



EXPRI, Agent EXPRI: Licence to Explain

Francesca Mosca, Stefan Sarkadi, Jose M. Such and Peter McBurney

King's College London


EXTRAAMAS 2020

Multi-user Privacy

Such *et al.*, 2017
Wisniewski *et al.*, 2012



 Friends

 Friends
of friends

Privacy regards the information we disclose about ourselves but also whatever others can disclose about us.

 Private

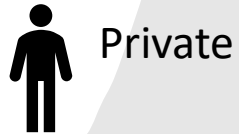
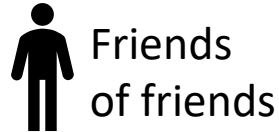
Public 



 Private

Multi-user Privacy

Such *et al.*, 2017
Wisniewski *et al.*, 2012



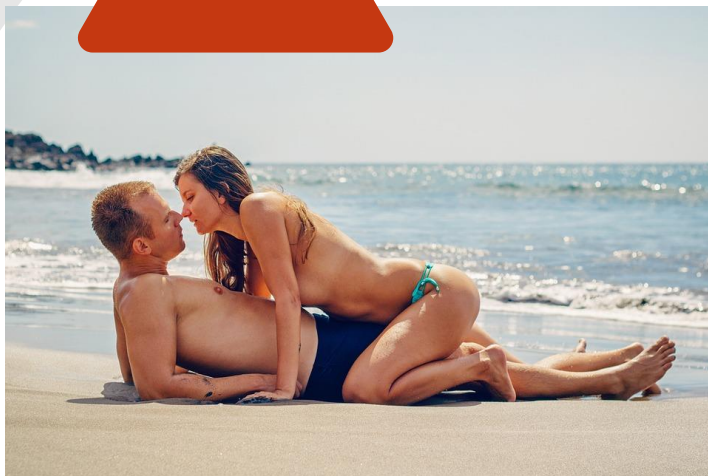
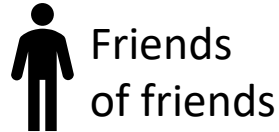
Privacy regards the information we disclose about ourselves but also whatever others can disclose about us.

96% of the participants to a large-scale study reported a **Multi-user Privacy Conflict (MPC)**:

- In 75% of the cases approach “all-or-nothing”.
- In 50% of the cases co-owners do not even complain.
- 70% of the conflicts was solved - general collaborative attitude.
- Sometimes there are no acceptable solutions.

Multi-user Privacy

Such *et al.*, 2017
Wisniewski *et al.*, 2012



Privacy regards the information we disclose about ourselves but also whatever others can disclose about us.

96% of the participants to a large-scale study reported a **Multi-user Privacy Conflict** (MPC):

- In 75% of the cases approach “all-or-nothing”.
- In 50% of the cases co-owners do not even complain.
- 70% of the conflicts was solved - general collaborative attitude.
- Sometimes there are no acceptable solutions.

Research to support and incentivise the search of a **compromise**.

Multi-user Privacy

Such *et al.*, 2017
Wisniewski *et al.*, 2012



Friends



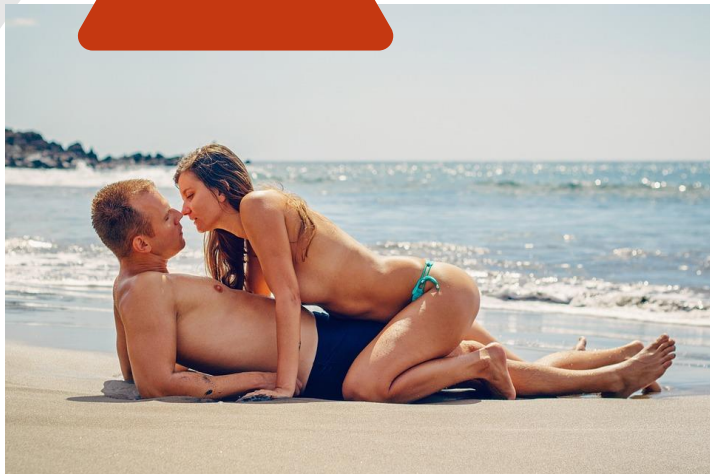
Friends
of friends



Private



CONFLICT!



Public



Private

Privacy regards the information we disclose about ourselves but also whatever others can disclose about us.

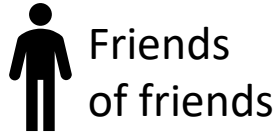
96% of the participants to a large-scale study reported a **Multi-user Privacy Conflict** (MPC):

- In 75% of the cases approach “all-or-nothing”.
- In 50% of the cases co-owners do not even complain.
- 70% of the conflicts was solved - general collaborative attitude.
- Sometimes there are no acceptable solutions.

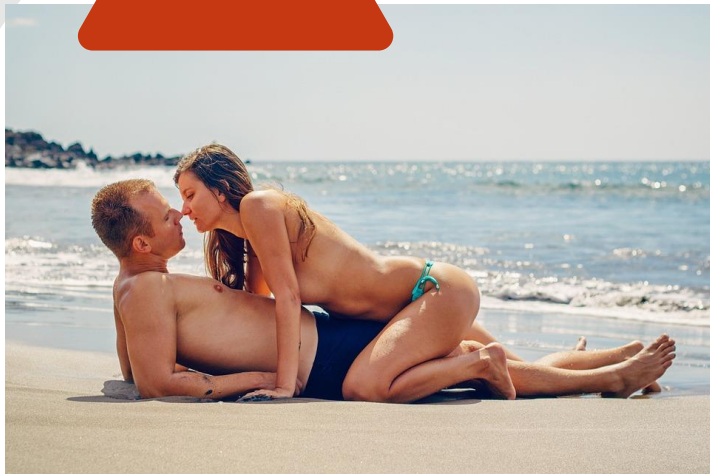
Research to support and incentivise the search of a **compromise**.

