Engineering Ethical Multiagent Systems

Nirav Ajmeri

Department of Computer Science
University of Bristol

Examples of Ethical and Privacy Concerns

Phone ringing: Intrusion of solitude, disapprobation and disclosure

US Senator's phone rings during important meeting - can you guess his embarrassing ringtone?

11:56, 17 APR 2015 | BY KARA O'NEILL

It's bad enough when your phone goes off in a silent meeting room, but it's even worse when your ringtone is as embarrassing as this one

A US Senator was left red-faced after his phone went off during a finance meeting - but it was his choice of ringtone that really raised some eyebrows.

Examples of Ethical and Privacy Concerns

Phone ringing: Intrusion of solitude, disapprobation and disclosure

US Senator's phone rings during important meeting - can you guess his embarrassing ringtone?

11:56, 17 APR 2015 | BY KARA O'NEILL

It's bad enough when your phone goes off in a silent meeting room, but it's even worse when your ringtone is as embarrassing as this one

A US Senator was left red-faced after his phone went off during a finance meeting - but it was his choice of ringtone that really raised some eyebrows.

- Intrusion
- Disapprobation
- Disclosure

[Westin, 1967; Solove, 2006]

Concepts

Sociotechnical system is a cyberphysical system where multiple stakeholders (humans, organizations, and agents) interact

Social norm governs the interactions between two stakeholders

- Commitment: *Meeting attendee are committed to keep their phones on silent*
- Prohibition: *Library prohibits visitors to answer phone calls*

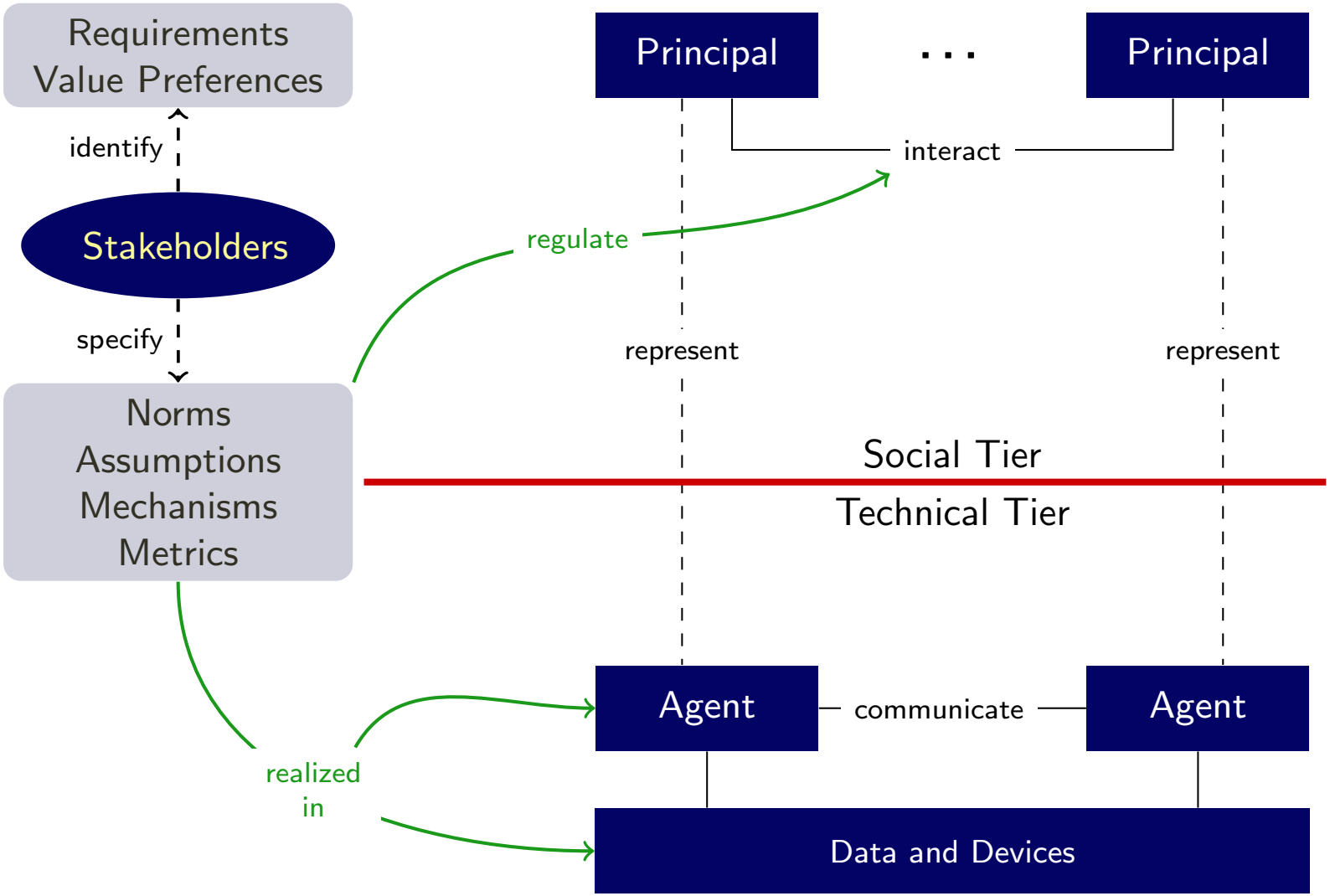
Deviation is a perceived violation of a norm

Social context is the circumstance under which the interaction takes place

Values are guiding principles of humans

- Ethics: subsumed in the theory of values
- Privacy: a value with an ethical import

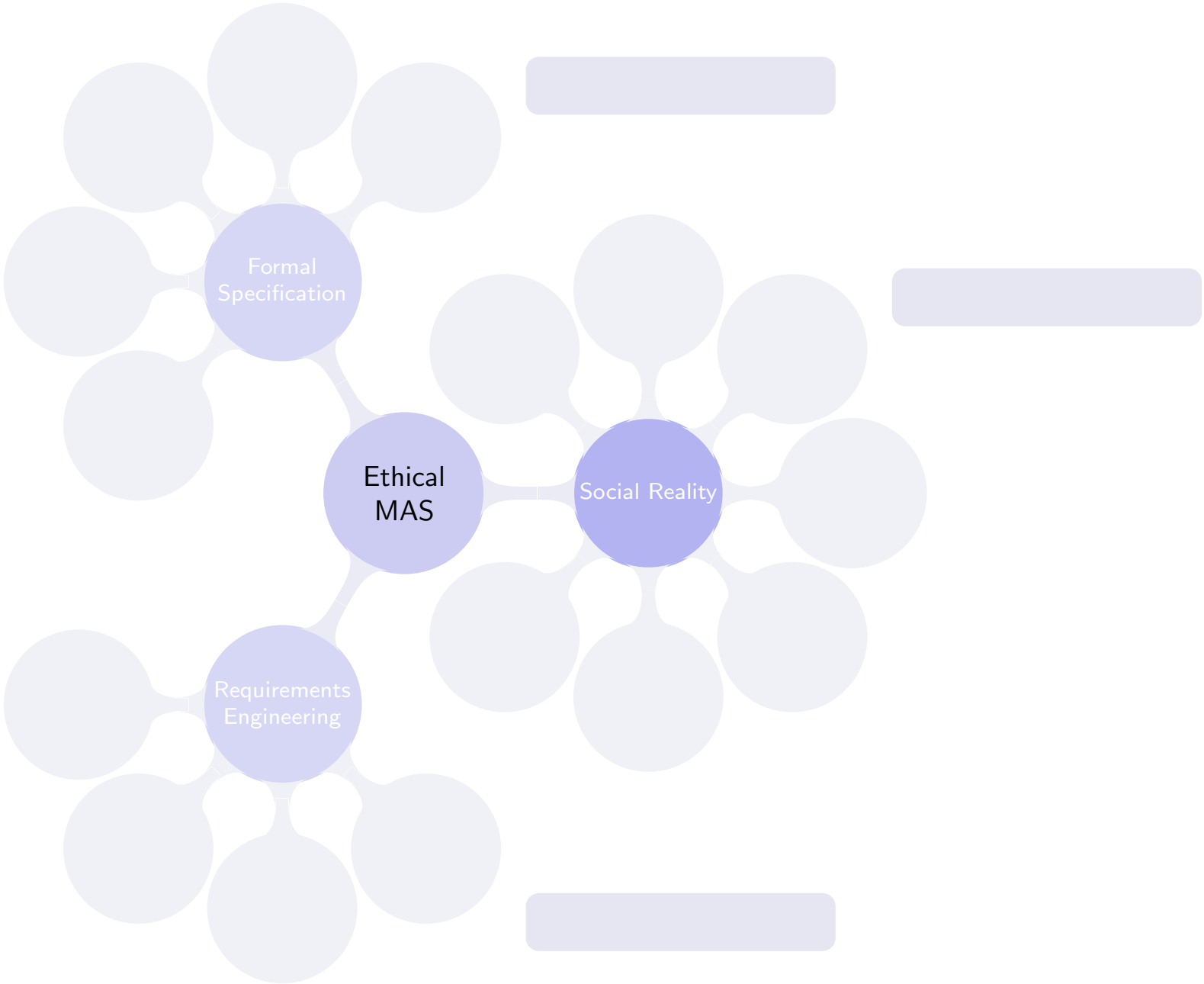
Schematic of a Sociotechnical System (STS)

