Engineering Multiagent Systems for Ethics and Privacy-Aware Social Computing

Nirav Ajmeri
(Under the guidance of Professor Munindar P. Singh)

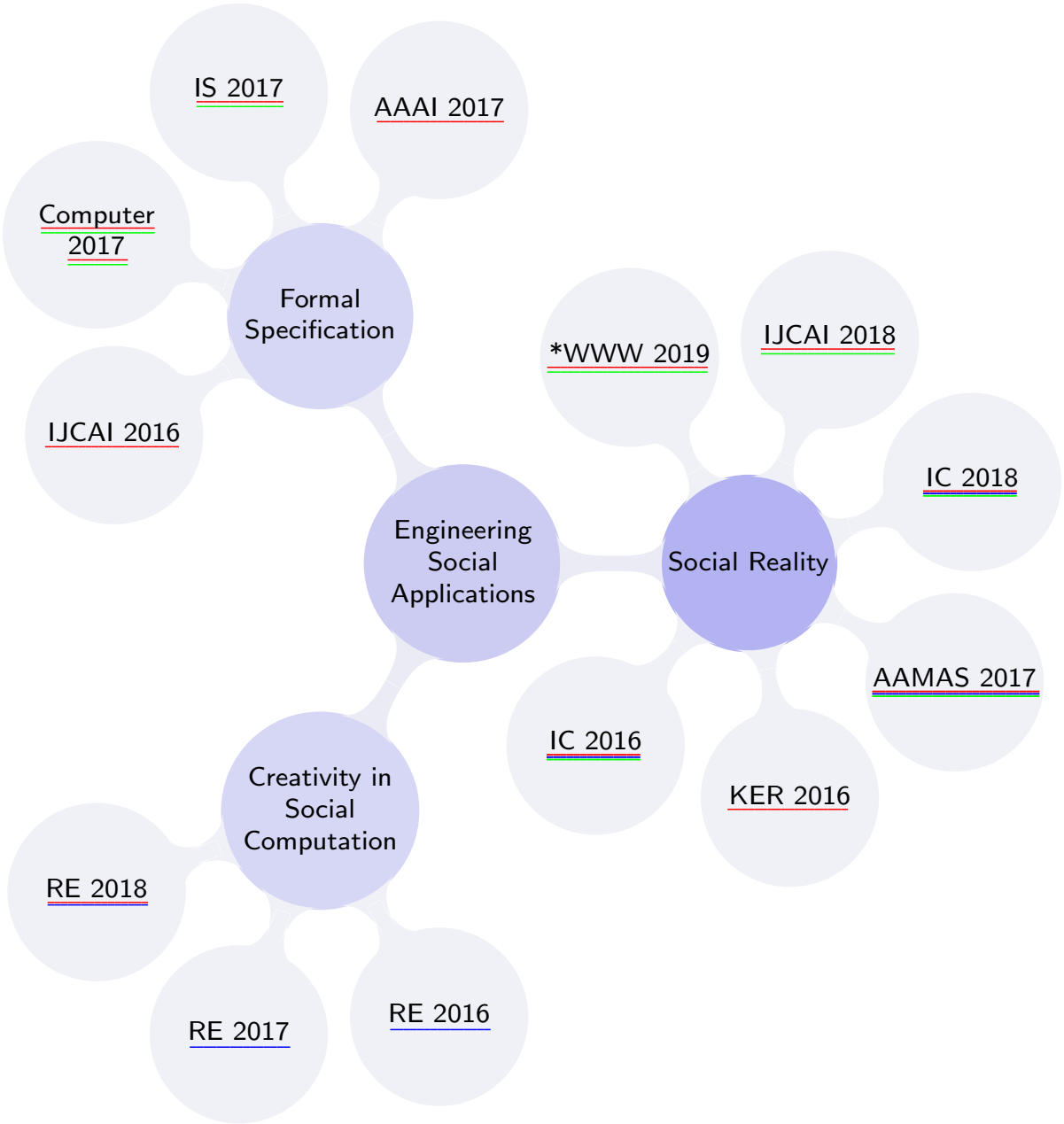
Department of Computer Science
North Carolina State University

December 2018

Outline

NSF's "Dear Colleague Letter" on FEAT (NSF 19-016)

- Fairness *in decision-making*
- Ethics *via incorporating values*
- Accountability *by social norms*
- Transparency *via understanding social context*



AI, SE, Privacy, * in-review

Examples of Ethical Concerns

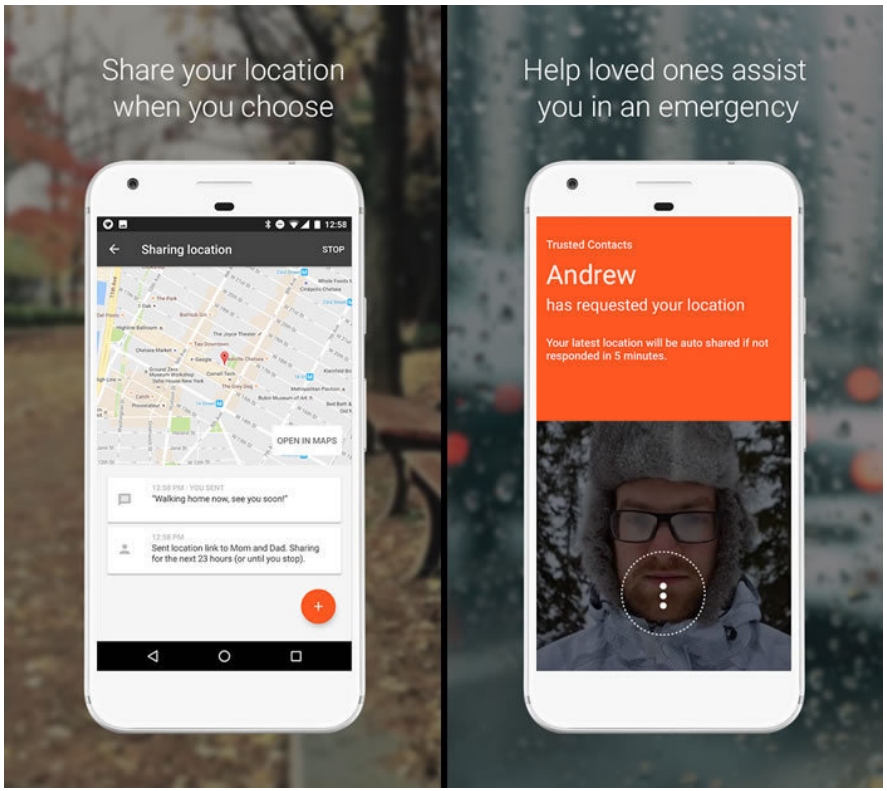
Audio leaking: Intrusion of solitude and disclosure of music taste

Source: `https://twitter.com/akokitamura/status/728521725172846592`

Examples of Privacy Concerns

Location sharing

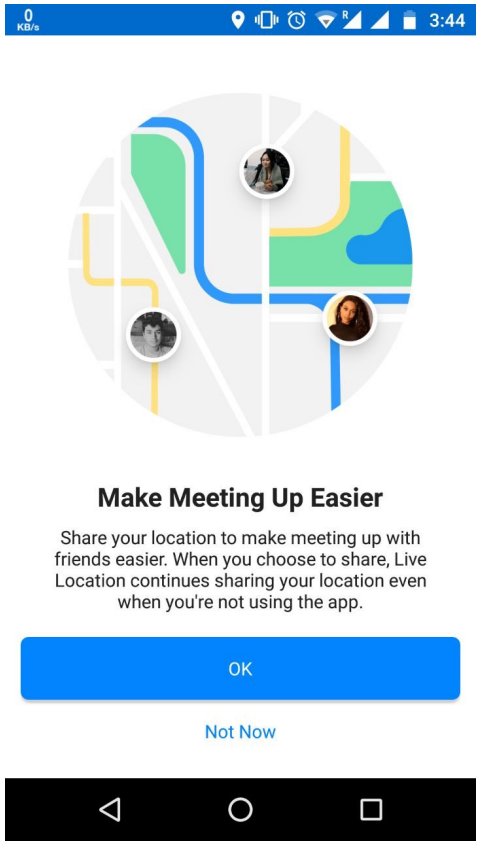
Google: Location sharing



Source: `https://www.csoonline.com`

Your latest location is auto shared if you do not respond in 5 minutes

Messenger: Live location



When you choose to share, Live Location continues sharing your location even when you are not using the app