Multi-user Privacy

Such *et al.*, 2017
Wisniewski *et al.*, 2012



CONFLICT!



Privacy regards the information we disclose about ourselves but also whatever others can disclose about us.

96% of the participants to a large-scale study reported a **Multi-user Privacy Conflict** (MPC):

- In 75% of the cases approach “all-or-nothing”.
- In 50% of the cases co-owners do not even complain.
- 70% of the conflicts was solved - general collaborative attitude.
- Sometimes there are no acceptable solutions.

Research to support and incentivise the search of a **compromise**.

Related Work



Fogues *et al.*, (2017)

Kökciyan *et al.*, (2017)



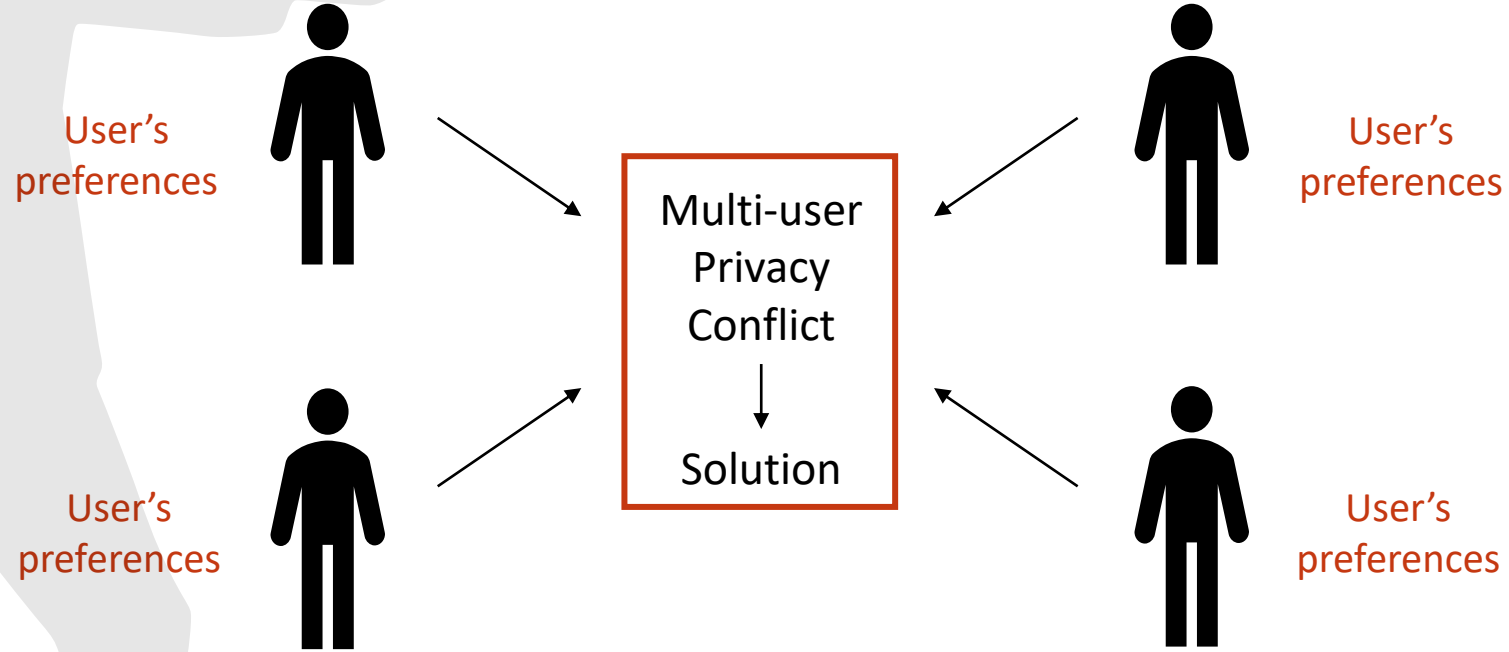
Explainability is crucial for autonomous systems to foster the users' trust

Numerous efforts to solve MPCs in the literature, but lack of explainability

- Fogues *et al.*, (2017): A recommendation system identifies the solution by considering a set of arguments
- Kökciyan *et al.*, (2017): Ontologies, semantic rules and persuasion dialogues

