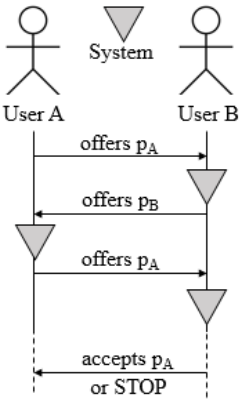

Figure 1: Message Sequence Chart showing the negotiation process between the uploader A and the co-owner B.

- *benevolence, universalism (be):* the user is willing to consider any proposal that is coming from a close friend (benevolence) or anyone (universalism).

We formalise the relative importance of the values for each user as follows.

Definition 2. A *value order* is a particular order $o \in \mathcal{O}$ over the values $\mathcal{V} = \{be, po, se, sd, tr\}$ that determines the relative relevance the user believes each value has on her behaviour, where $\mathcal{O}$ is the space of all the possible total or partial orders over $\mathcal{V}$.

Example: A user α has a total order over her values o_α : $po \succ se \succ sd \succ tr \succ be$, meaning that she considers power as the most important value to guide her behaviour, followed by security as the second most influential and so on.

Note that, in practice, the importance that each user attributes to each of these values can be elicited using the Schwartz Value Survey or the Portrait Values Questionnaire validated by Schwartz (Schwartz 2012). Also note that, while the example shows a total order, our model would also work with partial orders, e.g., users having some preferences of values over the others but not for all values.

Negotiation process

The communication between the users involved follows a number of steps (or negotiation rounds) until a decision is agreed. For simplicity and without loss of generality, we will consider in this paper just two users and an alternated proposals negotiation framework (Fatima, Kraus, and Wooldridge 2014), depicted in the Figure 1. The uploader A starts the dialogue offering a certain policy p_α to the co-owner B. B evaluates the received offer: if B accepts, then the negotiation is over and the content is shared with policy p_α; if B is not satisfied by the offer, then B can make a counter-proposal, which is evaluated by A and so on. However, there might be no possibility of reaching an agreement, for instance if both parties keep offering the same policy without trying to accommodate each other. In this case, the negotiation is considered as failed and no content is shared.

In order to support the user while taking decisions during the negotiating process, at every iteration the model suggests to each user the action that is the most coherent one with the ordering of their values, as detailed in the next section in Definition 6. As it is important to keep the user in the loop and in full control if the user desires so, as otherwise some users will feel out of control of their privacy (Vihavainen et al. 2014), the model can also produce other options together with how these other options promote/demote particular values. After a user suggests a policy, the other user has to decide if she accepts it. The condition for accepting depends on the distance between the received policy proposal and the one which would be suggested by the other user in case of rejection. Again, this may be left as a default distance value or to be decided by the user.

Definition 3. The *distance* $\epsilon_{\alpha\beta}$ between two policies p_α and p_β is defined as the Manhattan distance:

$$\epsilon_{\alpha\beta} = |d_\alpha - d_\beta| + |i_\alpha - i_\beta|,$$

where $p_\alpha = \langle d_\alpha, i_\alpha \rangle$ and $p_\beta = \langle d_\beta, i_\beta \rangle$ are respectively the preferred policy by users A and B.

If such distance is within a reasonable range, for instance it is equal to 0 so that the policies are the same or equivalent, then the offer is accepted and the negotiation ends. Otherwise, the other user communicates her counter-proposal and the dialogue proceeds until either the convergence is reached or until it is recognised as impossible, i.e. when both users have tried out all their strategies and cannot help but keep suggesting the same policies. If an agreement is found, then the item in discussion is shared according to the last offered policy. Otherwise, the content remains private.

Generation of a new policy proposal

We now introduce some definitions that help us understand the process through which the model suggests to the user the policy most coherent with her values at each negotiation round.

Definition 4. A *value-function* $f_v : \mathcal{P}^3 \longrightarrow \mathcal{P}^3$ defines the cumulative effect of a value on the policy to be suggested. It takes as inputs the preferred policies p_α, p_β of the two users, and p_v, a policy which memorises the outcome of the other value-functions previously combined. The codomain of f_v is a subset of its domain: i.e. given three policies p_α, p_β, p_v and defining

$$p_1 = \langle \min(d_\alpha, d_\beta, d_v), \max(i_\alpha, i_\beta, i_v) \rangle = \langle d_1, i_1 \rangle$$
$$p_2 = \langle \max(d_\alpha, d_\beta, d_v), \min(i_\alpha, i_\beta, i_v) \rangle = \langle d_2, i_2 \rangle$$