Concepts

Social norm as defined by Singh [2013], is a relation between two parties, a subject and an object, and involves an antecedent (which brings a norm in force) and a consequent (which brings the norm to satisfaction or violation)

Social context is the circumstance under which an agent takes an action [Dey, 2001]

Deviation is a perceived violation of a norm [Nardin et al., 2016]

Values are guiding principles of humans [Schwartz, 2012; Friedman et al., 2008; Rokeach, 1973]

Ethics is subsumed in the theory of values [Friedman et al., 2008]

Privacy is a value with an ethical import [Langheinrich, 2001; Taylor, 2002;]

Research Objective

To help software developers in engineering personal agents that deliver an ethical and privacy-respecting social experience to stakeholders via *modeling and reasoning about social norms, social context, and value preferences*

Socially Intelligent Personal Agent (SIPA)

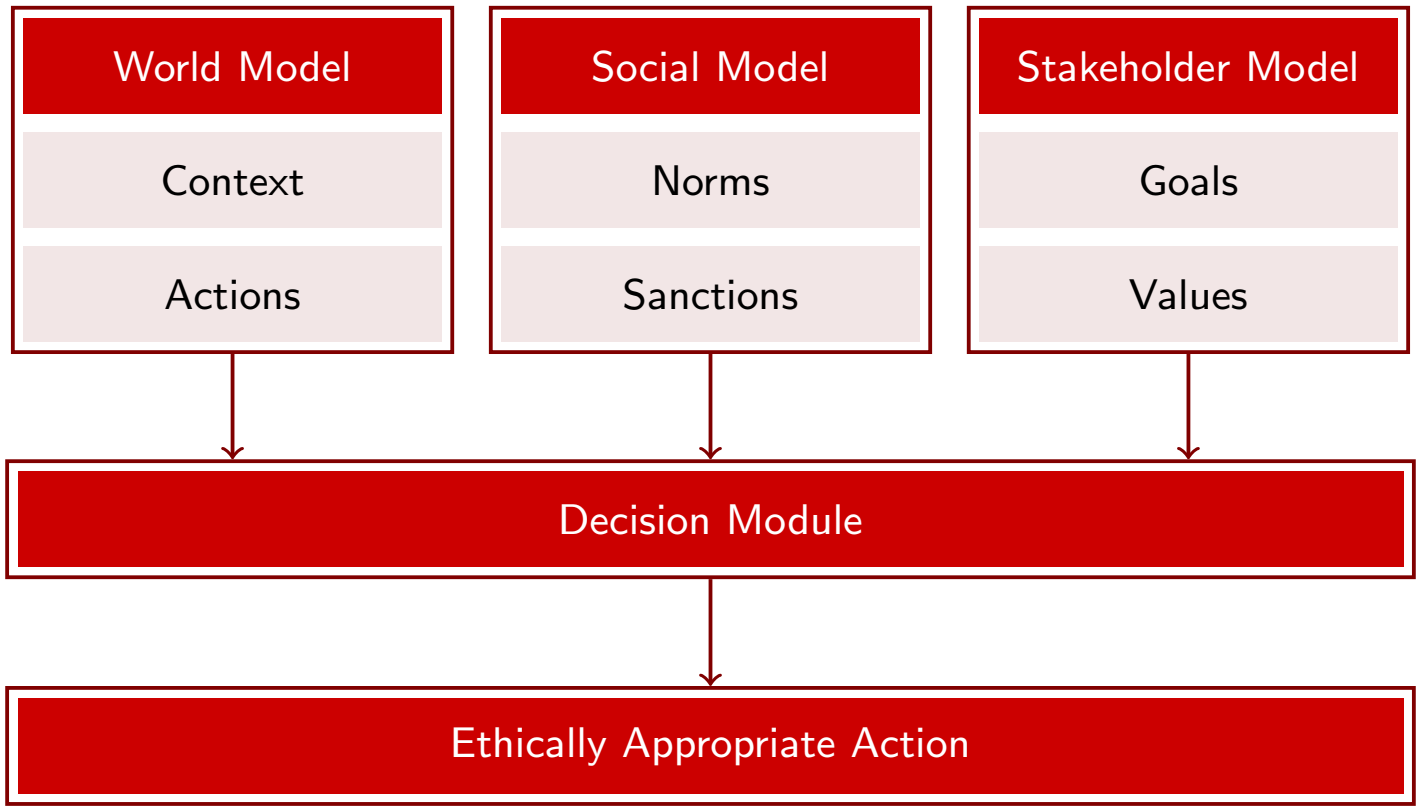
A SIPA adapts to social context and suppports meeting social expectations

- Ethical: Seeks to balance needs of
  - Primary stakeholder (user), who directly interacts with the agent
  - Secondary stakeholders, who are affected by the agent's actions

Challenge: Understanding Social Reality

- Modeling social intelligence
- Understanding social context
- Reasoning about values stakeholders

A SIPA: Schematically



Research Questions

RQ Social intelligence: How can modeling social intelligence in a SIPA help deliver a social experience and respects its stakeholders' privacy?

- Arnor, a software engineering method

RQ Context: How can SIPAs share and adapt to deviation contexts, and learn contextually relevant norms?

- Poros, a context reasoning approach

RQ Values: Does an ability to reason about values promoted or demoted by actions and an understanding of preferences among these values help a SIPA deliver a value-driven social experience to all its stakeholders?

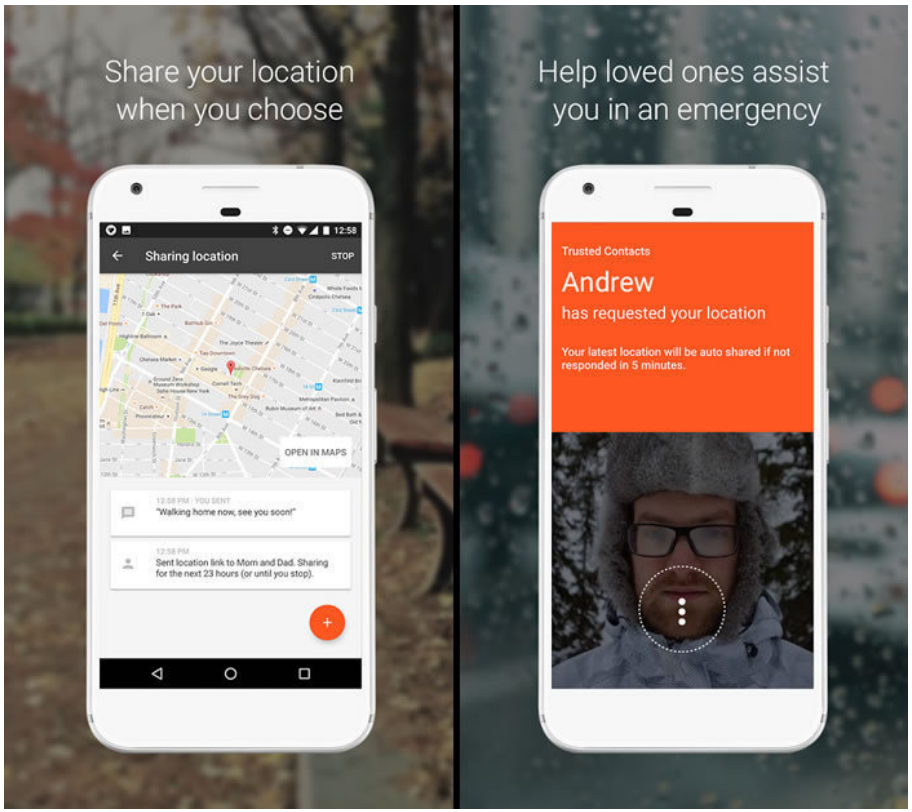
- Ainur, a decision-making framework

Outline

Norms and Values

RQ_{Values}: Does an ability to reason about values promoted or demoted by actions and an understanding of preferences among these values help a SIPA deliver a value-driven social experience to all its stakeholders?