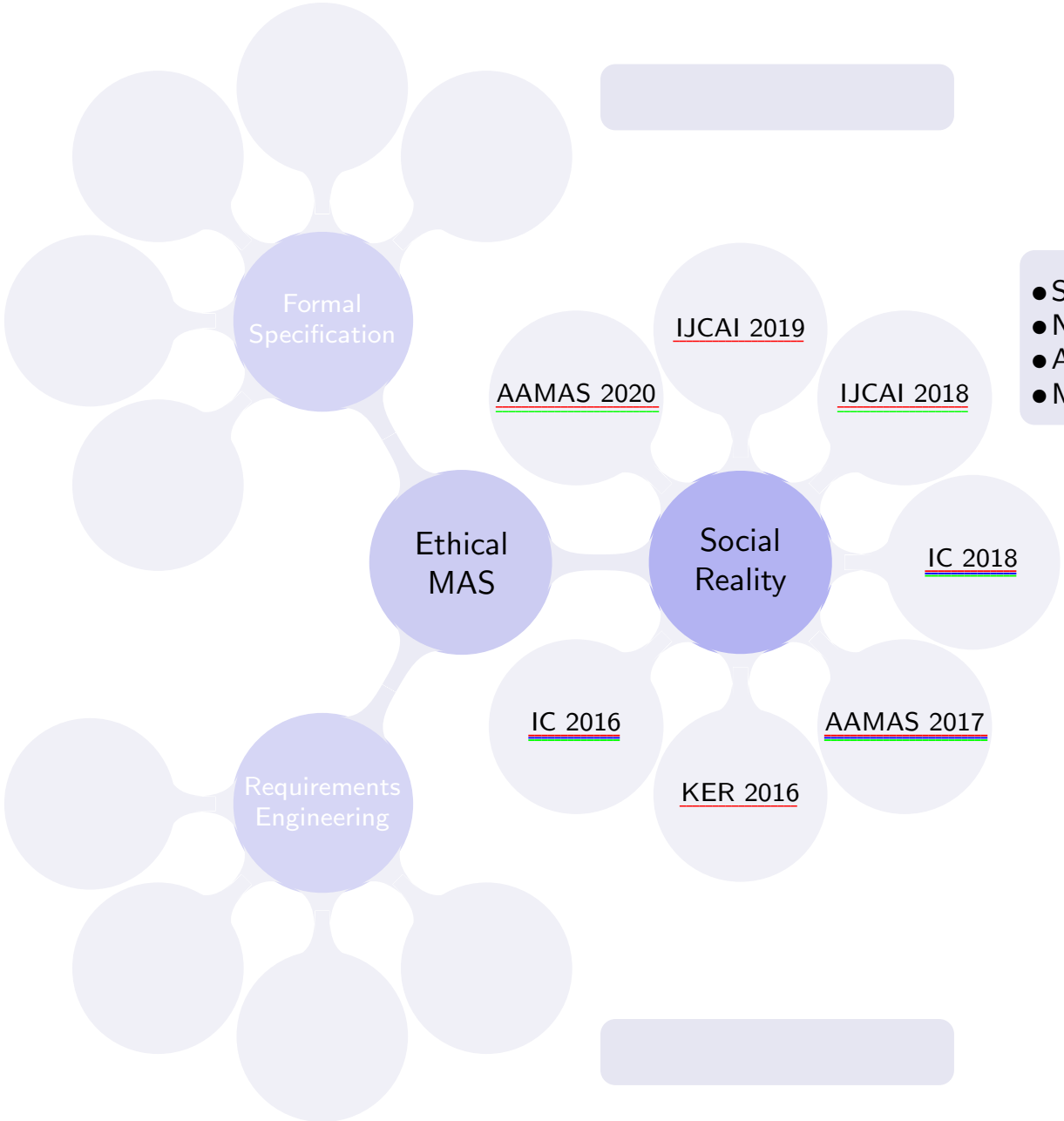


[IS 2016] Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. 2016. Revani: Revising and Verifying Normative Specifications for Privacy. IEEE Intelligent Systems (IS) 31(5) 8–15.

Research Objective

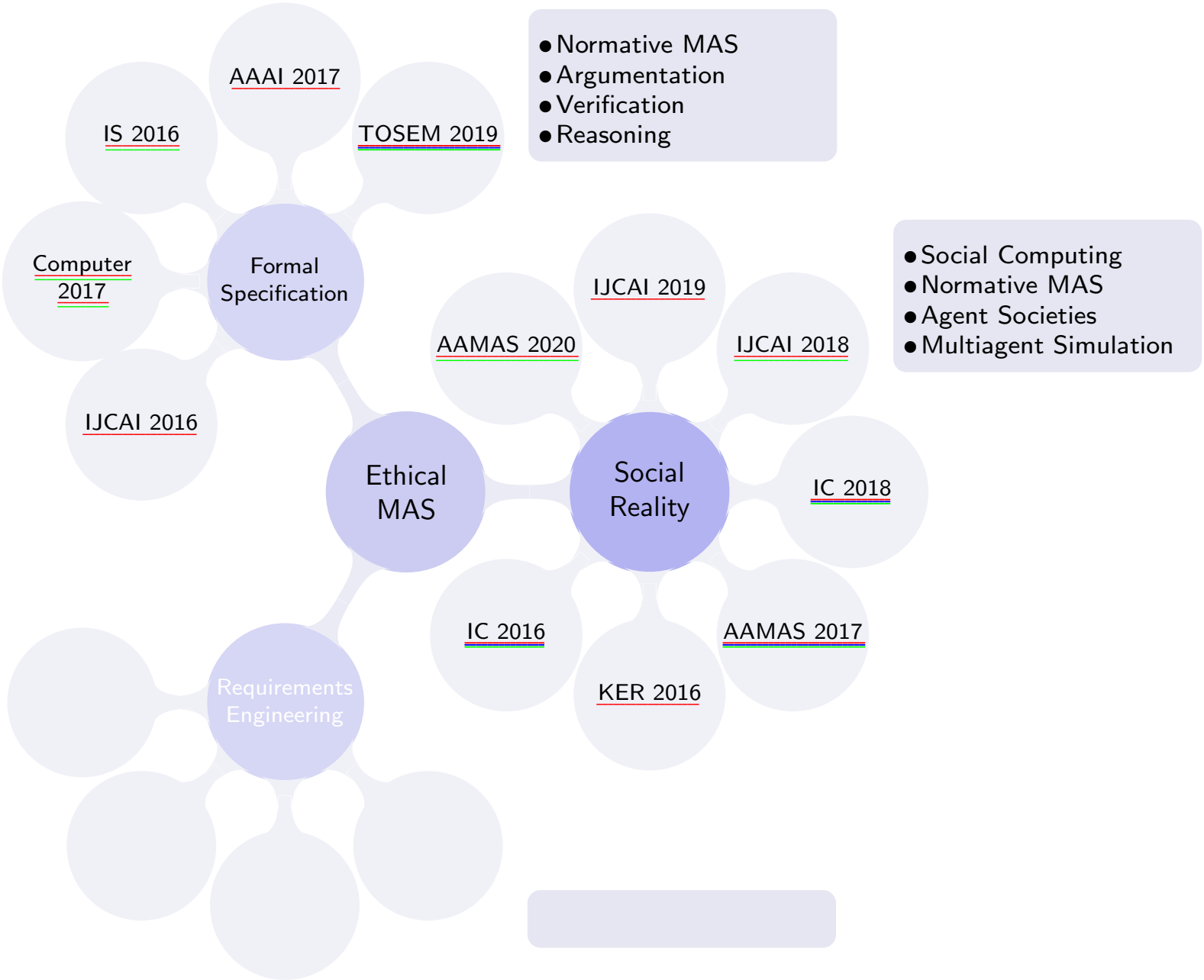
To aid software practitioners in realizing ethical multiagent systems (MAS) from the ground up by *developing secure computing technologies and infrastructure*