as the tuples having respectively the most and the least restrictive components over p_α, p_β, p_v, then

$$f_v(p_\alpha, p_\beta, p_v) \in ([d_1, d_2] \times [i_2, i_1])^3. \quad (1)$$

There are four f_v in the space of the value-functions $\mathcal{F} = \{f_{be}, f_{po}, f_{se}, f_{tr}\}$ representing respectively the influence of benevolence, power, security and tradition on the policy offer. A possible instance of the f_v functions is shown in Table 1. For the case of the self-direction value, we do not model it as a value function but instead we model it as a different function that changes the particular value ordering, as whenever the user believes in self-direction, the importance of being creative and open-minded while researching a solution is what matters (Schwartz 2010).

Definition 5. The *sd-function* $f_{sd} : \mathcal{O} \longrightarrow \mathcal{O}$ defines the influence of the value *self-direction* over the entire user's strategy. Considering the order provided by the user, it returns another order where the value self-direction is removed and the two following values, if any, are swapped at every other iteration of the negotiation.

The sd-function allows the user to be flexible in her negotiation process by employing eventually two strategies alternatively, i.e. two different orders over benevolence, power, security and tradition.

We are finally ready to introduce the crucial part of the model, that is the generation of a new proposal. At each step of the negotiation after the first one, a user receives an offer that needs to be evaluated. To do so, the system generates for the user what should be her next suggestion, given her preferred policy and order over values.

Definition 6. A *proposal generator* $g : \mathcal{P}^3 \longrightarrow \mathcal{P}$ is a function which provides the policy that should be suggested according to the particular value ordering and the initial policy preferences of the two users. In particular, it is a composition of value-functions $f_i \in \mathcal{F}$, where the order of the composition[1] is given by the order over values $o \in \mathcal{O}$, and the projection operator Π_3, which selects only the third policy from the last obtained tuple.

The proposal generator g computes for the user A what the next policy proposal $\tilde{p} = g(p_\alpha, p_\beta; o_\alpha)$ should be, so that it is the policy most consistent with the user's A order over values and the initial preferred sharing policies (an equivalent optimal policy is computed also for the user B, whenever it is her time to evaluate an offer from A). Then, according to the distance introduced in Definition 3, the user decides whether to accept the other user's proposal or to make a counter-offer, in which case the system suggests the last generated policy $\tilde{p}$.

Example Negotiation

Consider a situation where users A and B discuss about sharing some content on an online collaborative platform. Their preferred policies are respectively $p_\alpha = \langle 5, 8 \rangle$ and $p_\beta = \langle 1, 1 \rangle$. According to the sensitivity of the content, in general people would suggest the policy $p_\sigma = \langle 2, 6 \rangle$. A and B also provide to the system their order over the values:

$$se \succ_\alpha be \succ_\alpha tr \succ_\alpha po \succ_\alpha sd$$
$$be \succ_\beta tr \succ_\beta se \succ_\beta sd \succ_\beta po.$$

We consider the instances of the value-functions listed in Table 1, and that at the beginning of each step $p_v = null$. Here, both users have a single strategy, because *self-direction* is in

[1] In case the value order is a total order, then the composition order of the value functions is trivial (the Example Negotiation section). If the value order is a partial order, different solutions can be applied, for instance picking randomly only one of the equivalent value-functions.

f_v
$f_{po}(p_\alpha, p_\beta, p_v) = (p_\alpha, p_\beta, avg(p_\alpha, p_v))$
$f_{be}(p_\alpha, p_\beta, p_v) = (p_\alpha, p_\beta, avg(p_\beta, p_v))$
$f_{se}(p_\alpha, p_\beta, p_v) = (p_\alpha, p_\beta, avg(\langle \min_{p_\alpha, p_\beta} d, \max_{p_\alpha, p_\beta} i \rangle, p_v))$
$f_{tr}(p_\alpha, p_\beta, p_v) = (p_\alpha, p_\beta, avg(p_\sigma, p_v))$

Table 1: Possible instances of f_v functions. Note that p_σ refers to the policy a majority of people would select for content with the same sensitivity, which can be elicited as shown in (Fogues et al. 2017b). Also, note that since the distance and the intimacy must be integer numbers, rounding is performed towards the one own policy in general, and towards the other user's policy when the value-function is f_{be}.

the last or second-last position in the order and therefore it has no values to swap. In this situation, the condition to accept an offer is to reach the same common policy, so that the distance is $\epsilon = 0$ from the newly generated one.