Pichu: A location sharing SIPA



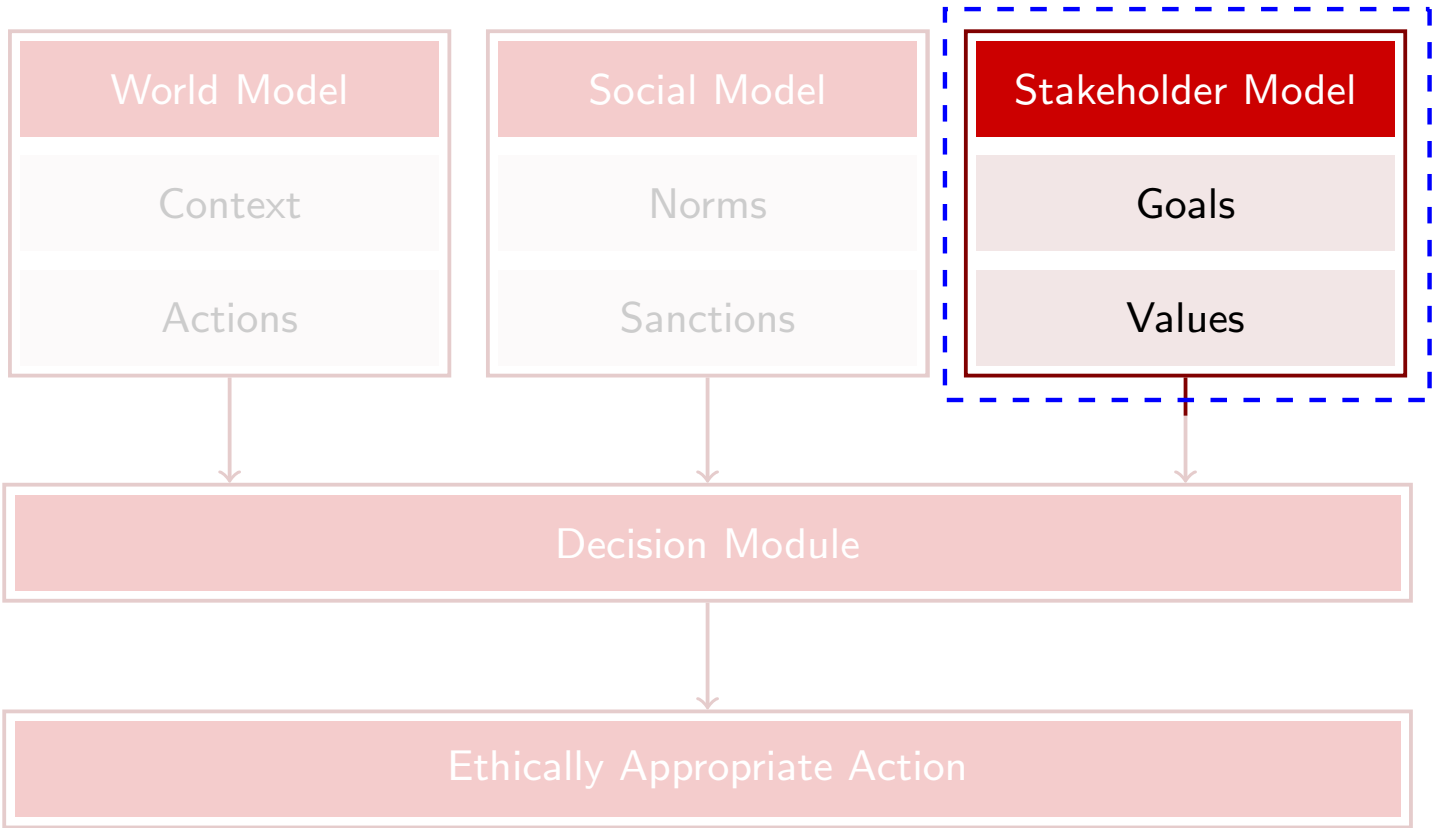
Source: `https://www.csoonline.com/article/3147286/security/google-launches-trusted-contacts-location-sharing-app.html`

Stakeholders

- Frank, a high school student; prefers pleasure and recognition
- Andrew, Frank's father; prefers safety
- Hope, Frank's aunt and also an intelligence analyst; prefers privacy