Formal Specification
Ethical MAS
Social Reality
Requirements Engineering

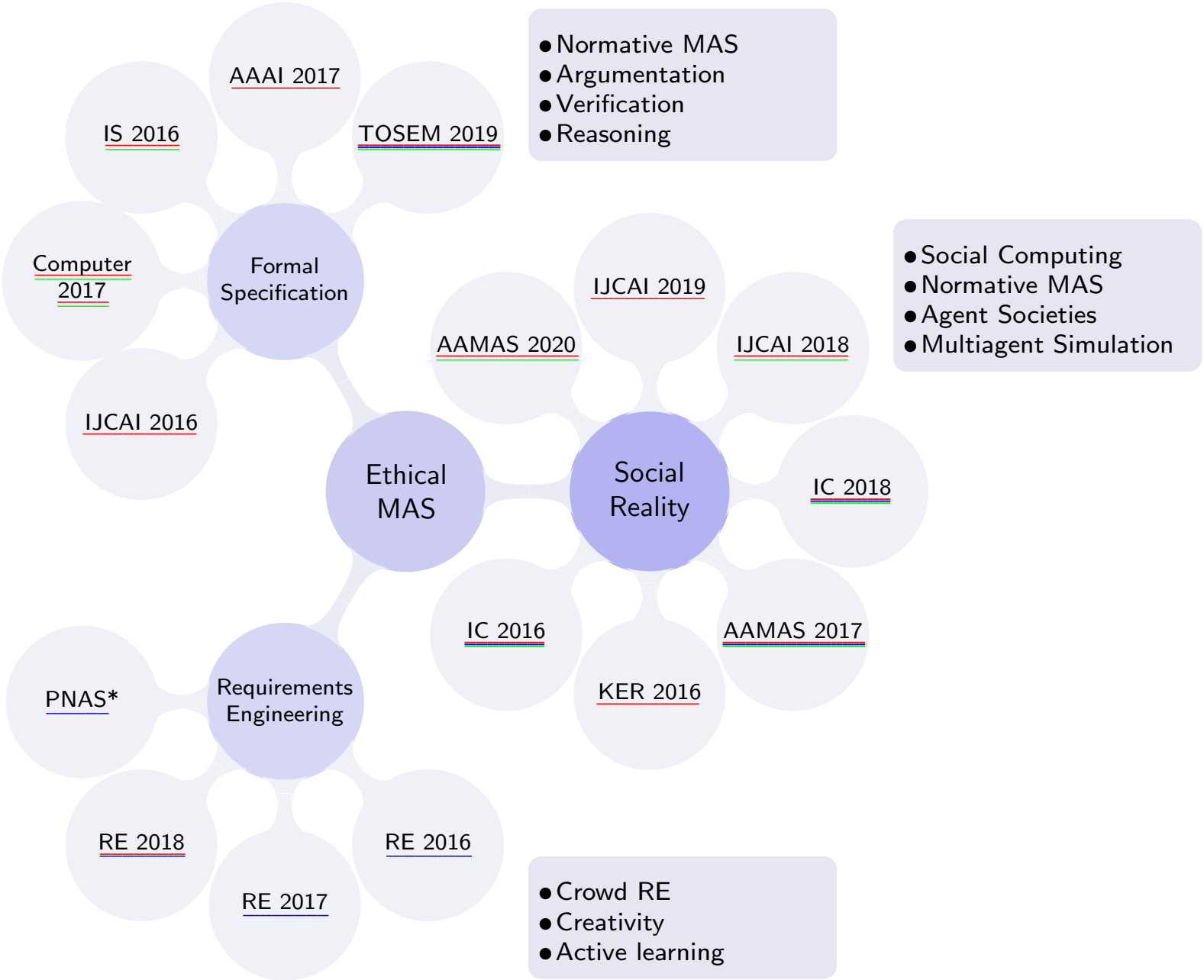


- Social Computing
- Normative MAS
- Agent Societies
- Multiagent Simulation

AI, SE, Privacy, * in-preparation



AI, SE, Privacy, * in-preparation



AI, SE, Privacy, * in-preparation

Socially Intelligent Personal Agent (SIPA)

A SIPA adapts to social context and supports meeting social norms

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

Challenge: Understanding Social Reality

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

Research Questions

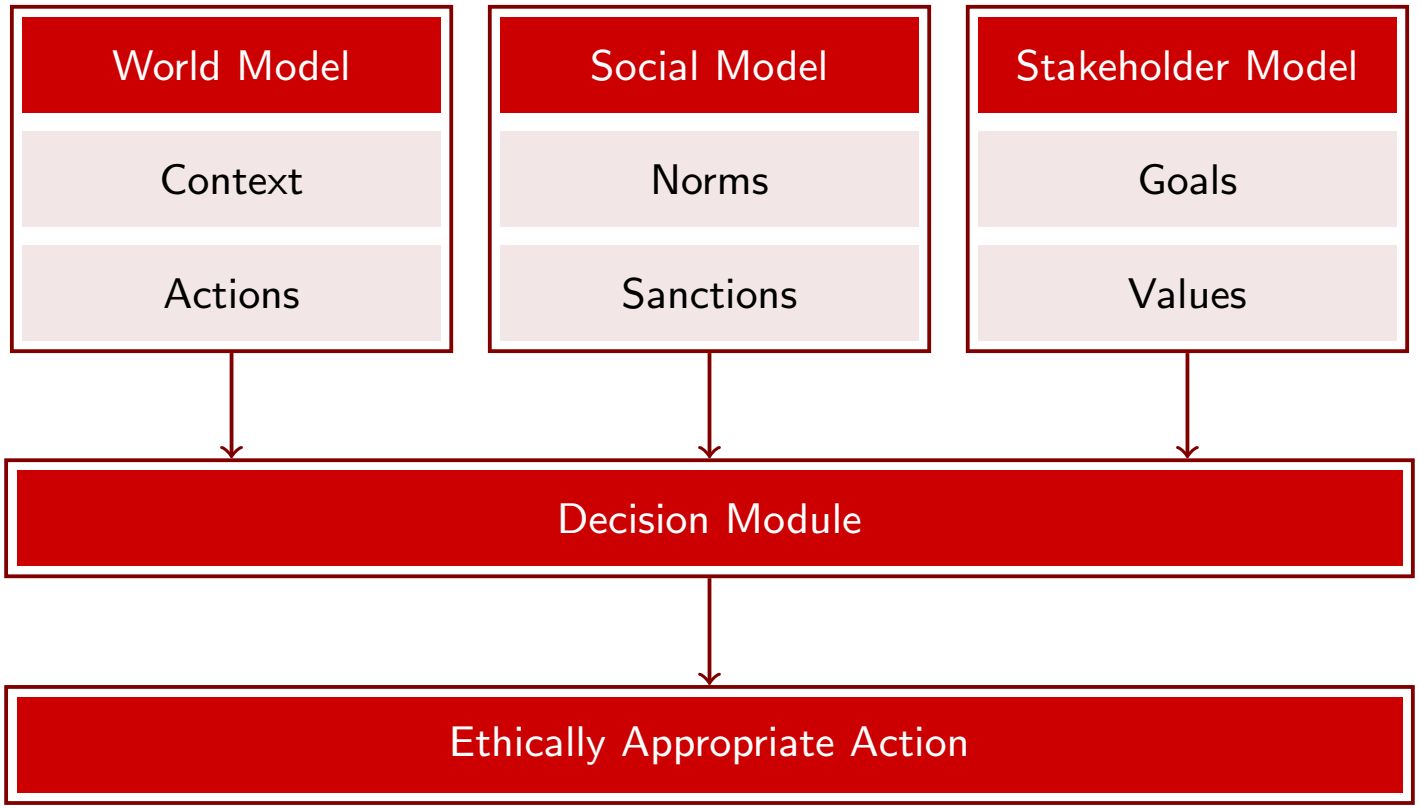
RQ Social intelligence: How can we model social intelligence in a SIPA to help it deliver a satisfactory experience to its stakeholders?

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

RQ Ethics: How can we enable SIPAs to select actions which are *just*, with respect to the applicable social norms and value preferences of their stakeholders?