At $t = 0$, A suggests the policy $p_0 = p_\alpha = \langle 5, 8 \rangle$.

At $t = 1$, B decides whether to accept or reject the offer. To do this, the system computes the best (according to B's values) policy that B would eventually counter-offer and then, if this coincides with p_0, the system would suggest to accept:

$$\begin{aligned} g(p_\alpha, p_\beta; o_\beta) =& \Pi_3 \circ f_{be} \circ f_{tr} \circ f_{se} \circ f_{po}(p_\alpha, p_\beta, \textit{null}) \\ =& \Pi_3 \circ f_{be}(f_{tr}(f_{se}(\langle 5, 8 \rangle, \langle 1, 1 \rangle, \langle 1, 1 \rangle))) \\ =& \Pi_3 \circ f_{be}(f_{tr}(\langle 5, 8 \rangle, \langle 1, 1 \rangle, \langle 1, 4 \rangle)) \\ =& \Pi_3 \circ f_{be}(\langle 5, 8 \rangle, \langle 1, 1 \rangle, \langle 1, 5 \rangle) \\ =& \Pi_3(\langle 5, 8 \rangle, \langle 1, 1 \rangle, \langle 3, 7 \rangle) \\ =& \langle 3, 7 \rangle. \end{aligned}$$

Since $\epsilon_{\alpha\beta} = |5 - 3| + |8 - 7| > 0$, the system suggests B to reject the offer and to propose $p_1 = \langle 3, 7 \rangle$. However, at every iteration, the decision about following the system's suggestion is left to the user. Future studies should look into how to convey to the users some interpretation and comments on each of the possible actions they might take (Acquisti et al. 2017), for instance which values might be promoted or demoted by suggesting another policy. For simplicity here we are describing a case when each user always performs the action which is suggested by the system as the most coherent one given her own values. So, B rejects p_0 and offers $p_1 = \langle 3, 7 \rangle$.

At $t = 2$, it's A's time to evaluate B's offer:

$$\begin{aligned} g(p_\alpha, p_1; o_\alpha) =& \Pi_3 \circ f_{se} \circ f_{be} \circ f_{tr} \circ f_{po}(p_\alpha, p_1, null) \\ =& \Pi_3 \circ f_{se}(f_{be}(f_{tr}(\langle 5, 8 \rangle, \langle 3, 7 \rangle, \langle 5, 8 \rangle))) \\ =& \Pi_3 \circ f_{se}(f_{be}(\langle 5, 8 \rangle, \langle 3, 7 \rangle, \langle 4, 7 \rangle)) \\ =& \Pi_3 \circ f_{se}(\langle 5, 8 \rangle, \langle 3, 7 \rangle, \langle 3, 7 \rangle) \\ =& \Pi_3(\langle 5, 8 \rangle, \langle 3, 7 \rangle, \langle 3, 8 \rangle) \\ =& \langle 3, 8 \rangle. \end{aligned}$$

Since $\epsilon_{\alpha\beta} = |3 - 3| + |8 - 7| > 0$, A rejects the offer and proposes $p_2 = \langle 3, 8 \rangle$.

At $t = 3$, B reasons about the last offer received from A:

$$\begin{aligned} g(p_2, p_1; o_\beta) =& \Pi_3 \circ f_{be} \circ f_{tr} \circ f_{se} \circ f_{po}(p_2, p_1, \textit{null}) \\ =& \Pi_3 \circ f_{be}(f_{tr}(f_{se}(\langle 3, 8 \rangle, \langle 3, 7 \rangle, \langle 3, 7 \rangle))) \\ =& \Pi_3 \circ f_{be}(f_{tr}(\langle 3, 8 \rangle, \langle 3, 7 \rangle, \langle 3, 7 \rangle)) \\ =& \Pi_3 \circ f_{be}(\langle 3, 8 \rangle, \langle 3, 7 \rangle, \langle 3, 7 \rangle) \\ =& \Pi_3(\langle 3, 8 \rangle, \langle 3, 7 \rangle, \langle 3, 8 \rangle) \\ =& \langle 3, 8 \rangle. \end{aligned}$$

Since B obtains the same policy that is last offered by A, B accepts and the content is shared with policy $p = \langle 3, 8 \rangle$.

Properties of the model

Given the practical problem we aim to solve, it is crucial for the model to present some properties that allow its implementation on real systems, like termination in a finite time, correctness and completeness.