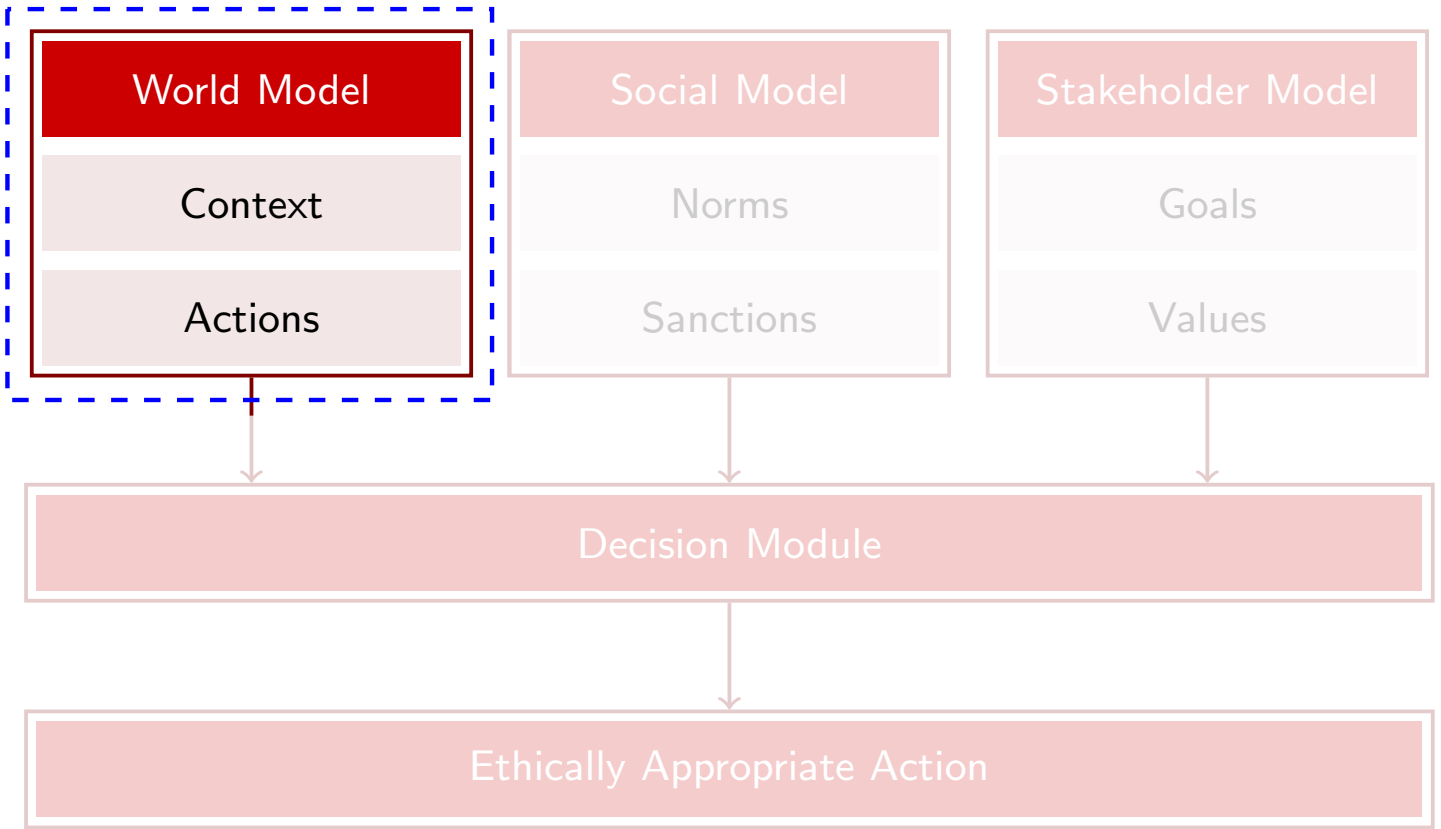
Stakeholder Model

A SIPA's stakeholders and their goals and values



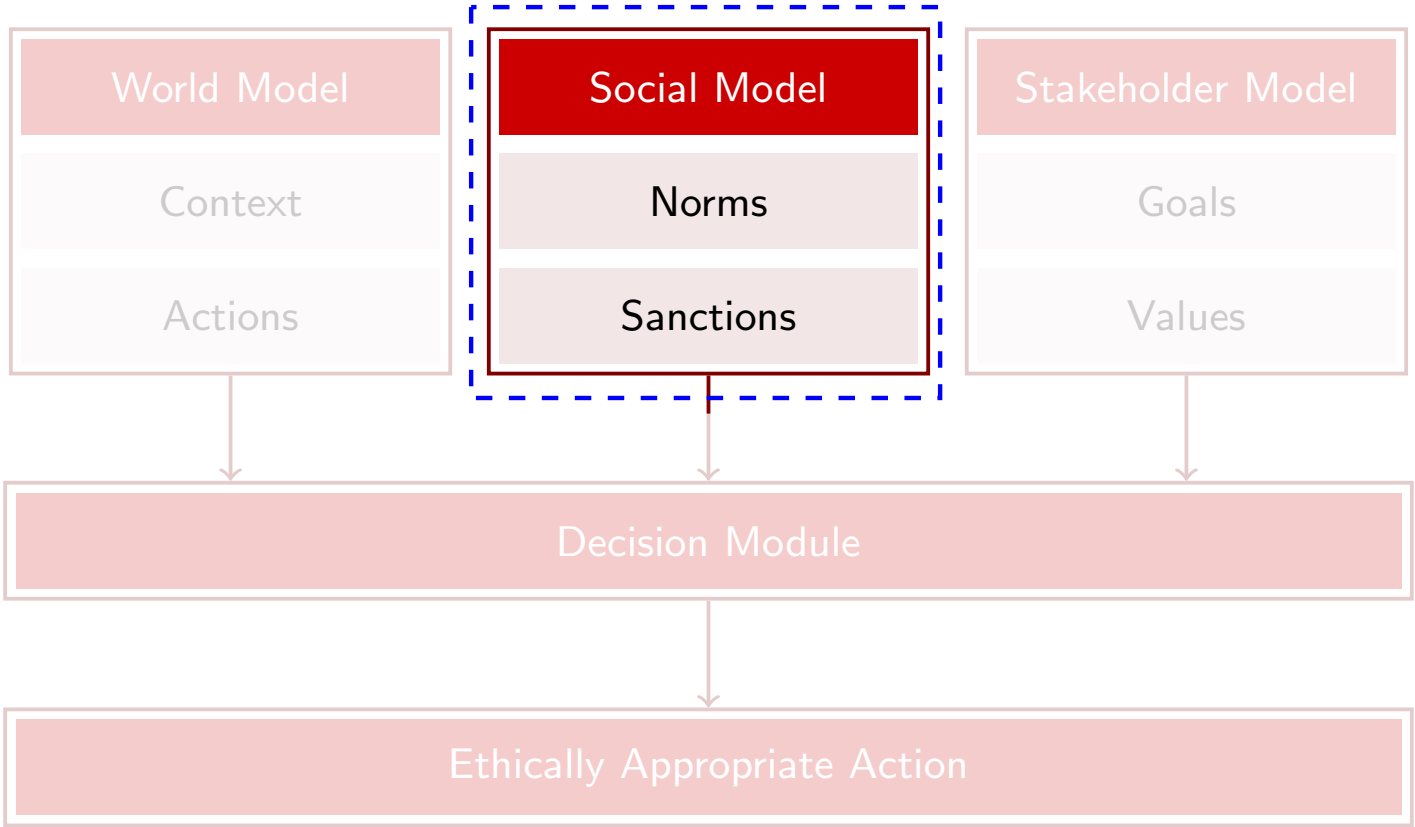
World Model

Context in which a SIPA acts



Social Model

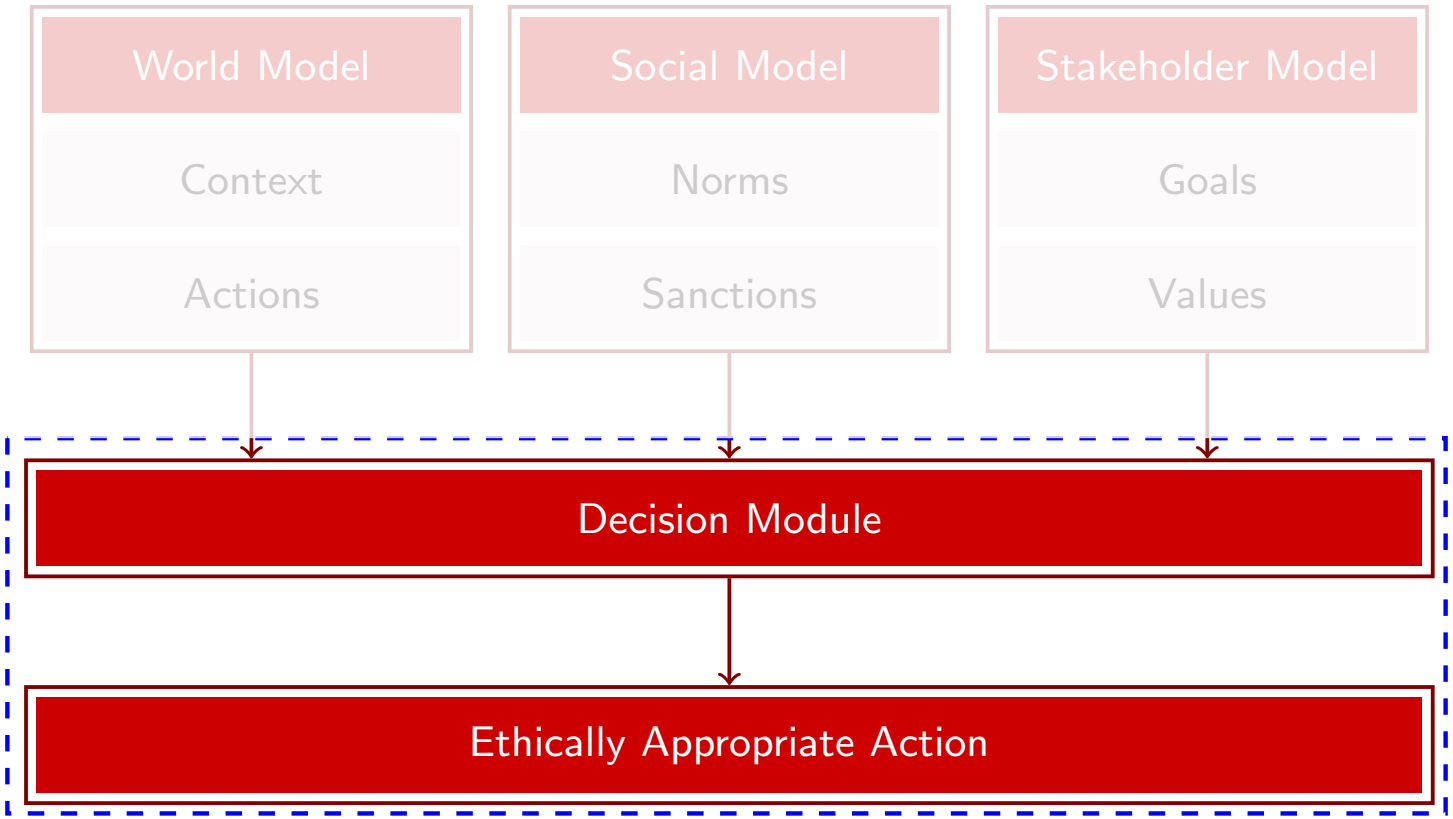
Norms governing a SIPA's interactions in a society and the associated sanctions



Decision Module

Incorporates VIKOR [Opricovic and Tzeng, 2004], a multicriteria decision-making method

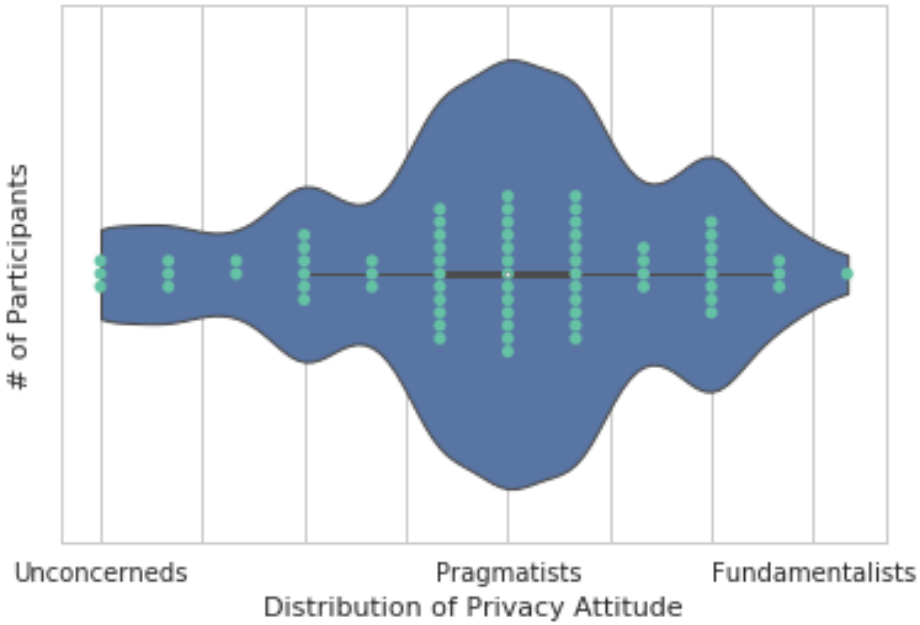
- Norms may conflict with actions
- Stakeholders' value preferences may not align