RQ: Modeling Social Intelligence in a SIPA

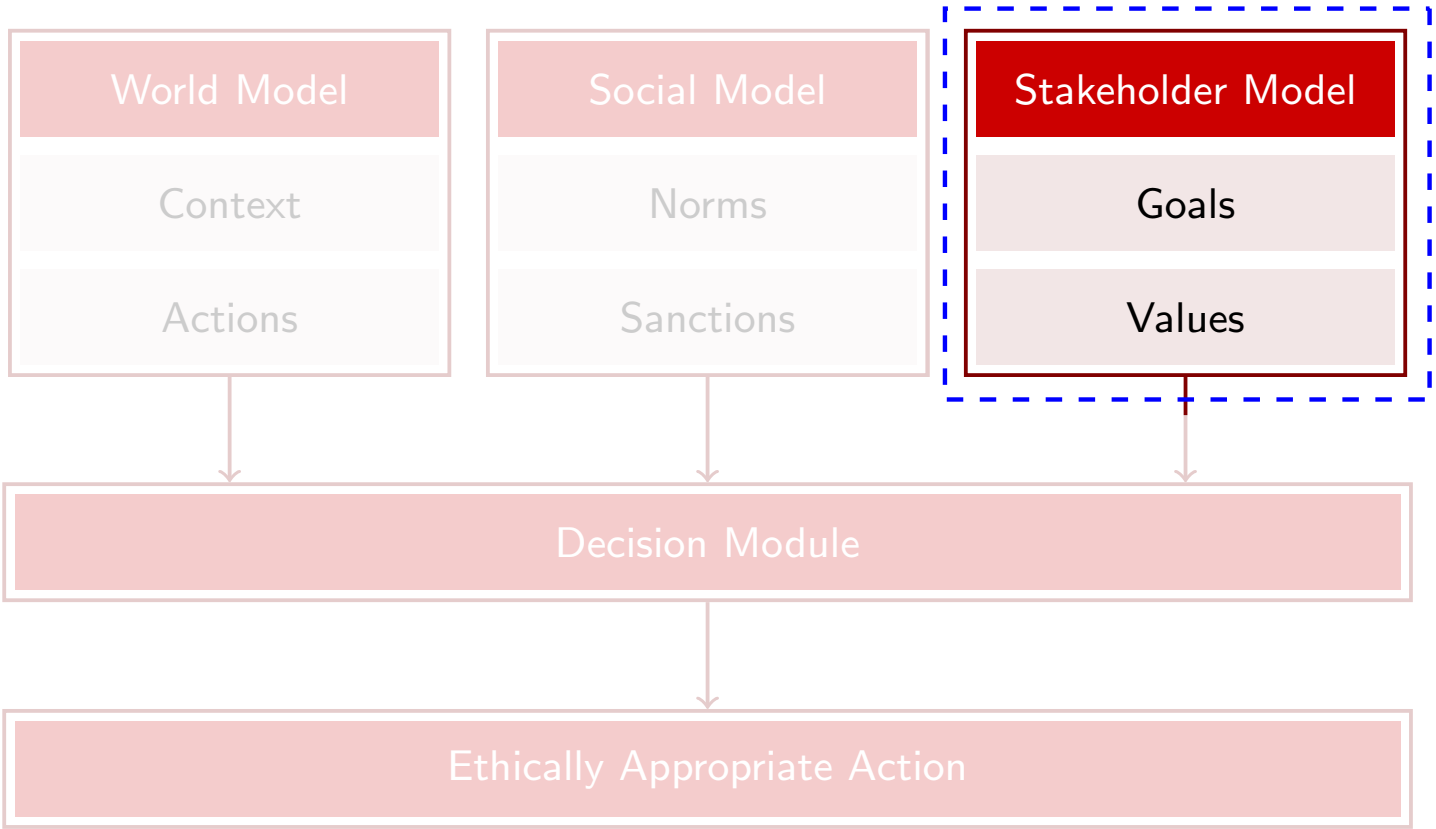
Arnor: A method to model social intelligence



[AAMAS 2020] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in Norm-Aware Agents. In Proc. AAMAS, 1–9.
[AAMAS 2017] Nirav Ajmeri, Pradeep K. Murukannaiah, Hui Guo, and Munindar P. Singh. Arnor: Modeling Social Intelligence via Norms to Engineer Privacy-Aware Personal Agents. In Proc. AAMAS, 230–238.

RQ: Modeling Social Intelligence in a SIPA

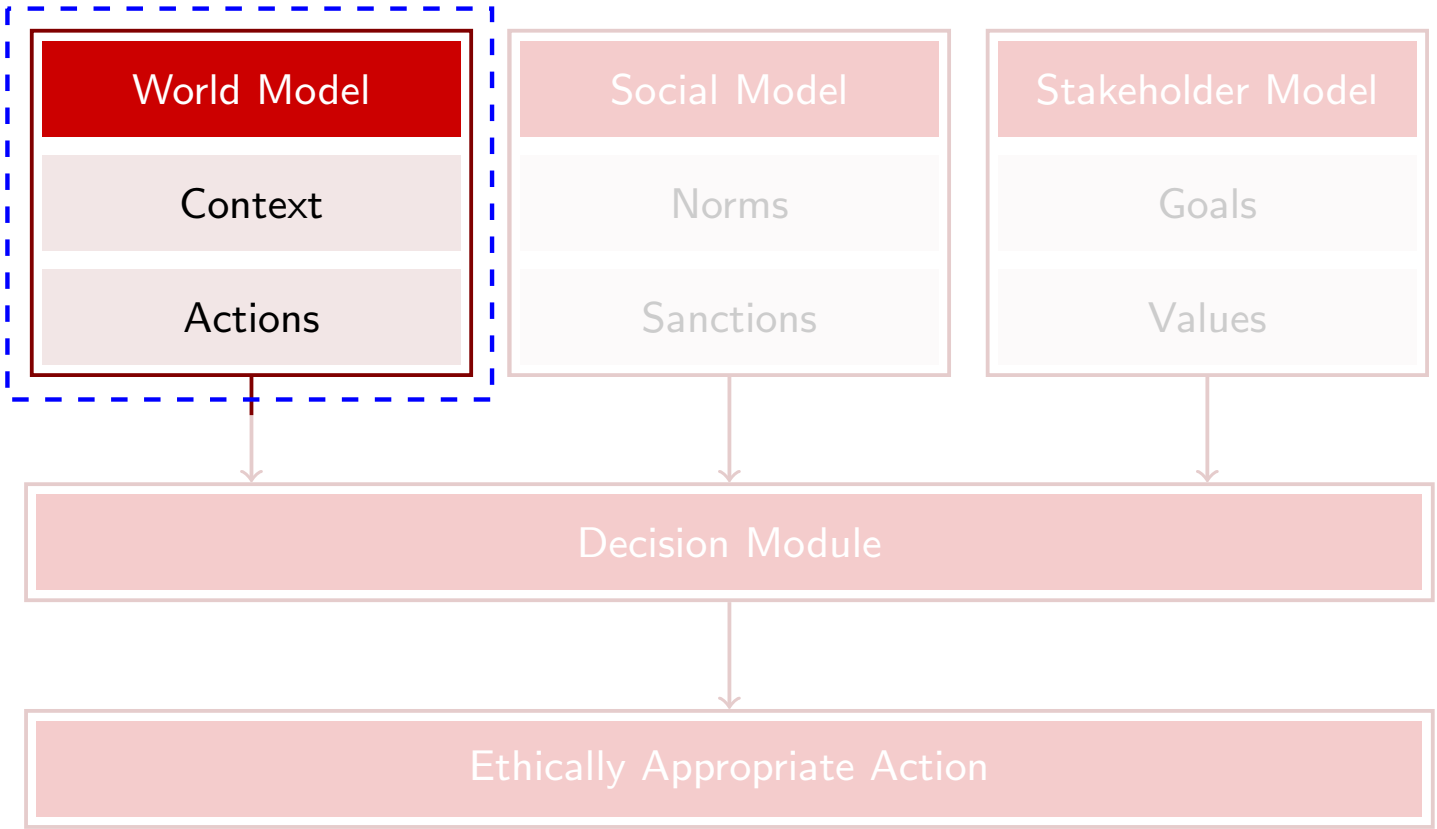
Stakeholder model: A SIPA's stakeholders and their goals and values



[AAMAS 2020] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in Norm-Aware Agents. In Proc. AAMAS, 1–9.
[AAMAS 2017] Nirav Ajmeri, Pradeep K. Murukannaiah, Hui Guo, and Munindar P. Singh. Arnor: Modeling Social Intelligence via Norms to Engineer Privacy-Aware Personal Agents. In Proc. AAMAS, 230–238.