Lemma 1. Termination On the assumption that neither party withdraws, in a finite time, the offers p_t suggested by the system always converge towards an agreement $\tilde{p}$

$$d(p_t, \tilde{p}) \to 0 \text{ for } t \to N < +\infty \tag{2}$$

or the system recognises the impossibility of reaching an agreement.

Proof. During the negotiation process, a user can either maintain her position or accommodate the other user's preference. Let us consider each case separately:

(a) Both users want to accommodate each other: let us hypothesise, by absurd, that (2) is false; this means that the distance between the new suggestion and the final agreement may increase at each iteration or that the convergence may happen in an infinite number of iterations. By the definitions of the functions f_v, the output of each f_v is always within the range defined by the most and the least restrictive tuples of each iteration (see Equation (1)). If the users are both trying to accommodate each other's preference, it means that the new suggestion is a tuple whose at least one element is internal to the domain; i.e. the domain of $f_v(p_\alpha, p_\beta, p_v)$ becomes one of the following:

$$\begin{array}{ll} [d_1, d_2) \times (i_2, i_1] & (d_1, d_2] \times (i_2, i_1] \\ [d_1, d_2) \times [i_2, i_1) & (d_1, d_2] \times [i_2, i_1) \\ (d_1, d_2) \times (i_2, i_1). & \end{array}$$

Noting that these are all subsets of $\mathbb{N}^2$, it follows that the width of the domain of the new suggestion decreases at every iteration. Therefore, the distance between the new suggestion and the final deal can only decrease, as they both are elements of the domain; this contradicts our initial hypothesis. Also, given the fact that the domain is a finite and bounded subset of $\mathbb{N}^2$, the convergence happens in a finite number of iterations. Therefore, (2) is valid.

(b) One user wants to accommodate, the other user holds her position: the reasoning is similar to the previous case, but now the contraction of the domain happens only at every other iteration, i.e. whenever a user makes an accommodating offer. In fact, when a user sticks to her preference,

she keeps offering a tuple whose elements are on the border of the domain. Eventually with a slower speed than in the previous case, the domain does get contracted and, given its finite dimension and its boundary, it converges in a finite time. So, (2) is valid.

(c) At some point, both users start holding their positions: if both users stick to their preferences, it means that both of them keep suggesting policies whose elements are on the border of the domain

$$f_v(p_\alpha, p_\beta, p_\sigma) \in [d_1, d_2] \times [i_2, i_1] \in \mathbb{N}^2;$$

therefore, the domain cannot contract. The system interrupts the negotiation whenever both users have tried all their strategies after receiving the same inputs. Since every user has at most two strategies, this happens at latest at the 5th iteration of the algorithm with no changes in the offers: that is, the termination of the algorithm is realised in a finite time. □

Lemma 2. Correctness Assuming that the users always follow the system's suggestions, the outcome of the negotiation is consistent with their initial preferences and orders over the basic value.

Proof. This can be proved by contradiction. Let us assume that the outcome of the negotiation is not consistent with the users' initial preferences over the sharing policies and the values, even though both users always followed the system's suggestions. This implies that, in at least one step of the negotiation process, the system provided a policy recommendation which was not consistent with the users' inputs. A new policy proposal is defined from the function g (see Definition 6) as the composition of, excluding the final projection, only f_v functions (see Equation (1)). If a new suggestion given by the function g is not consistent, it means that either (i) the composition order or (ii) the f_v functions are not consistent with the inputs. However, (i) the order for composing the f_v functions is defined exclusively by the order over values provided by the users: this assures that different relevance is given to different f_v functions according to the priority that the user assigns to the values that each f_v function represents. Therefore, the composition order is by definition consistent with the user's preference. On the other hand, (ii) the f_v functions are defined in such a way that they reflect the interpretation of each value in the negotiation context given the initial policy preferences; for instance, in Table 1 f_{po} reflects power and is influenced by whatever the output of the negotiation is up to that point (p_v) and by the last policy preference of the user (for the user A, p_α in the first step and then the last policy proposed by A). So, by definition, the value-functions are consistent with the initial preferences of the users. In both (i) and (ii) we reached a contradiction, therefore we can say that, given a coherent behaviour from the users' side, every step of the negotiation must be consistent with the initial preferences of the users. We know that every outcome of the model is reached through a sequence of such consistent steps; therefore every outcome whether it is a deal or no deal, must be consistent with the initial preferences of the users. □

Lemma 3. Completeness An optimal agreement is reached by a sequence of optimal negotiation steps; the outcome of a single step is optimal if it is coherent with the user's initial preferences. If an optimal agreement exists, then the system is able to find it.