The Agent Architecture

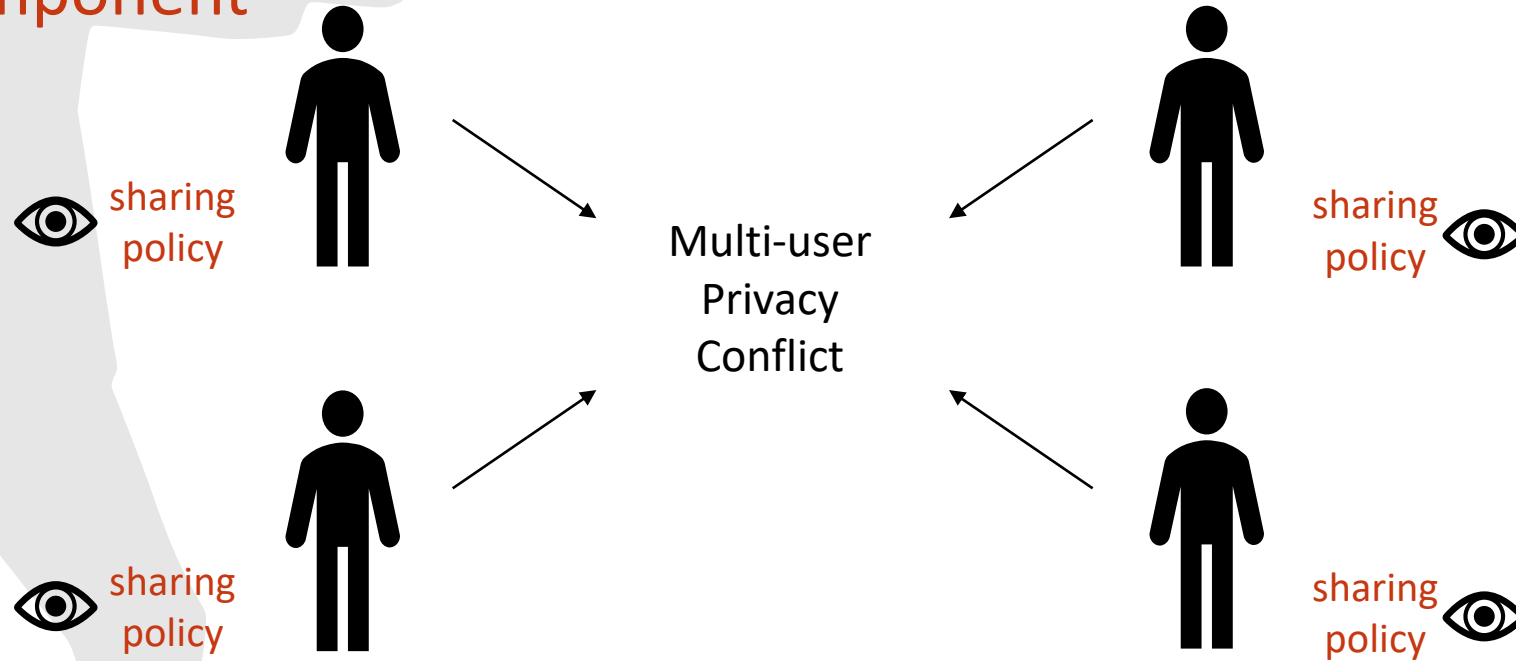
EXPRI



The agents *Ag* act on behalf of their users, according to their preferences, in order to identify a compromise acceptable for everyone.

The Agent Architecture

Utility Component

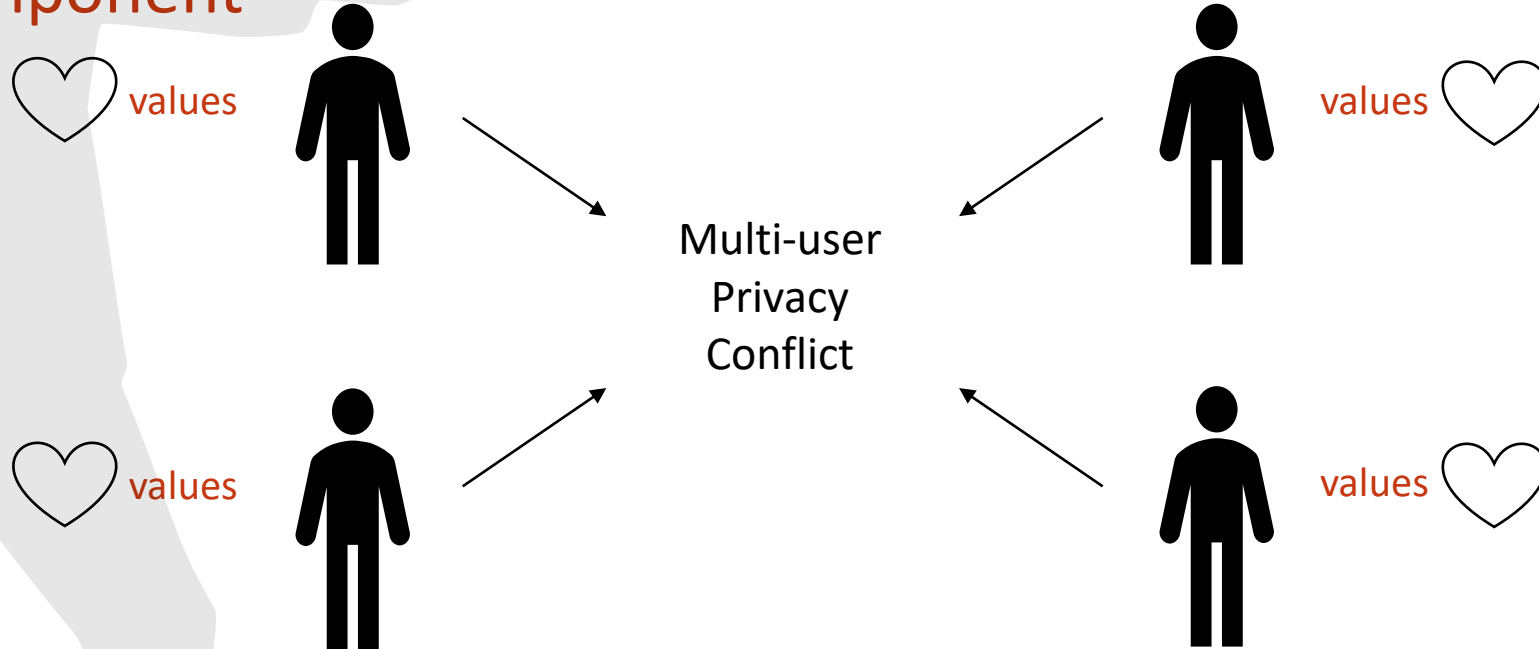


The preferred sharing policy is elicited for each user $k \in Ag$ and compared with each candidate solution in SP . Each $sp \in SP$ can generate for each user a gain or a loss in utility:

utility: $u_{k,sp}$

The Agent Architecture

Value Component



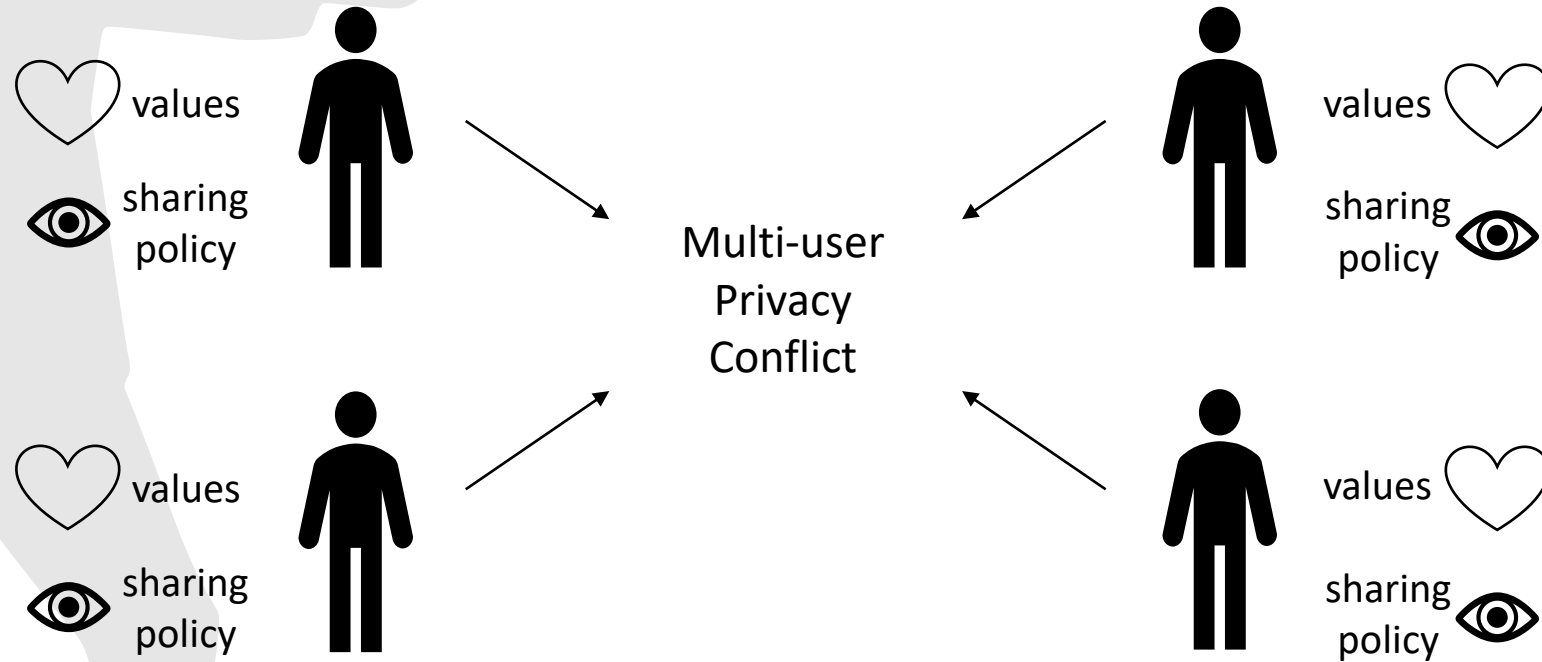
Schwartz, 2012
Mosca *et al.*, 2019



According to the Schwartz Theory of Basic Values, we define the user's morality according to 4 value-directions. We interpret the value-direction in the MPC context and we evaluate whether the values of the user $k \in Ag$ are promoted by selecting each candidate solution:

value promotion: $v_{k,sp}$

The Agent Architecture



Each agent $k \in Ag$ computes for each candidate solution $sp \in SP$ the **individual score** that represents the user's appreciation of the solution in terms of utility and value promotion. The individual scores are aggregated into the **collective score** for each $sp \in SP$:

$$s_{k,sp} = u_{k,sp} \cdot v_{k,sp}$$

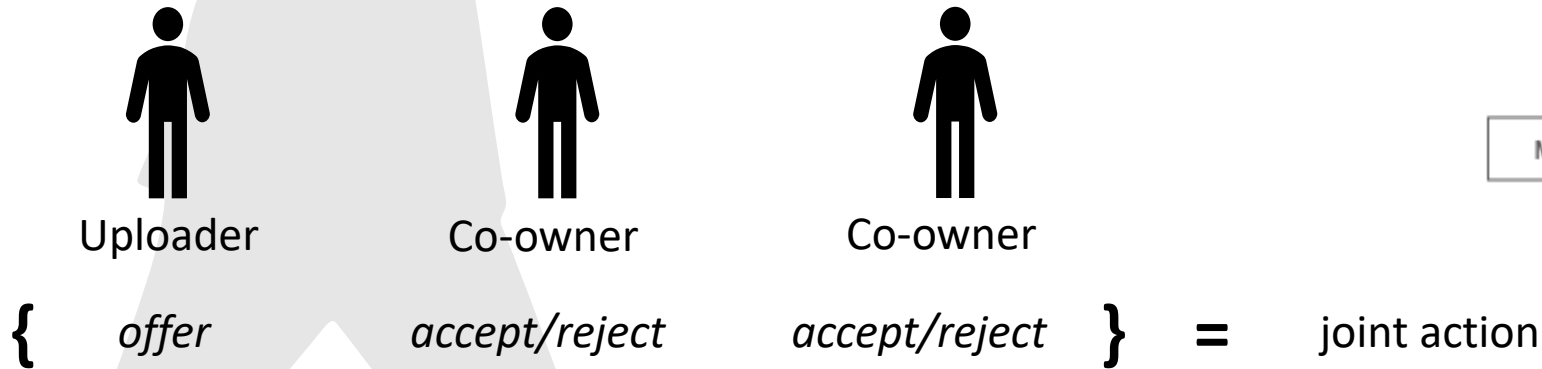
$$s_{sp} = \sum_{k \in Ag} s_{k,sp}$$

Explainable Agents

Cognitive Process

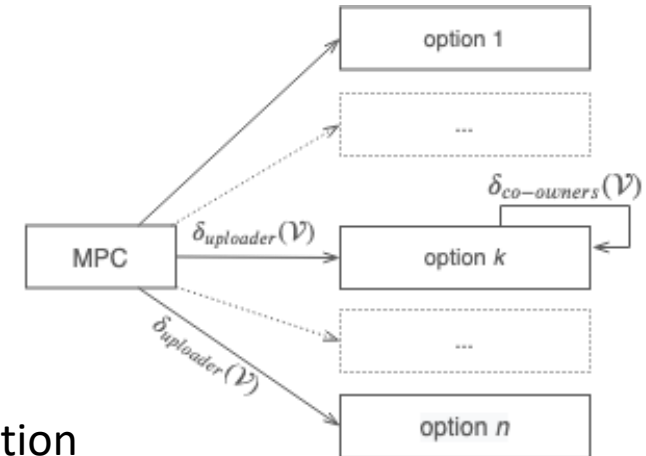
Miller, 2018

Atkinson and Bench-Capon, 2007



Value-based arguments and critical questions supporting/challenging each individual action

Practical reasoning techniques and computational argumentation



AS-U: Given the current conflict, I should offer the sharing policy sp , that will be accepted by the co-owners and therefore will solve the conflict, that will provide the score s_{sp} and that will promote my values V .