RQ: Modeling Social Intelligence in a SIPA

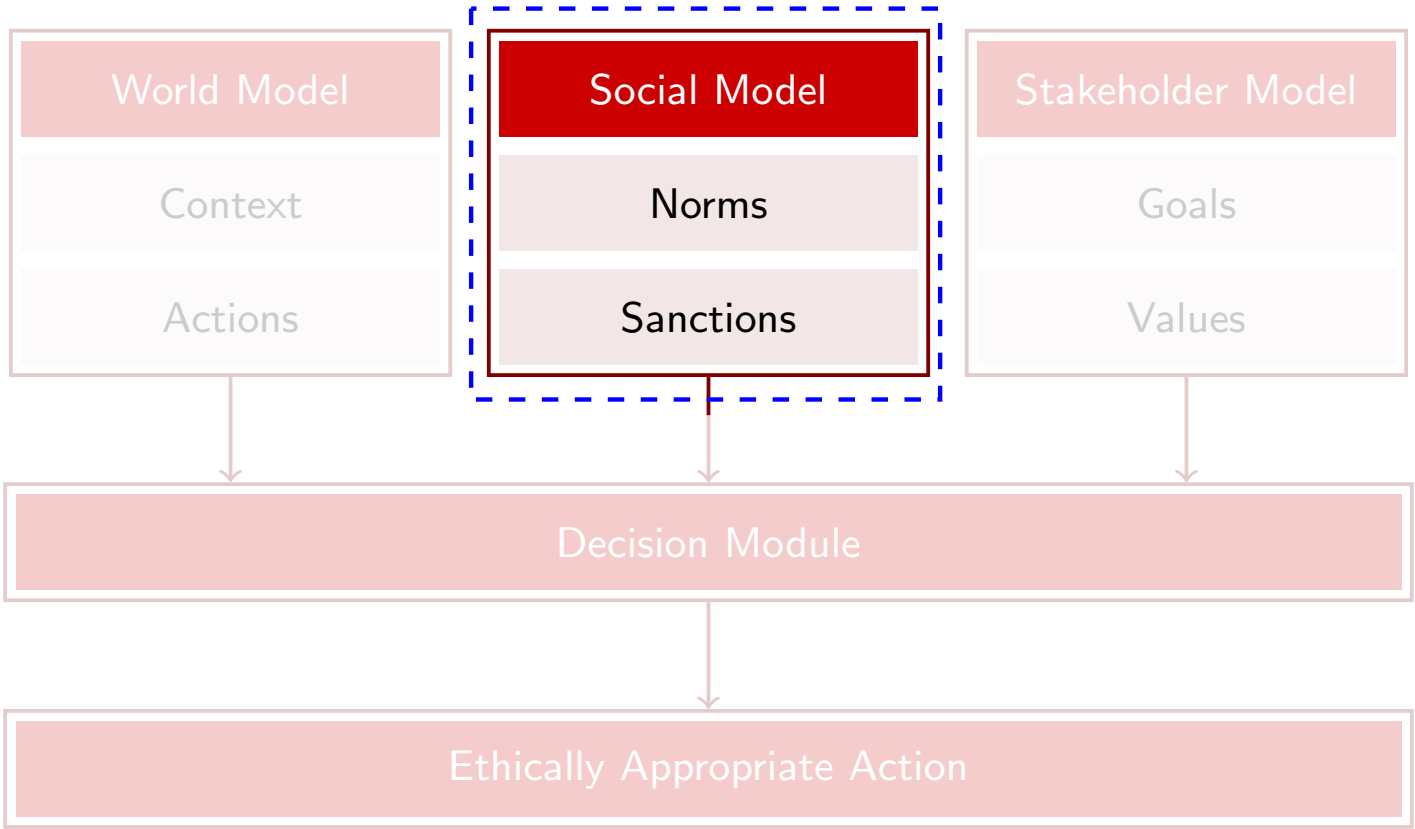
World model: Context in which a SIPA acts



[AAMAS 2020] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in Norm-Aware Agents. In Proc. AAMAS, 1–9.
[AAMAS 2017] Nirav Ajmeri, Pradeep K. Murukannaiah, Hui Guo, and Munindar P. Singh. Arnor: Modeling Social Intelligence via Norms to Engineer Privacy-Aware Personal Agents. In Proc. AAMAS, 230–238.

RQ: Modeling Social Intelligence in a SIPA

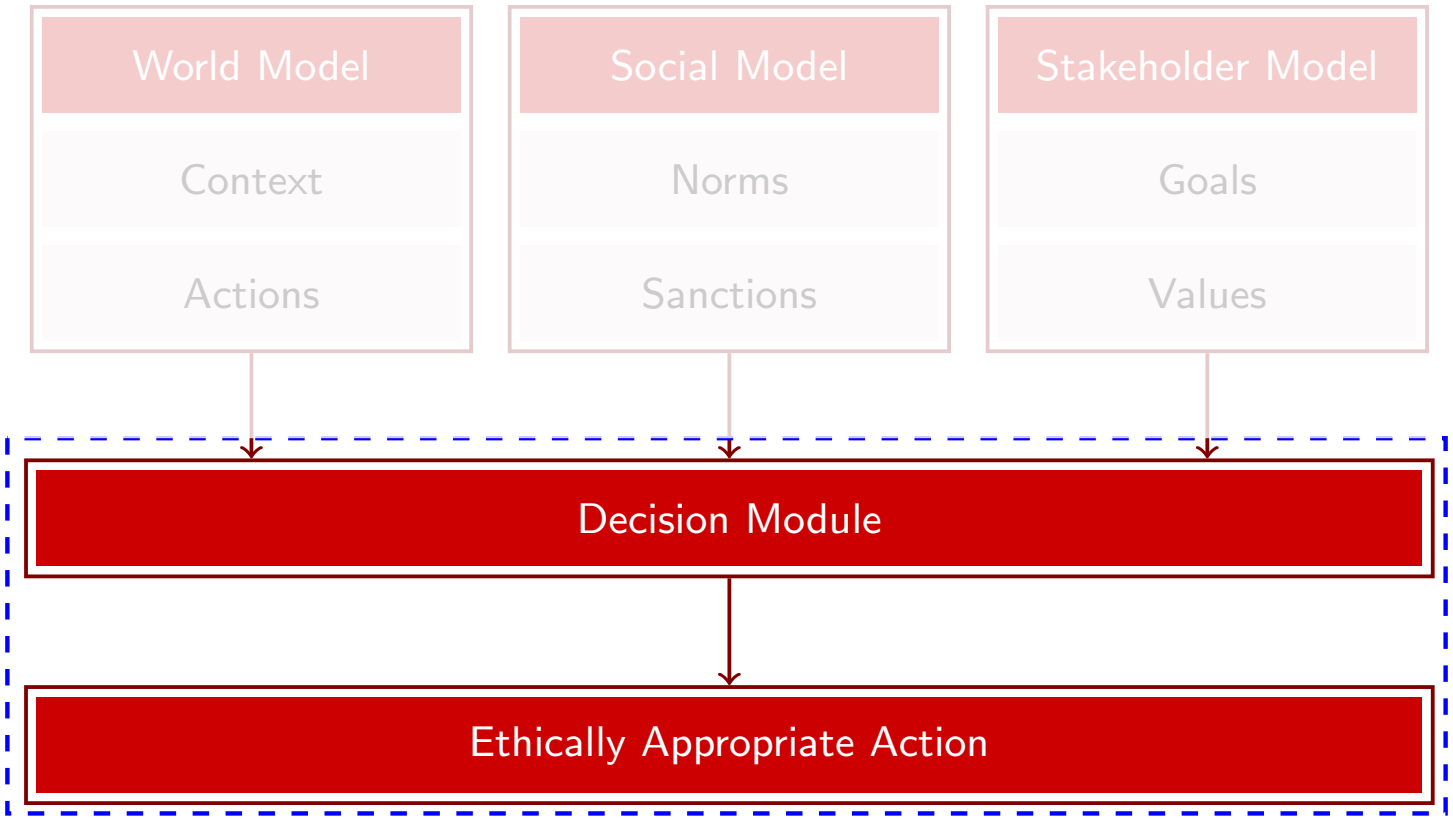
Social model: Norms governing the interactions in a society and the associated sanctions



[AAMAS 2020] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in Norm-Aware Agents. In Proc. AAMAS, 1–9.
[AAMAS 2017] Nirav Ajmeri, Pradeep K. Murukannaiah, Hui Guo, and Munindar P. Singh. Arnor: Modeling Social Intelligence via Norms to Engineer Privacy-Aware Personal Agents. In Proc. AAMAS, 230–238.

RQ: Modeling Social Intelligence in a SIPA

Social experience model: Identifying actions which yield a satisfactory social experience



[AAMAS 2020] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in Norm-Aware Agents. In Proc. AAMAS, 1–9.
[AAMAS 2017] Nirav Ajmeri, Pradeep K. Murukannaiah, Hui Guo, and Munindar P. Singh. Arnor: Modeling Social Intelligence via Norms to Engineer Privacy-Aware Personal Agents. In Proc. AAMAS, 230–238.