Evaluation: Crowdsourcing Study

Participants: 58 students enrolled in a mixed graduate and undergraduate-level computer science course

Privacy attitude survey: Level of comfort in sharing personal information

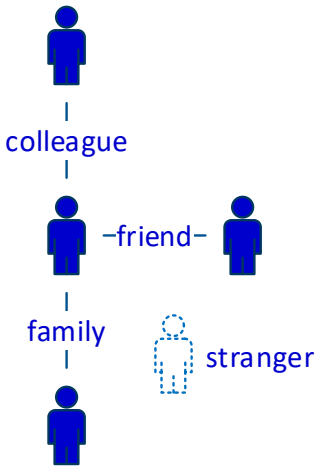


Context sharing surveys: Select context sharing policy

- Phase 1. Based on context, including place and social relationship
- Phase 2. Based on context and values (pleasure, privacy, recognition, safety)

Evaluation: Simulation

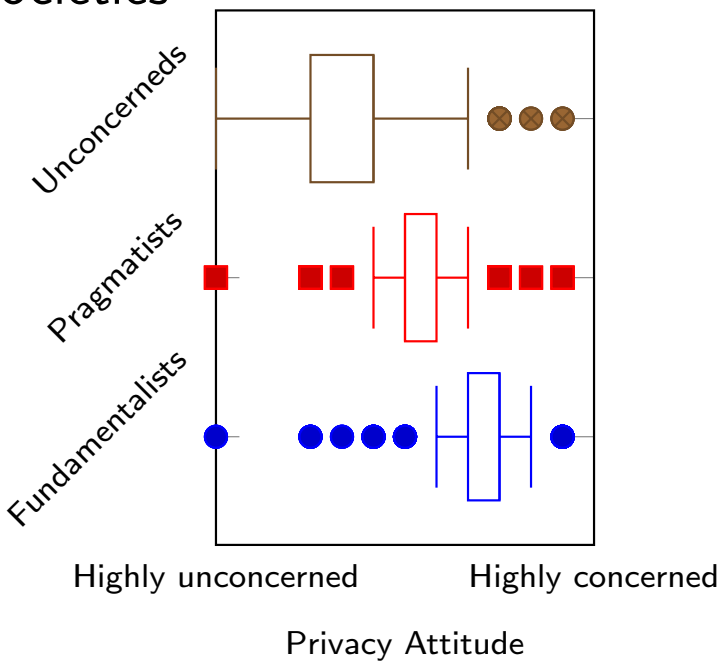
Study unit: Pichu SIPA



Simulated societies

- Mixed
- Fundamentalists,
- Pragmatists
- Unconcerneds

Privacy attitude distribution of societies



Decision-making strategies:

S_{Ainur}: Policy based on VIKOR

$S_{primary}$: Policy based on primary stakeholder's preferences

$S_{conservative}$: Least privacy-violating sharing policy

$S_{majority}$: Most common sharing policy

Metric

Mean social experience is the mean utility obtained by a society as a whole based on context sharing policy decisions

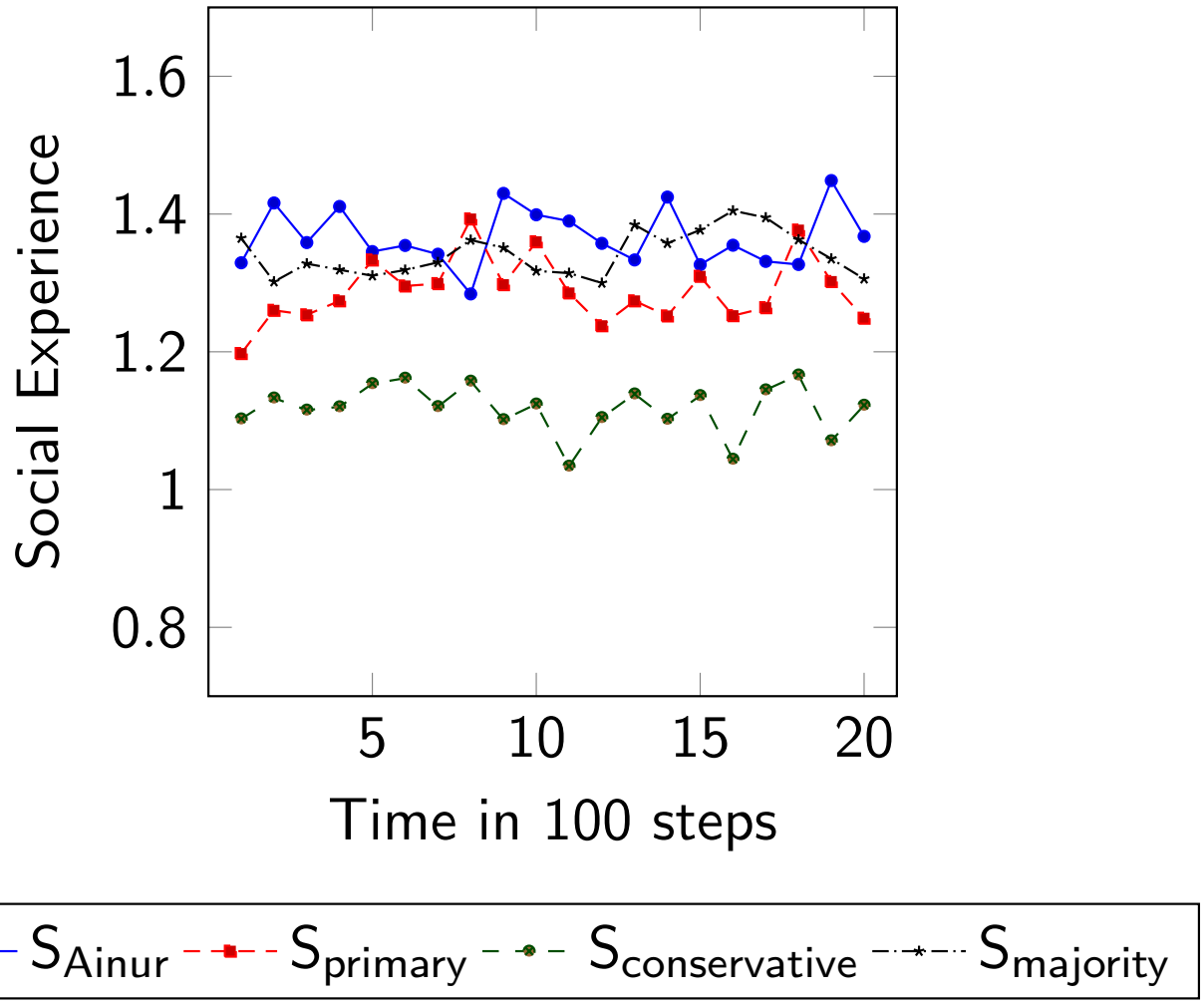
Best individual experience is the maximum utility obtained by one or more of the SIPA's stakeholders during a single interaction

Worst individual experience is the minimum utility obtained by one or more of the SIPA's stakeholders during a single interaction

Fairness is the reciprocal of the difference between the best and worst individual experience