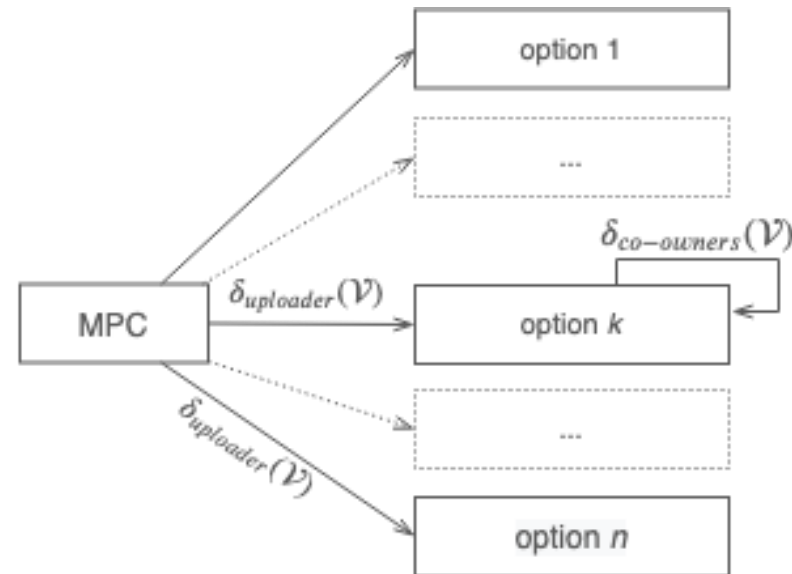
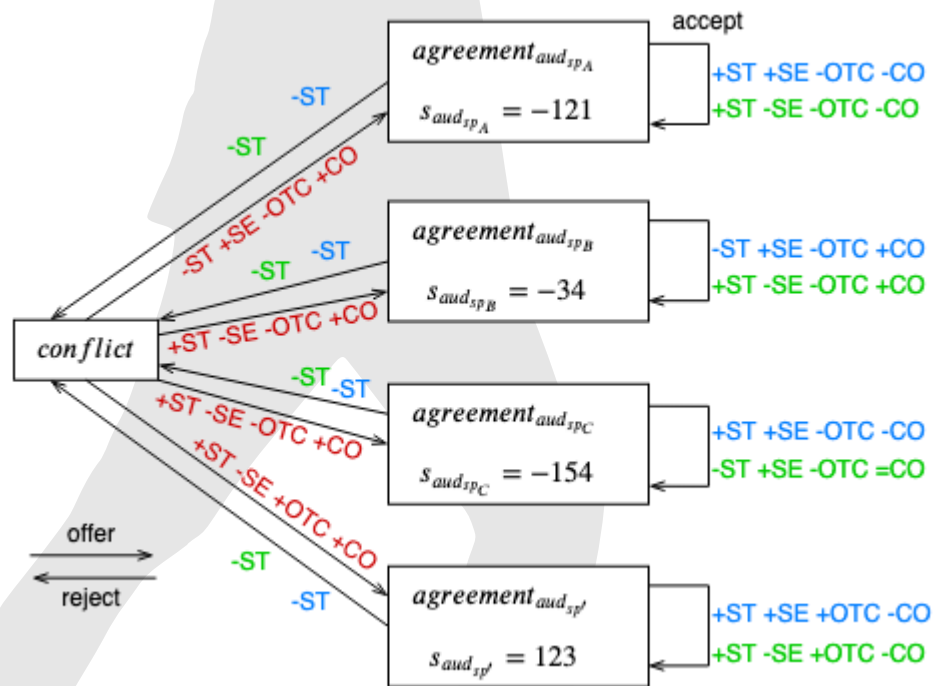
Explainable Agents