Evaluation: Developer Study

Result: Developers who model social intelligence find developing a SIPA to be easy. They expend less time in development than those who follow a prior method

Participants: 30 developers

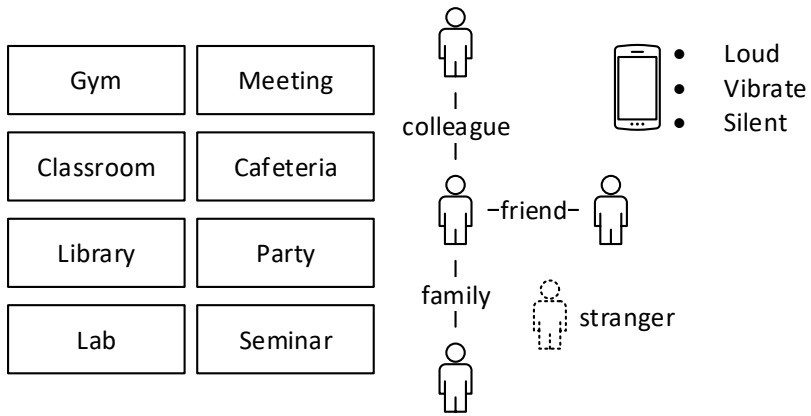
Mechanics: One factor; two alternatives

- 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

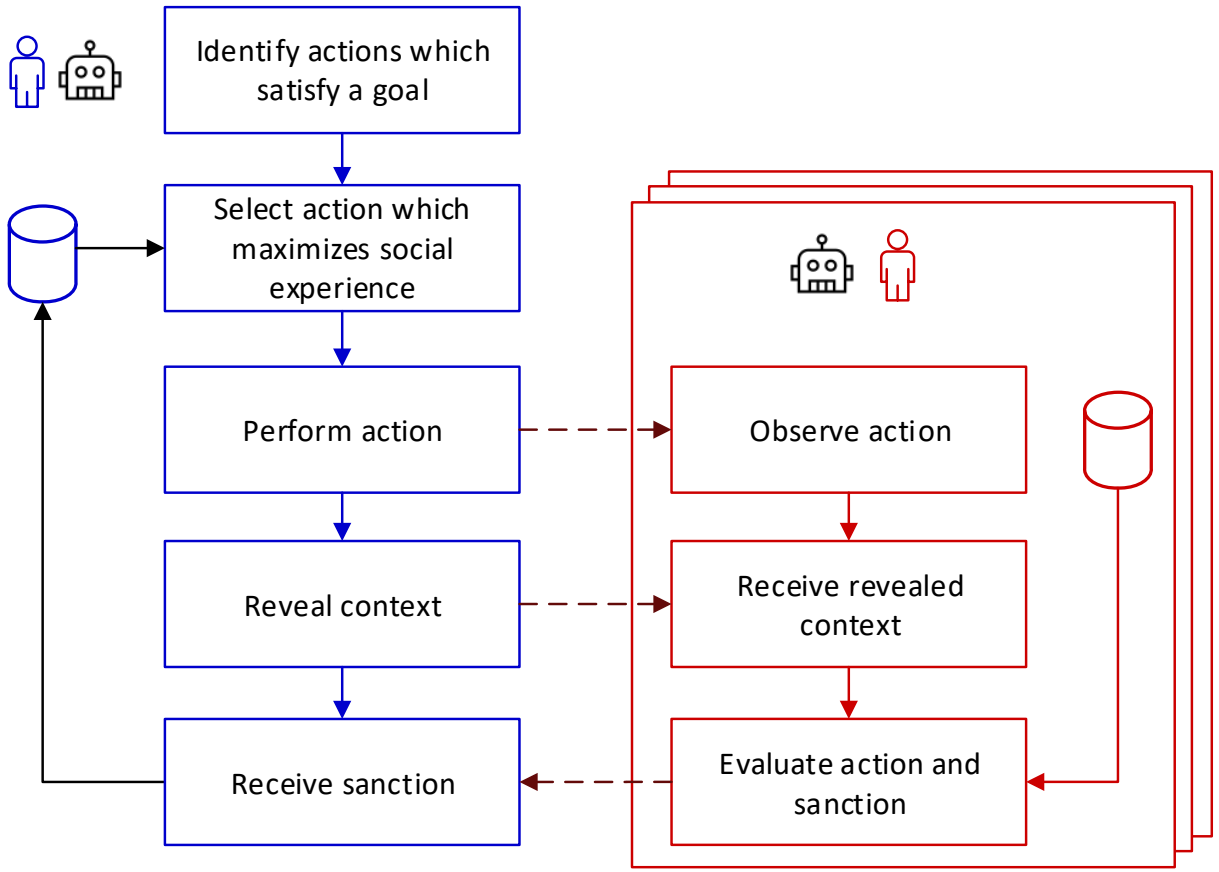


Possible Next Steps?

- Model affect and incorporate emotions with respect to norms
- Model guilt in socially intelligent agents

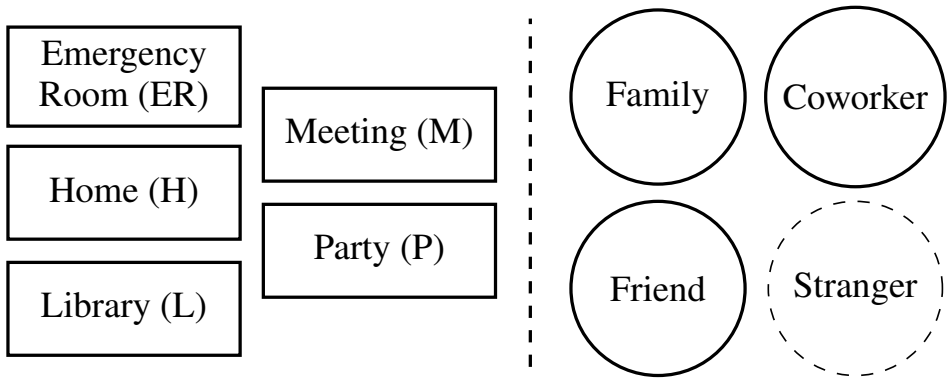
RQ: Understanding Social Context

Interaction and learning in Poros: Deviating SIPAs explain their deviations by sharing elements of their contexts



[IJCAI 2018] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Robust Norm Emergence by Revealing and Reasoning about Context: Socially Intelligent Agents for Enhancing Privacy. In Proc. IJCAI, 28–34.

Evaluation: The Ringer Environment



Agent societies:

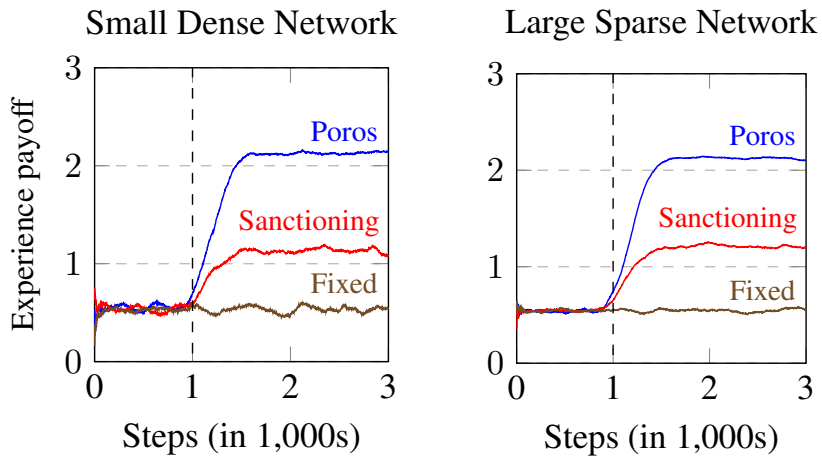
- Pragmatic
- Considerate
- Selfish

Learning strategies:

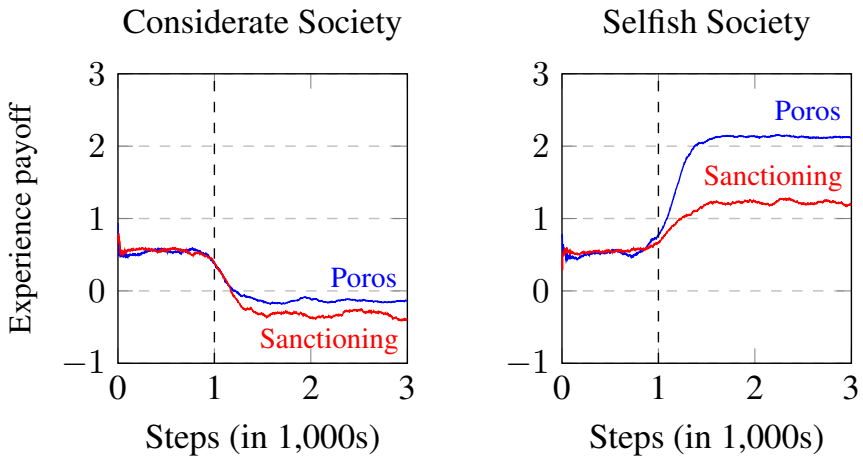
- Fixed
- Sanctioning
- Poros

Experiments on Pragmatic Agents (Varying Network Types)

Result: Social cohesion and social experience offered by Poros agents (which understand social context) are significantly better than those offered by Fixed and Sanctioning agents