Proof. An optimal agreement is the outcome of a chain of actions that are coherent with the user's initial preferences, i.e. when the user follows the system's suggestion instead of acting impulsively. Whenever the chain of coherent events is altered by an impulsive choice, it becomes impossible to explain or predict this outcome using the system's rules without introducing an inconsistency. Therefore the system is complete. □

Conclusion

In the past years, an increasing attention has been paid to MPCs on online collaborative platforms, especially on social networks. Reparative solutions are proved not to be efficient and many attempts have been performed by scholars to tackle this problem, while considering collaborative decision making strategies. However, the majority of the suggested solutions fail to provide simple and understandable reasons about the generation of their suggestions, sometimes hindering the user's endorsement of the recommended policy provided by the model. Following previous studies on usable security (Acquisti et al. 2017), we hypothesise that the transparency and the interpretability of a model can play a crucial role in solving MPCs. On this basis, we built a model to support the users while interacting with each other in order to collaboratively define access control based on values. Following the model's suggestions at every negotiation step, each user is recommended to act coherently with her values, and when an agreement is achieved, the agreement is guaranteed to be the most consistent with their initially preferred sharing policies and her values.

Despite the rigorous and formally proven definition of the model in this paper, this line of research is still work-in-progress. There are some particular directions we would like to follow. First, we would like to validate the model here presented and refine it with empirical evidence coming from user studies. Particularly, we would like to test the hypothesis that users are able to understand better the support both during negotiations and after an outcome is agreed and that a value-based approach leads to better user satisfaction when resolving MPCs. Before we can perform the user study, it is crucial that we define the best way to maximise the usability of any interfaces based on our model: this includes studying the best trade-off between the autonomy of the system and the user's effort, and identifying the most suitable presentation of actions promoting/demoting values for each action, which we will do following evidence from privacy nudges and notifications (Acquisti et al. 2017). Secondly, the aim of this model is to capture a single collaborative decision processes. Given that interactions over a OSN have an evolutionary nature, it seems interesting for us to analyse the influence over time that the outcomes of particular negotiations have on the overall users' behaviour, i.e. if it is possible to identify an historical component of the decision making

process defined by the previous experiences. Also, the formal syntax and semantics for the negotiation protocol need to be defined (McBurney and Parsons 2009).

References

Acquisti, A.; Adjerid, I.; Balebako, R.; Brandimarte, L.; Cranor, L. F.; Komanduri, S.; Leon, P. G.; Sadeh, N.; Schaub, F.; Sleeper, M.; et al. 2017. Nudges for privacy and security: Understanding and assisting users choices online. *ACM Computing Surveys (CSUR)* 50(3):44.

Acquisti, A.; Brandimarte, L.; and Loewenstein, G. 2015. Privacy and human behavior in the age of information. *Science* 347(6221):509–514.

Besmer, A., and Richter Lipford, H. 2010. Moving beyond untagging: photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1563–1572. ACM.

Bilsky, W.; Janik, M.; and Schwartz, S. H. 2011. The structural organization of human values-evidence from three rounds of the european social survey (ess). *Journal of Cross-Cultural Psychology* 42(5):759–776.

Criado, N., and Such, J. M. 2015. Implicit Contextual Integrity in Online Social Networks. *Information Sciences* 325:48–69.

Fatima, S.; Kraus, S.; and Wooldridge, M. 2014. *Principles of automated negotiation*. Cambridge University Press.

Fogués, R. L.; Such, J. M.; Espinosa, A.; and Garcia-Fornes, A. 2014. Bff: A tool for eliciting tie strength and user communities in social networking services. *Information Systems Frontiers* 16(2):225–237.

Fogues, R. L.; Murukannaiah, P. K.; Such, J. M.; and Singh, M. P. 2017a. Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making. *ACM Transactions on Computer-Human Interaction* 24(1):5:1–5:29.

Fogues, R. L.; Murukannaiah, P. K.; Such, J. M.; and Singh, M. P. 2017b. Sosharp: Recommending sharing policies in multiuser privacy scenarios. *IEEE Internet Computing* 21(6):28–36.

Ilia, P.; Polakis, I.; Athanasopoulos, E.; Maggi, F.; and Ioannidis, S. 2015. Face/Off: preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15*, 781–792. New York, New York, USA: ACM Press.