Cognitive Process

Atkinson and Bench-Capon, 2007



1. Problem Formulation



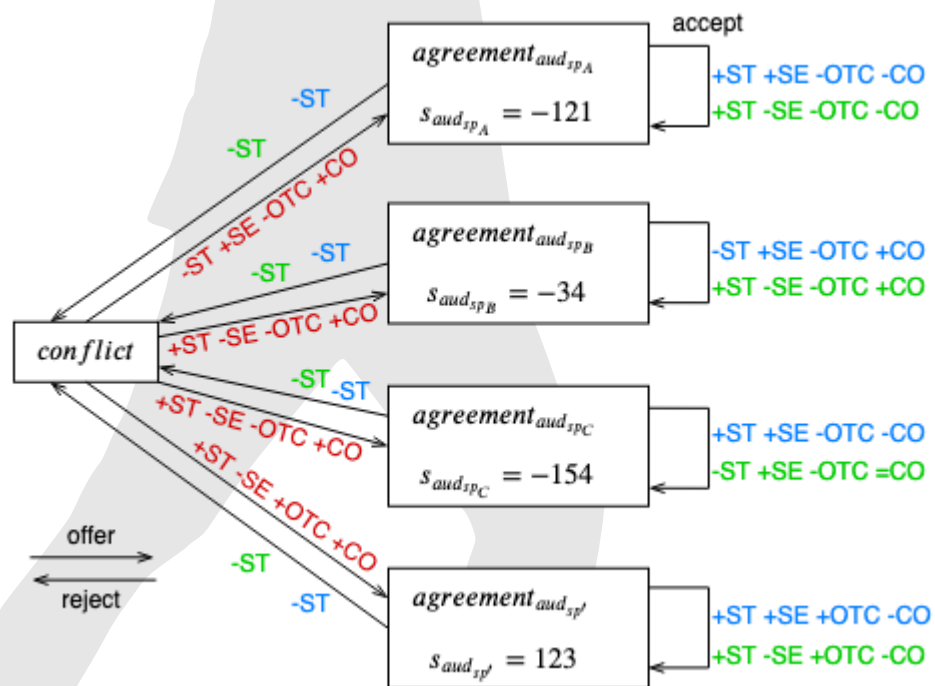
Explainable Agents

Cognitive Process

Atkinson and Bench-Capon, 2007

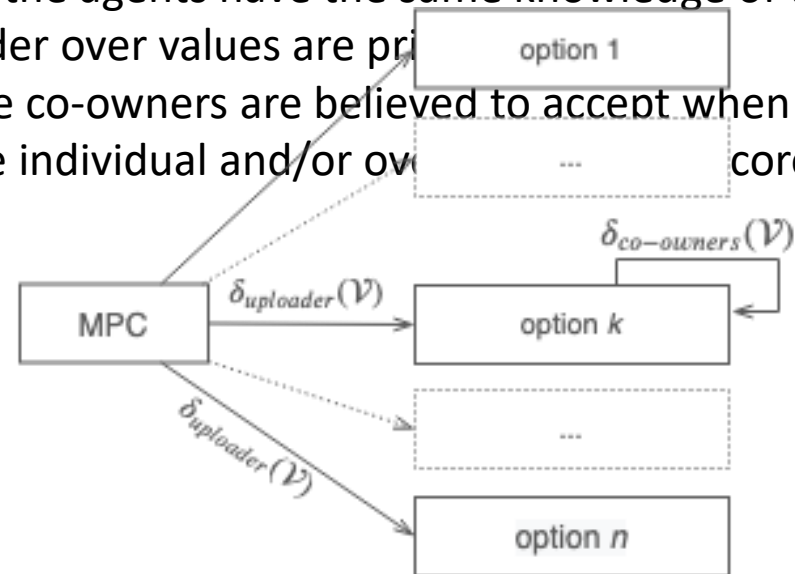


1. Problem Formulation



2. Epistemic assumptions:

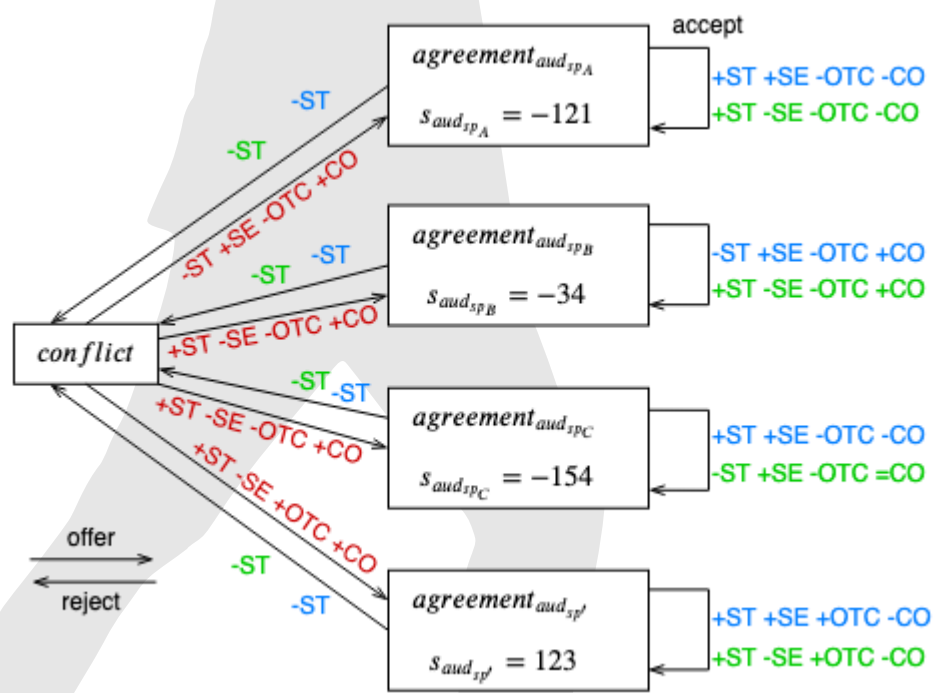
- All the agents have the same knowledge of the system (only the order over values are prioritized)
- The co-owners are believed to accept when the offer matches the individual and/or overall core



Explainable Agents

Cognitive Process

1. Problem Formulation

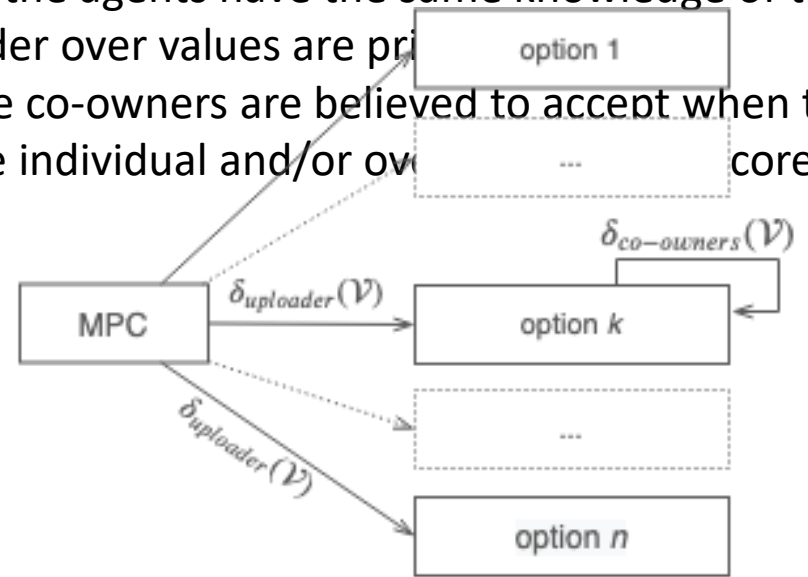


3. Choice of Action

AS-U and AS-C, and their critical questions, allow the creation of an argumentation framework, that provides the justification for action.