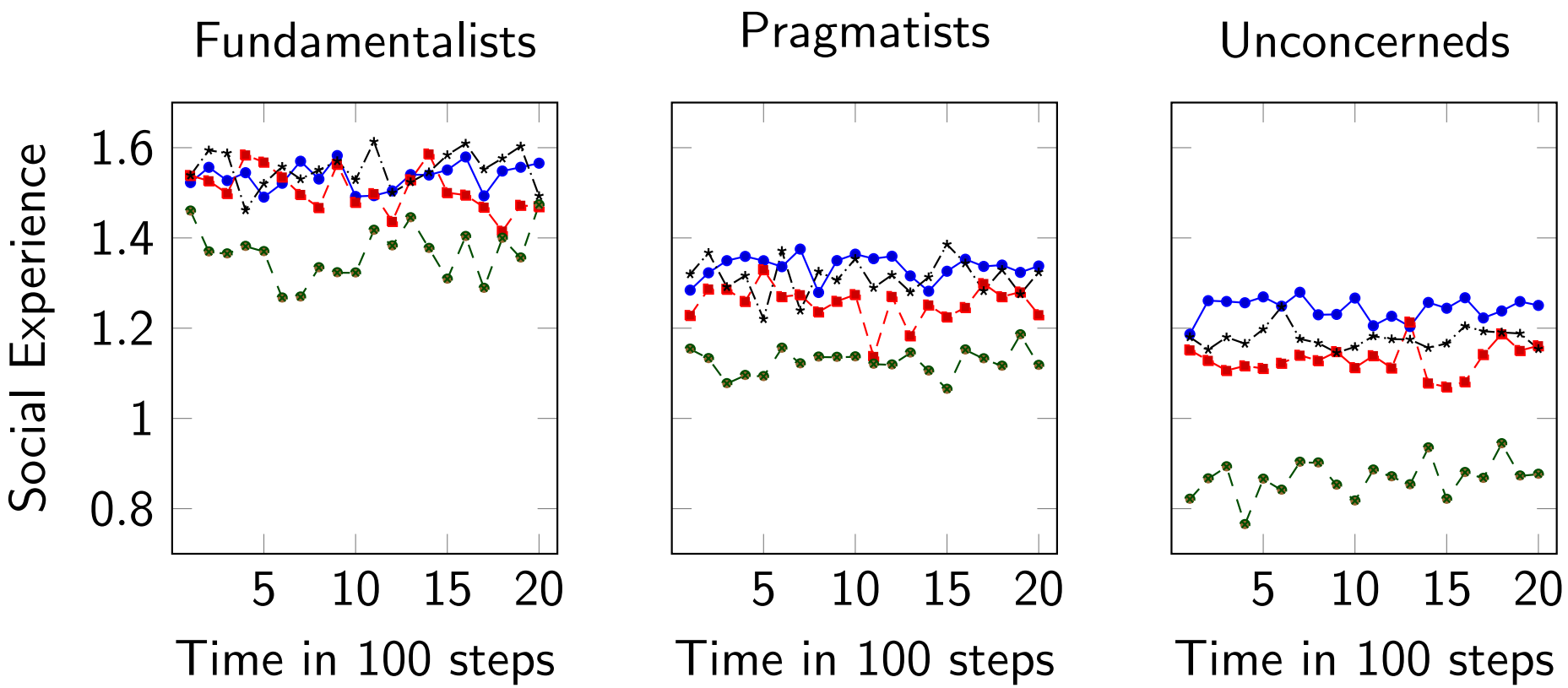
Experiment with Mixed Privacy Attitudes

Result: Ainur yields better mean social experience, mean worst individual experience, and fairness than other decision-making strategies



Experiments with Majority Privacy Attitudes

Result: Ainur maximizes the worst individual experience and yields better fairness than other decision-making strategies



Threats to Validity and Mitigation

Threats:

- Simulation as an evaluation methodology
- Unreliability of self-reported attitudes
- Survey sample not representative of actual population

Limitations (because of logistical reasons):

- Limited set of predetermined situations
- Limited set of actions

Outline

Conclusions and Relationship to FEAT

- Seeking to advance the science of privacy by tackling nuanced notions of privacy (understood as an ethical value) in personal agents
- Contributions:

  Modeling social intelligence: Arnor, a software engineering method to engineer privacy-aware personal agents (**Fairness**; **Accountability**)

  Understanding social context: Poros, an approach that enables personal agents to infer contextually relevant social norms that preserve privacy (**Accountability**; **Transparency**)

  Understanding value preferences: Ainur, a decision-making framework to design personal agents that can reason about values and act ethically (**Fairness**; **Ethics**)

Possible Directions for Future Dissertations

- Artificial Intelligence

  Social reality: White lies and affect in personal agents (building on IJCAI 2018 and Trust 2014 works)

  Formal specification: Argumentation and value-based reasoning (building on Computer 2017 and IJCAI 2016 works)

- Software Engineering

  Creativity: CrowdRE for privacy requirements (building on RE 2016 and RE 2018 works)

  Social reality: RE for ethical systems (building on AAMAS 2017)

- Privacy

  Social reality: Middleware based on Ainur as a privacy-enhancing technology to support ethical decision-making

  Social reality: Usable privacy and ethics

Acknowledgments

- Adviser: Dr. Munindar P. Singh
- Committee: Drs. Jon Doyle, Will Enck, Chris Mayhorn, Jessica Staddon, and Laurie Williams
- Past and present collaborators

  Engineering privacy and ethics in social applications with PK Murukannaiah and H Guo [IC 2016, AAMAS 2017, IC 2018, IJCAI 2018], and MB van Riemsdjik and P Pasotti;
  Reasoning about normative conflicts with J Jiang, R Chirkova, and J Doyle [IJCAI 2016; HotSoS 2016];
  Sanctions and cybersecurity with H Du, BY Narron, S Al-Amin, S Goyal, E Berglund, and J Doyle [HotSoS 2015, ACySe 2015, SIMPAT 2018];
  Norms and sociotechnical systems with Ö Kafalı [IS 2016, AAAI 2017];
  Sanction typology with LG Nardin, T Balke-Visser, AK Kalia, and JS Sichman [KER 2016];
  Trust and emotions with AK Kalia, KS Chan, JH Cho, and S Adalı [TRUST 2014];
  Argumentation and secure service policies with CW Hang and SD Parsons [Computer 2017];
  Analytic workflow with G Yuan, C Allred, PR Telang, and M Wilson [RCIS 2015];
  Creativity, personality, crowdsourcing, and teamwork with PK Murukannaiah [RE 2016, RE 2017];
  App review mining with VT Dhinakaran, R Pulle, and PK Murukannaiah [RE 2018], and H Guo and Z Zhang (ongoing)
  Collective intelligence with AK Kalia, PK Murukannaiah, R Pandita, and H Du (ongoing);
  Analysis of privacy news with K Sheshadri and J Staddon [PST 2017];
  Preserving probe trajectory privacy with R Balu, B Xu, and M Stroila;
  Agile requirements evolution with S Ghaisas et al. [JSS 2013, MaRK 2013, MaRK 2011, MaRK 2010, RSSE 2010];

- Labmates at Multiagent Systems and Service-Oriented Computing Lab
- Science of Security Lablet at North Carolina State University
- Laboratory for Analytic Sciences
- Family and friends

Appendix

Arnor: A Method to Model Social Intelligence

$RQ_{\text{Social intelligence}}$: How can modeling social intelligence in a SIPA help deliver a social experience and respects its stakeholders' privacy?

Goal modeling: identifying a SIPA's stakeholders, their goals, and plans

Context modeling: identifying the social contexts in which a SIPA's stakeholders interact

- Context helps in deciding which goals to bring about or plans to execute

Social expectation modeling: identifying norms and sanctions that govern stakeholders' goals and plans

Social experience modeling: identifying a SIPA's actions that improve social experience, i.e., choosing plans, goals, and norms

Evaluation: Developer Study

Participants: 30 developers

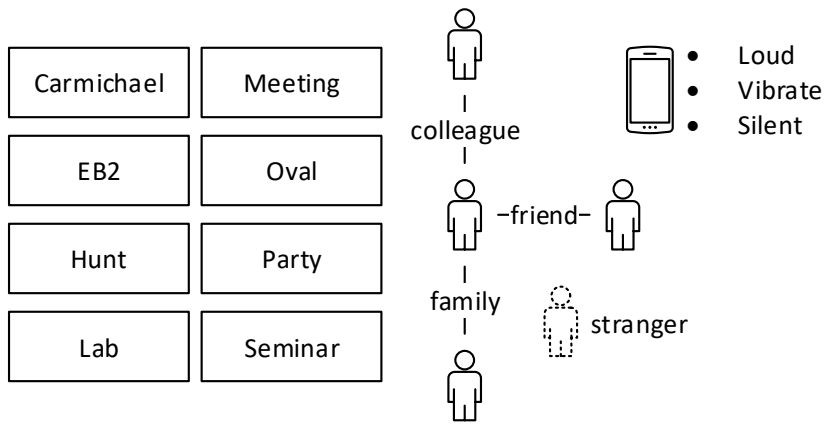
Mechanics: One factor; two alternatives

- Two groups (Arnor and Xipho, a prior method) balanced on skills developed RINGER SIPAs in six weeks
- Model, Implement, Test

Metrics:

- Coverage and correctness
- Time and difficulty to develop

Study Unit: RINGER SIPAs



Result

Developers who follow Arnor feel it is easier to develop a SIPA and expend less time, than those who follow Xipho

Evaluation: User Study (Simulations)

Developed RINGER SIPAs simulated in varying adaptation scenarios:

- Fixed norms
- Changing norms
- Changing context
- Changing sanction

Metrics:

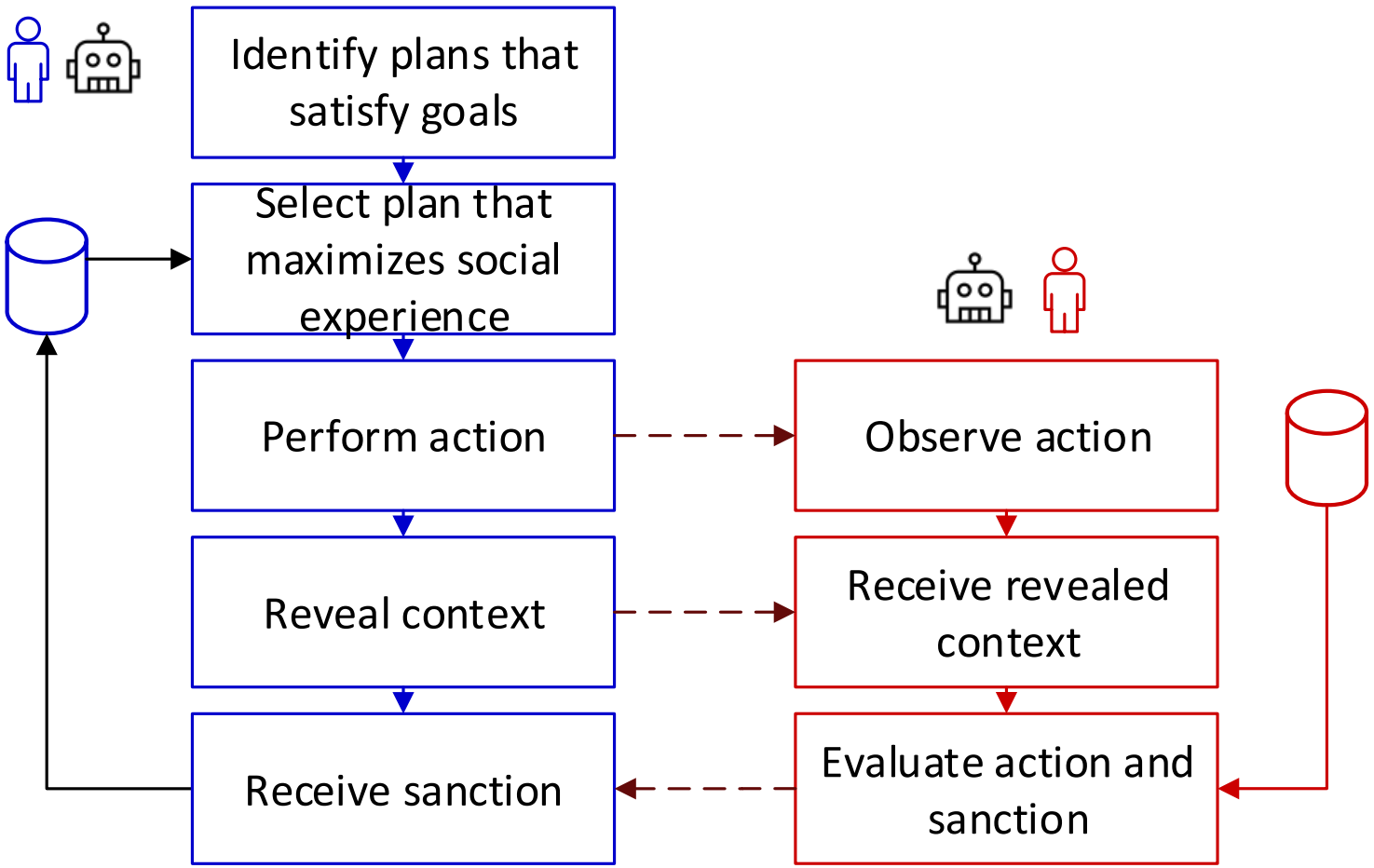
- Adaptability coverage and correctness
- Norm compliance
- Proportion of positive sanctions

Result

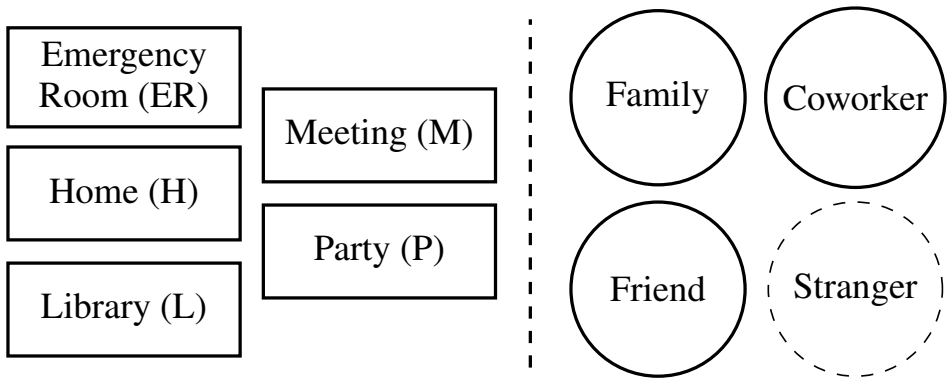
SIPAs developed using Arnor yield lower sanction proportions than SIPAs developed using Xipho (a previous approach)

Interaction and Learning in Poros

$RQ_{Context}$: How can SIPAs share deviation contexts and adapt to them, and learn contextually relevant norms?



Evaluation: The Ringer Environment



Agent Societies

- Pragmatic
- Considerate
- Selfish

Agent Types

- Fixed
- Sanctioning
- Poros

Evaluation: Social Simulations

Metric:

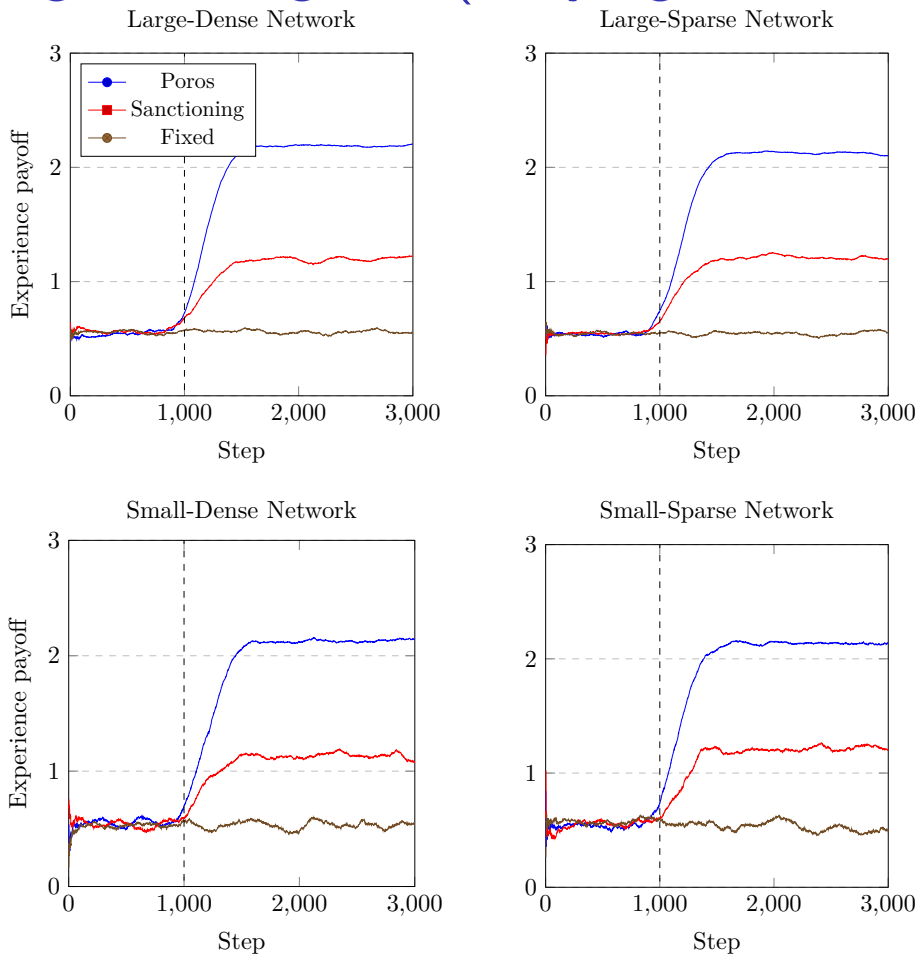
Social cohesion measures the proportion of agents that perceive actions as norm compliant. Higher the social cohesion, lower is the number of negative sanctions