Kökciyan, N.; Yaglikci, N.; and Yolum, P. 2017. An argumentation approach for resolving privacy disputes in online social networks. *ACM Transactions on Internet Technology (TOIT)* 17(3):27.

Lampinen, A.; Lehtinen, V.; Lehmuskallio, A.; and Tamminen, S. 2011. We're in it together. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, 3217. New York, New York, USA: ACM Press.

Liang, K.; Liu, J. K.; Lu, R.; and Wong, D. S. 2015. Privacy concerns for photo sharing in online social networks. *IEEE Internet Computing* 19(2):58–63.

McBurney, P., and Parsons, S. 2009. Dialogue games for agent argumentation. In *Argumentation in artificial intelligence*. Springer. 261–280.

Misra, G., and Such, J. M. 2017. Pacman: Personal agent for access control in social media. *IEEE Internet Computing* 21(6):18–26.

Moyo, M.; Goodyear-Smith, F. A.; Weller, J.; Robb, G.; and Shulruf, B. 2016. Healthcare practitioners personal and professional values. *Advances in Health Sciences Education* 21(2):257–286.

Paci, F.; Squicciarini, A.; and Zannone, N. 2018. Survey on access control for community-centered collaborative systems. *ACM Computing Surveys* 51(1).

Patterson, F.; Prescott-Clements, L.; Zibarras, L.; Edwards, H.; Kerrin, M.; and Cousans, F. 2016. Recruiting for values in healthcare: a preliminary review of the evidence. *Advances in Health Sciences Education* 21(4):859–881.

Rajtmajer, S.; Squicciarini, A.; Such, J. M.; Semonsen, J.; and Belmonte, A. 2017. An ultimatum game model for the evolution of privacy in jointly managed content. In *International Conference on Decision and Game Theory for Security*, 112–130. Springer.

Schwartz, S. H. 2010. Basic values: How they motivate and inhibit prosocial behavior. *Prosocial motives, emotions, and behavior: The better angels of our nature* 14:221–241.

Schwartz, S. H. 2012. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture* 2(1):11.

Squicciarini, A. C.; Sundareswaran, S.; Lin, D.; and Wede, J. 2011. A3p: adaptive policy prediction for shared images over popular content sharing sites. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, 261–270. ACM.

Squicciarini, A. C.; Shehab, M.; and Paci, F. 2009. Collective privacy management in social networks. In *Proceedings of the 18th international conference on World wide web*, 521–530. ACM.

Stern, P. C.; Dietz, T.; and Guagnano, G. A. 1998. A brief inventory of values. *Educational and psychological measurement* 58(6):984–1001.

Such, J. M., and Criado, N. 2016. Resolving Multi-Party Privacy Conflicts in Social Media. *IEEE Transactions on Knowledge and Data Engineering* 28(7):1851–1863.

Such, J. M., and Criado, N. 2018. Multiparty privacy in social media. *Communications of the ACM* 61(8):74–81.

Such, J. M., and Rovatsos, M. 2016. Privacy Policy Negotiation in Social Media. *ACM Transactions on Autonomous and Adaptive Systems* 11(1):1–29.

Such, J. M.; Porter, J.; Preibusch, S.; and Joinson, A. 2017. Photo privacy conflicts in social media: a large-scale empirical study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3821–3832. ACM.

Such, J. M. 2017. Privacy and autonomous systems. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 4761–4767. AAAI Press.

Thøgersen, J.; Zhou, Y.; and Huang, G. 2016. How stable is the value basis for organic food consumption in china? *Journal of Cleaner Production* 134:214–224.

Vihavainen, S.; Lampinen, A.; Oulasvirta, A.; Silfverberg, S.; and Lehmuskallio, A. 2014. the clash between privacy and automation in social media. *Pervasive Computing, IEEE* 13(1):56–63.

Vishwamitra, N.; Li, Y.; Wang, K.; Hu, H.; Caine, K.; and Ahn, G.-J. 2017. Towards PII-based Multiparty Access Control for Photo Sharing in Online Social Networks. In *Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies - SACMAT '17 Abstracts*, 155–166. New York, New York, USA: ACM Press.

Walton, D., and Krabbe, E. C. 1995. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press.

Watson, J.; Lipford, H. R.; and Besmer, A. 2015. Mapping user preference to privacy default settings. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22(6):32.

Wisniewski, P.; Lipford, H.; and Wilson, D. 2012. Fighting for my space: Coping mechanisms for sns boundary regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 609–618. ACM.