2. Epistemic assumptions:

- All the agents have the same knowledge of the system (only the order over values are prioritized)
- The co-owners are believed to accept when the offer matches the individual and/or overall utility core



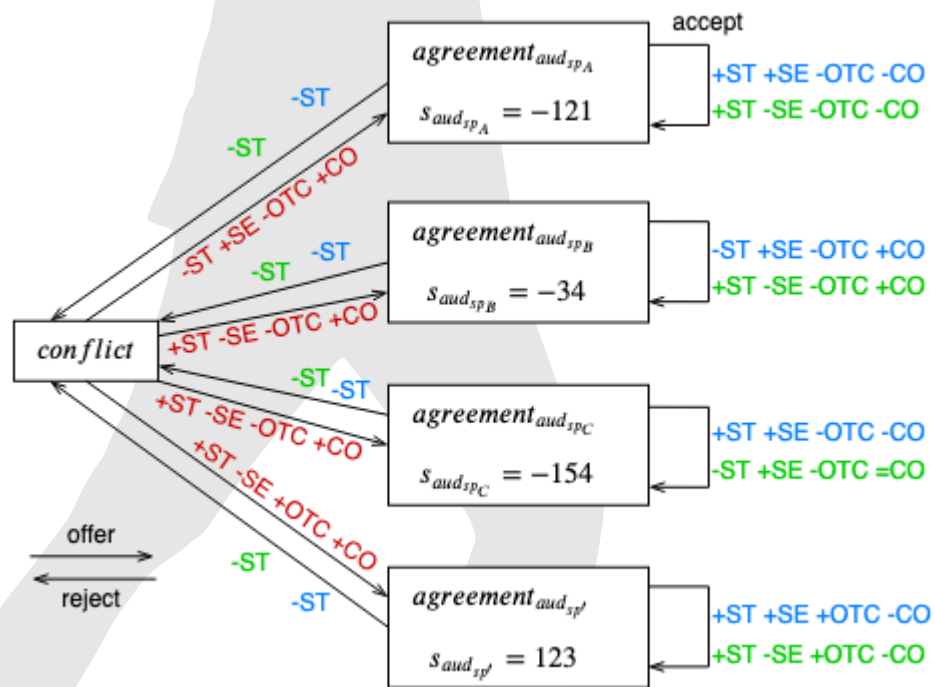
Explainable Agents

Cognitive Process

Atkinson and Bench-Capon, 2007



1. Problem Formulation

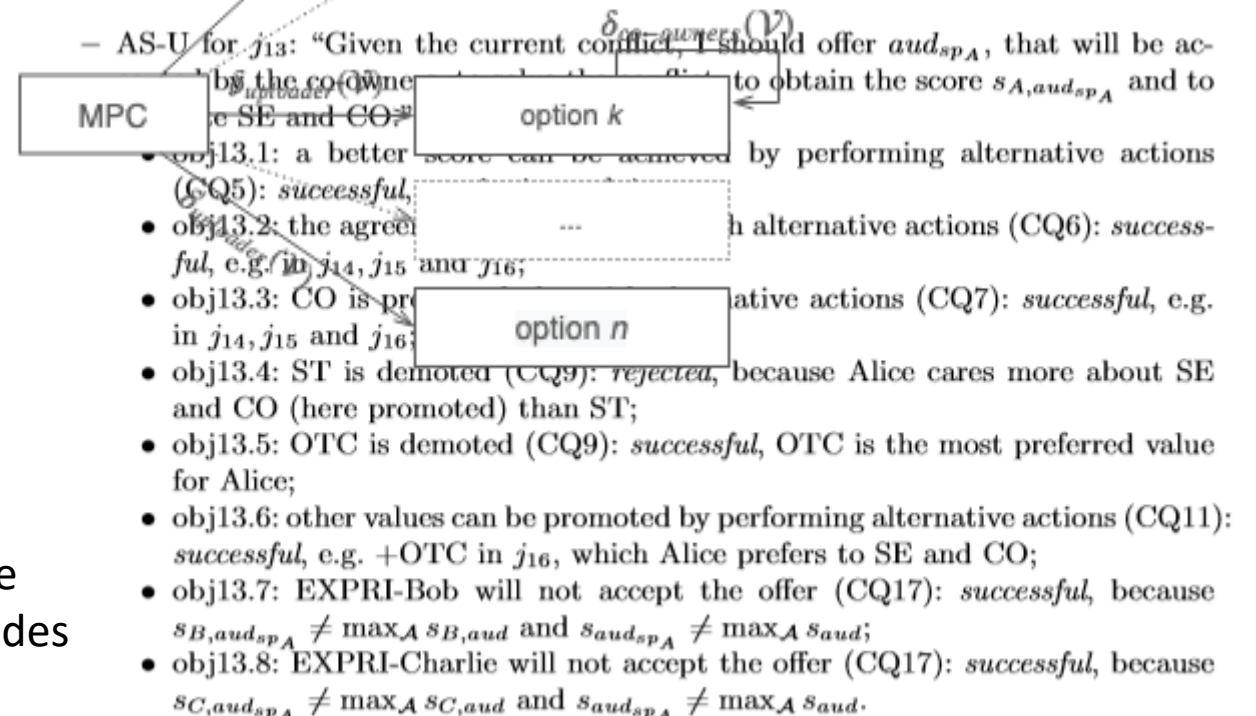


3. Choice of Action

AS-U and AS-C, and their critical questions, allow the creation of an argumentation framework, that provides the justification for action.