Social experience measures the goal satisfaction delivered by an agent (computed by aggregating payoffs for all stakeholders)

Results

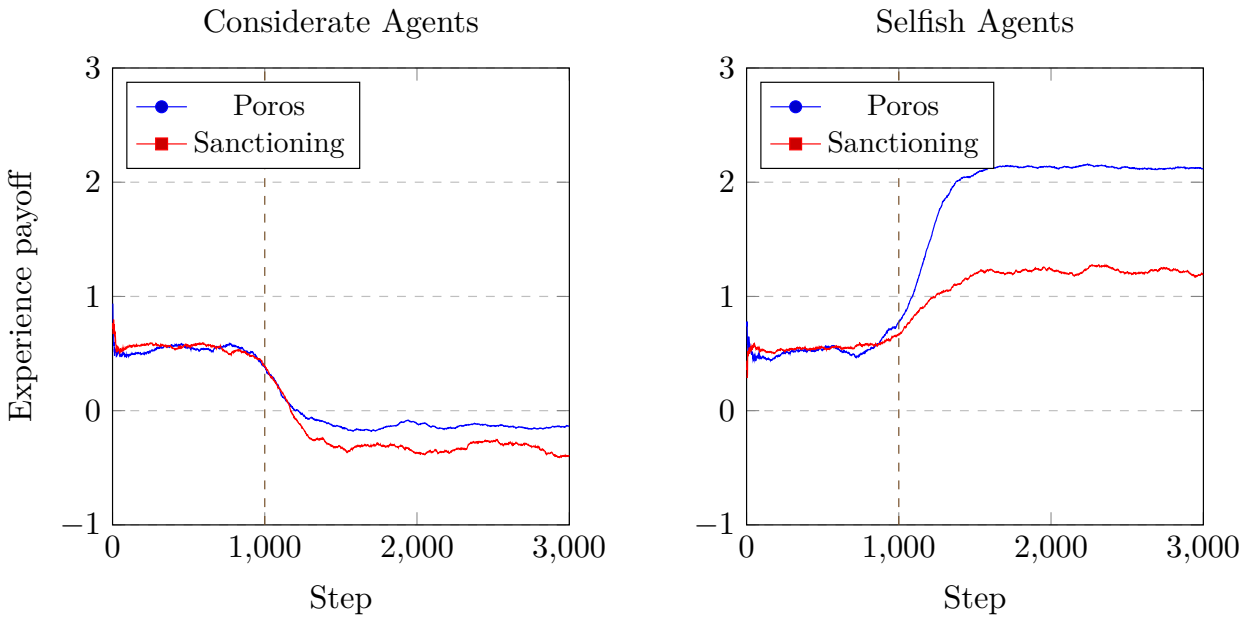
- Pragmatic society: Social cohesion and social experience offered by Poros agents are significantly better than those offered by Fixed and Sanctioning agents
- Considerate society: Average social experience drops for Sanctioning and Poros agents after they have gained enough confidence
- Selfish society: Plots are similar to those in the experiment with pragmatic agent societies, but with slightly lower stabilized values

Experiments on Pragmatic Agents (Varying Network Types)



Social cohesion and social experience offered by Poros agents are significantly better than those offered by Fixed and Sanctioning agents

Experiments on Considerate and Selfish Agents



- The average social experience drops for considerate Sanctioning and Poros agents after they have gained enough confidence
- Plots for selfish agents are similar to those in the experiment with pragmatic agents, but with slightly lower stabilized values

VIKOR Summary

1. Determine the best and worst numeric payoffs, f_x^* and f_x^- for each value preference x over the alternative actions y to bring about a goal. That is, $f_x^* = \max_y f_{xy}$, $f_x^- = \min_y f_{xy}$.
2. For each alternative action y, compute the weighted and normalized Manhattan distance [Opricovic and Tzeng, 2004]: $S_y = \sum_{x=1}^{n} w_x(f_x^* - f_{xy})/(f_x^* - f_x^-)$, where w_x is the weight for value preference x, which is subject to a stakeholder context and preferences over values. In particular, $S_y = 0$ when $f_x^* = f_x^-$.
3. Compute the weighted and normalized Chebyshev distance [Krause, 1973]: $R_y = \max_x[w_x(f_x^* - f_{xy})/(f_x^* - f_x^-)]$, where w_x is the weight for value preference x.
4. Compute $Q_y = k(S_y - S^*)/(S^- - S^*) + (1-k)(R_y - R^*)/(R^- - R^*)$, where $S^* = \min_y S_y$, $S^- = \max_y S_y$, $R^* = \min_y R_y$, $R^- = \max_y R_y$, and k is a weight of the strategy to maximum group or individual experience. We set $k = 0.5$ to select a consensus policy.
5. Rank alternative actions, sorting by the values S, R, and Q, in increasing order. The results are three ranked lists of actions.
6. Choose the alternative based on min Q as the compromise solution if it is better than the second best alternative by a certain threshold or also the best ranked as per S and R.

VIKOR Calculations

Policy Alternatives	Frank's Values				Hope's Values				S_y	R_y	Q_y
	Ple	Pri	Rec	Saf	Ple	Pri	Rec	Saf			
y_1 All	10	5	10	5	5	0	5	5	3.5	3	0.75
y_2 Common	5	5	5	10	5	0	5	5	0.4	3	1
y_3 Andrew	0	5	0	0	5	15	5	5	**0.3**	**1**	**0**
w_x	1	1	1	1	1	3	1	1			
f_x^*	1	0	1	1	0	1	0	0			
f_x^-	0	0	0	0	0	0	0	0			

$k = 0.5$, $w_{Hope-privacy} = 3$

Places in the Simulation

Place	Safe	Sensitive
Attending graduation ceremony	–	No
Presenting a conference paper	–	No
Studying in library	Yes	–
Visiting airport	Yes	–
Hiking at night	No	–
Being stuck in a hurricane	No	–
Visiting a bar with fake ID	–	Yes
Visiting a drug rehab center	–	Yes

Example Numeric Utility Matrix for a Stakeholder

Place	Companion	Policy	Value			
			Pleasure	Privacy	Recognition	Security
Graduation	Family	All	1	0	1	0
Conference	Co-workers	None	0	1	0	0
Library	Friends	All	1	0	0	0
Airport	Friends	Common	0	1	0	0
Hiking	Alone	All	1	0	0	1
Hurricane	Family	All	1	0	0	1
Bar	Alone	None	0	2	0	0
Rehab	Friends	None	0	2	0	0

Comparing Social Experience and Fairness for Mixed Privacy Attitudes

Strategy	Mean	Best	Worst	Fairness	p
S_{Ainur}	**1.361**	1.715	**0.767**	**1.05**	–
$S_{primary}$	1.286	1.789	0.579	0.83	<0.01
$S_{conservative}$	1.106	1.721	0.472	0.80	<0.01
$S_{majority}$	1.339	**1.836**	0.570	0.78	<0.01

Comparing Social Experience and Fairness for Majority Privacy Attitudes

Strategy	Fundamentalist				Pragmatist				Unconcerned			
	M.	B.	W.	F.	M.	B.	W.	F.	M.	B.	W.	F.
S_{Ainur}	1.535	1.664	**1.233**	**2.27**	**1.329**	1.531	**0.867**	**1.51**	**1.242**	1.457	**0.768**	**1.45**
$S_{pri.}$	1.506	1.766	1.082	1.46	1.253	1.592	0.679	1.10	1.129	1.466	0.584	1.13
$S_{cons.}$	1.366	1.745	1.059	1.46	1.093	1.519	0.608	1.10	0.870	1.338	0.454	1.34
$S_{maj.}$	**1.551**	**1.858**	1.007	1.18	1.318	**1.699**	0.575	0.89	1.176	**1.534**	0.518	0.98

Location Sharing Survey: Policy Selection

Companion	Check-in Policy			
	Share with all	Common friends	Companions	No one
Alone	o	o	o	o
Colleague	o	o	o	o
Friend	o	o	o	o
Family member	o	o	o	o
Crowd	o	o	o	o