2. Epistemic assumptions:

- All the agents have the same knowledge of the system (only the order over values are prioritized)
- The co-owners are believed to accept when the offer matches the individual and/or overall score



Explainable Agents

Social Process

Miller, 2018



Tailored explanations



Contrastive explanation



Explaining conflicts

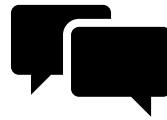
Explainable Agents

Social Process

Miller, 2018



Tailored explanations



Contrastive explanation



Explaining conflicts

Given the disagreement with Bob and Charlie about how to share your picture, to offer sp' is your most convenient action, because it would allow you to compromise with your friends (remember that openness-to-change is your most preferred value).

Why shouldn't I offer sp_A instead?

Because you could get a better score than the one guaranteed by sp_A (obj13.1), openness-to-change would be demoted (obj13.5), and because Bob and Charlie would most likely reject your offer (obj13.7 and obj13.8).

Explainable Agents

Social Process

Miller, 2018



Tailored explanations



Contrastive explanation



Explaining conflicts



Given the disagreement with Bob and Charlie about how to share your picture, to offer sp' is your most convenient action, because it would allow you to compromise with your friends (remember that openness-to-change is your most preferred value).

Why shouldn't I offer sp_A instead?

Because you could get a better score than the one guaranteed by sp_A (obj13.1), openness-to-change would be demoted (obj13.5), and because Bob and Charlie would most likely reject your offer (obj13.7 and obj13.8).

Explainable Agents

Social Process

Miller, 2018



Tailored explanations



Contrastive explanation



Explaining conflicts



Given the disagreement with Bob and Charlie about how to share your picture, to offer sp' is your most convenient action, because it would allow you to compromise with your friends (remember that openness-to-change is your most preferred value).



Why shouldn't I offer sp_A instead?



Because you could get a better score than the one guaranteed by sp_A (obj13.1), openness-to-change would be demoted (obj13.5), and because Bob and Charlie would most likely reject your offer (obj13.7 and obj13.8).

Explainable Agents

Social Process

Miller, 2018



Tailored explanations



Contrastive explanation



Explaining conflicts



Given the disagreement with Bob and Charlie about how to share your picture, to offer sp' is your most convenient action, because it would allow you to compromise with your friends (remember that openness-to-change is your most preferred value).



Why shouldn't I offer sp_A instead?



Because you could get a better score than the one guaranteed by sp_A (obj13.1), openness-to-change would be demoted (obj13.5), and because Bob and Charlie would most likely reject your offer (obj13.7 and obj13.8).

Discussion



- Explainability crucial for autonomous systems
- Explainability given by cognitive and social processes
- EXPRI's cognitive process guaranteed by performing practical reasoning
- EXPRI's social process to be further studied

References

F. Mosca, S. Sarkadi, J.M. Such, P. McBurney,
Agent EXPRI: Licence to Explain @ EXTRAAMAS 2020

- Acquisti et al., *Privacy and human behavior in the age of information*. Science 347,6221 (2015)
- Atkinson and Bench-Capon, *Practical reasoning as presumptive argumentation using action based alternating transition systems*. AIJ 171, 10-15 (2007)
- Cranefield et al., *Accountability for practical reasoning agents*. In: Proc. of AT (2018)
- Fogues et al., *Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making*. ACM TOCHI (2017)
- Kökciyan et al., *An argumentation approach for resolving privacy disputes in online social networks*. ACM TOIT (2017)
- Langley, *Explainable, normative, and justified agency*. In: Proc. of AAI (2019)
- Miller, *Explanation in artificial intelligence: Insights from the social sciences*, AIJ (2018)
- Mosca, *Value-Aligned and Explainable Agents for Collective Decision Making: Privacy Application*. DC AAMAS (2020)
- Mosca et al., *Towards a Value-driven Explainable Agent for Collective Privacy*. AAMAS Extended Abstract (2020)
- Paci et al., *Survey on access control for community-centered collaborative systems*. ACM CSUR 51,1 (2018).
- Schwartz, *An overview of the Schwartz theory of basic values*. Online readings in Psychology and Culture 2,1 (2012)
- Such et al 2017. *Photo privacy conflicts in social media: a large-scale empirical study*. In Proc. of CHI (2017)

References

F. Mosca, S. Sarkadi, J.M. Such, P. McBurney,
Agent EXPRI: Licence to Explain @ EXTRAAMAS 2020

- Acquisti et al., *Privacy and human behavior in the age of information*. Science 347,6221 (2015)
- Atkinson and Bench-Capon, *Practical reasoning as presumptive argumentation using action based alternating transition systems*. AIJ 171, 10-15 (2007)
- Cranefield et al., *Accountability for practical reasoning agents*. In: Proc. of AT (2018)
- Fogues et al., *Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making*. ACM TOCHI (2017)
- Kökciyan et al., *An argumentation approach for resolving privacy disputes in online social networks*. ACM TOIT (2017)
- Langley, *Explainable, normative, and justified agency*. In: Proc. of AAI (2019)
- Miller, *Explanation in artificial intelligence: Insights from the social sciences*, AIJ (2018)
- Mosca, *Value-Aligned and Explainable Agents for Collective Decision Making: Privacy Application*. DC AAMAS (2020)
- Mosca et al., *Towards a Value-driven Explainable Agent for Collective Privacy*. AAMAS Extended Abstract (2020)
- Paci et al., *Survey on access control for community-centered collaborative systems*. ACM CSUR 51,1 (2018).
- Schwartz, *An overview of the Schwartz theory of basic values*. Online readings in Psychology and Culture 2,1 (2012)
- Such et al 2017. *Photo privacy conflicts in social media: a large-scale empirical study*. In Proc. of CHI (2017)

Thank you for your attention! Please contact me for any questions: francesca.mosca@kcl.ac.uk