Experiments on Considerate and Selfish Agents

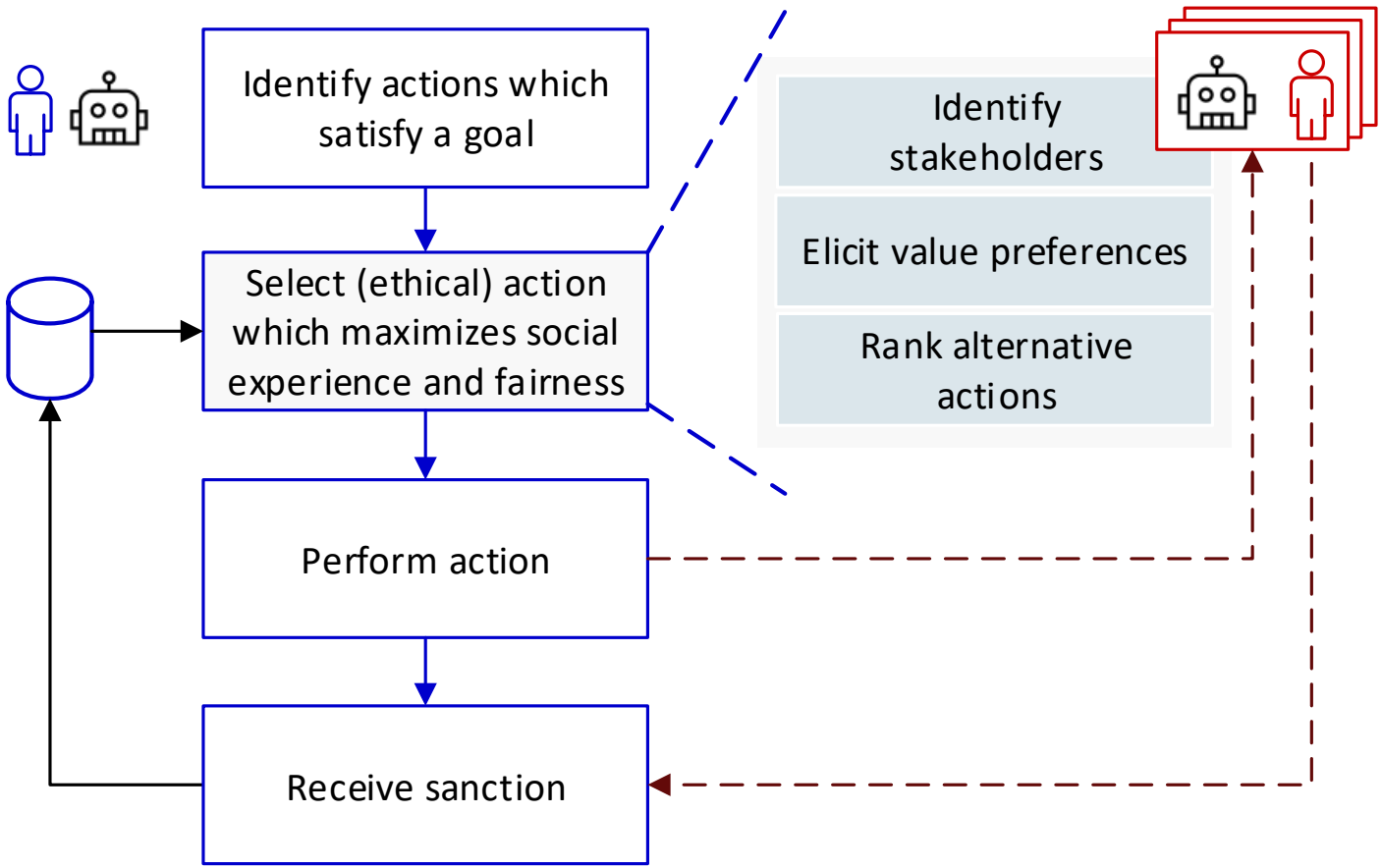


Possible Next Steps?

- Understand abstractions of the shared contexts
- Support white lies to promote privacy and social cohesion

RQ: Reasoning about Stakeholders' Value Preferences

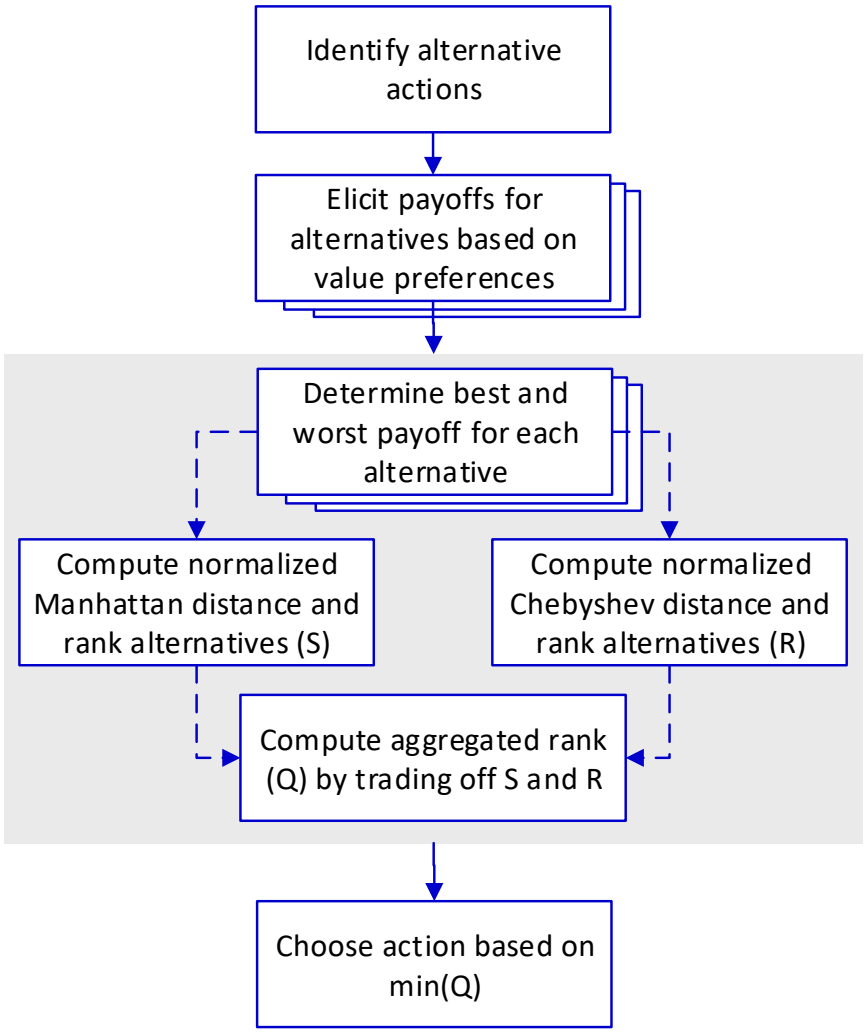
Interaction and decision making in Elessar: SIPAs aggregate value preferences of their stakeholders



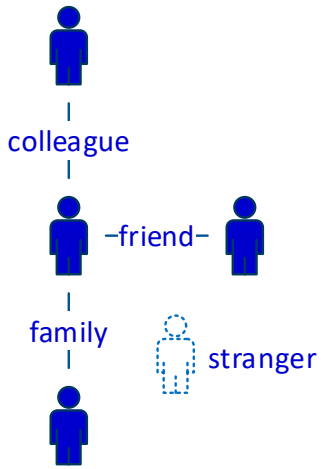
[AAMAS 2020] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in Norm-Aware Agents. In Proc. AAMAS, 1–9.

Choosing an Ethical Action

Elessar SIPAs adapt a multicriteria decision making method (VIKOR) to select ethically appropriate action — balancing *utilitarianism* and *egalitarianism*



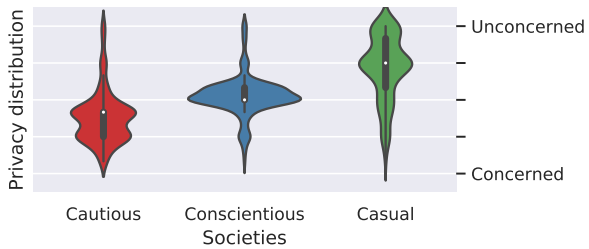
Evaluation: The Context-Sharing Environment



Simulated societies:

- Mixed
- Cautious
- Conscientious
- Casual

Privacy attitude:



Decision-making strategies:

$S_{Elessar}$: Policy based on VIKOR

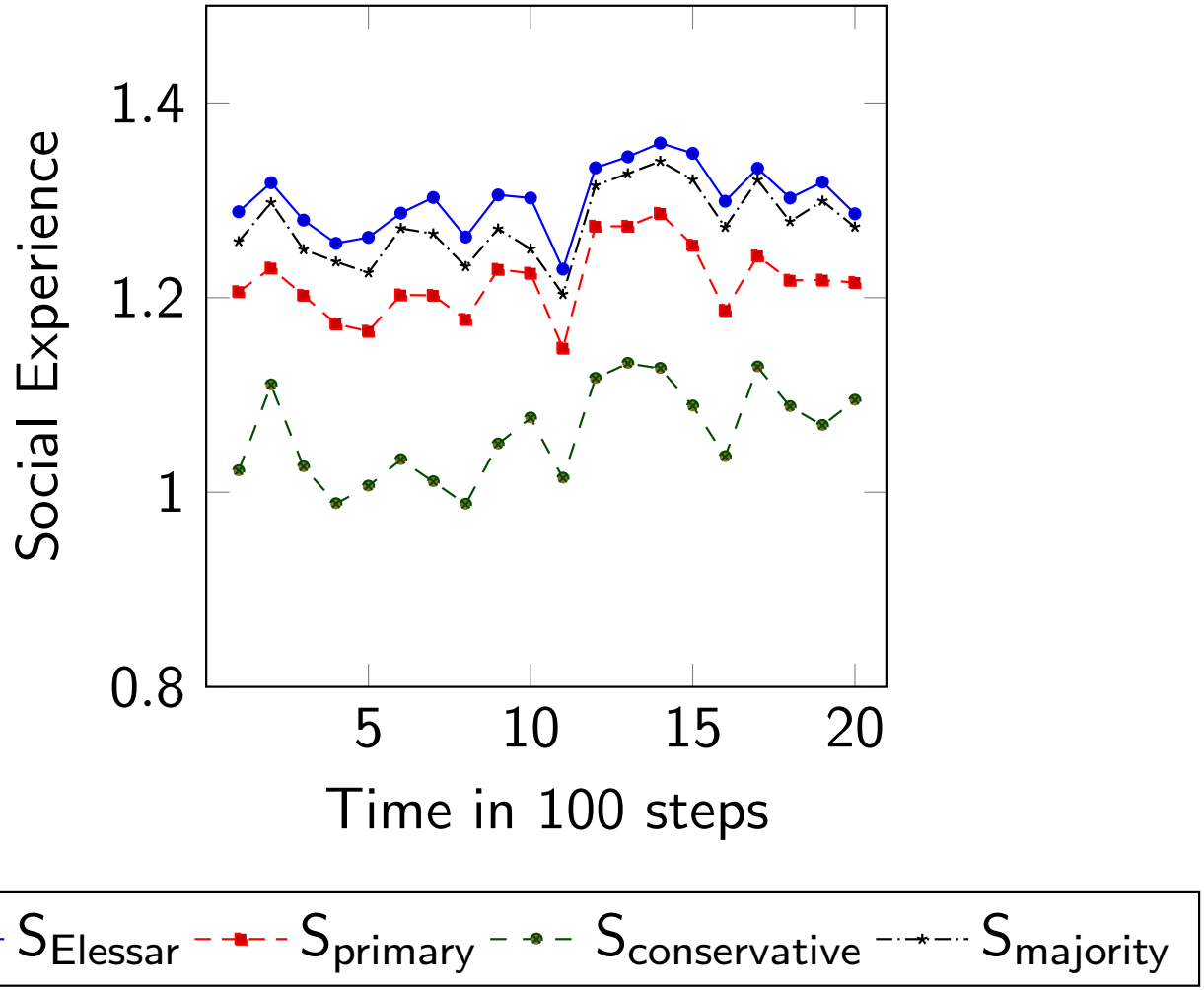
$S_{primary}$: Primary user's preference

$S_{conservative}$: Least privacy-violating

$S_{majority}$: Most common

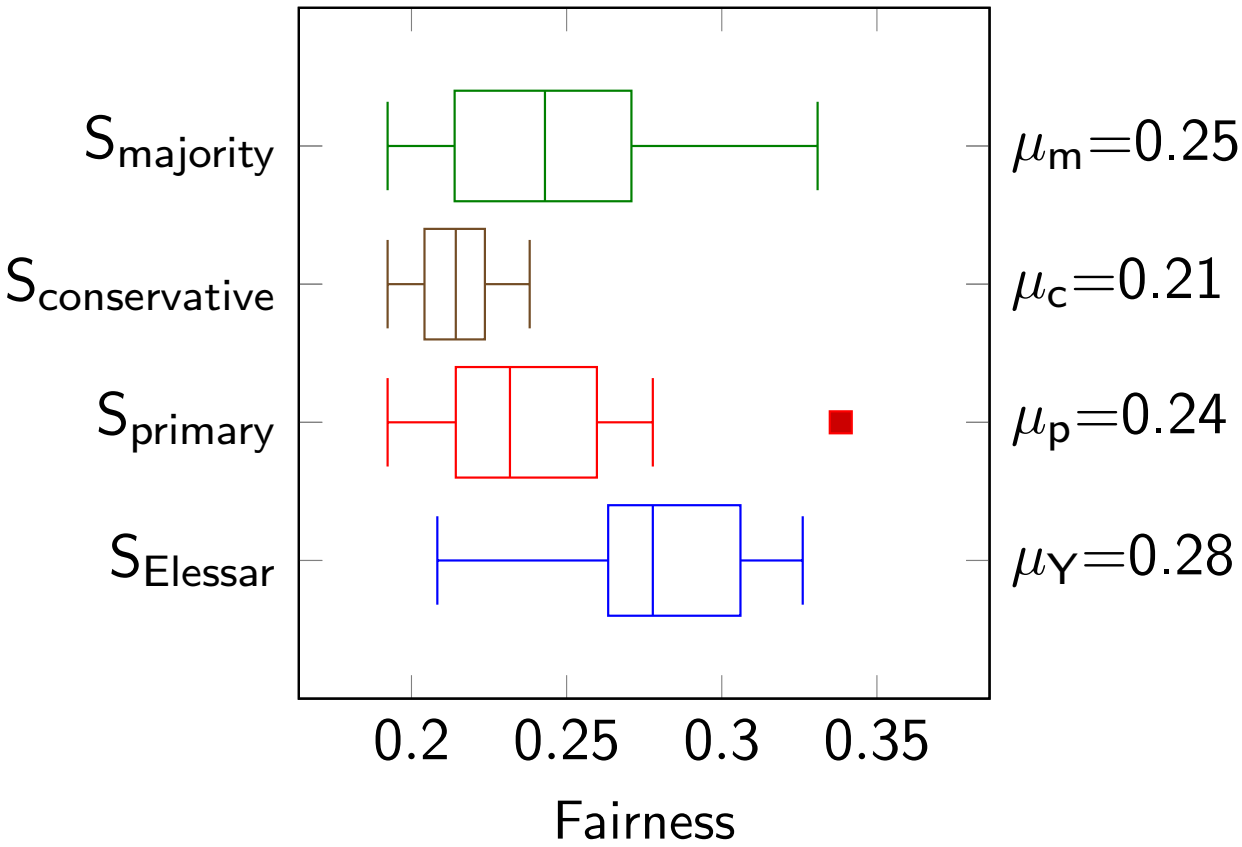
Experience: Experiment with Mixed Privacy Attitudes

Result: Elessar SIPAs (which reason about value preferences) yield higher social experience ($p < 0.01$; Glass' $\Delta > 0.8$ indicating large effect size) than the baselines



Fairness: Experiment with Mixed Privacy Attitudes

Result: Elessar SIPAs (which reason about value preferences) give significantly better ($p < 0.01$) fairness with large effect size (Glass' $\Delta > 0.8$) than the baseline methods



Possible Next Steps?

- Infer stakeholder's values and value preferences
- Formally verify if normative specification is optimal

Summary

- Ethics inherently involves looking beyond one's self interests
- Ethical considerations apply in mundane settings—anywhere agents of multiple stakeholders interact
- A multiagent understanding of ethics can provide a foundation for a science of security and privacy

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

Possible Directions (for Future Dissertations!)

- (Ethical) Artificial Intelligence

  Social reality: Guilt, white lies, and affect in personal agents

  Formal specification: Argumentation and value-based reasoning
- (Secure) Software Engineering

  Creativity: CrowdRE for privacy requirements

  Social reality: RE for secure and ethical systems
- (Usable) Privacy

  Social reality: Middleware to support ethical decision-making

  Social reality: Usable security, privacy, and ethics

Recent Collaborators

- Dr. Munindar Singh — North Carolina State University, US
- Dr. Laurie Williams — North Carolina State University, US
- Dr. Pradeep Murukannaiah — TU Delft, Netherlands
- Dr. Özgür Kafalı — University of Kent, UK
- Dr. Catholijn Jonker — TU Delft, Netherlands
- Dr. Pankaj Telang — SAS Institute Inc., US
- Dr. Anup Kalia — IBM Research, US
- Dr. Zhe Zhang — IBM Watson, US
- Dr. Chung-Wei Hang — IBM Watson, US
- Dr. Mehdi Mashayekhi — eBay Inc., US
- Dr. George List — North Carolina State University, US
- Dr. Jessica Staddon — Google, US

`https://niravajmeri.github.io/`
`nirav.ajmeri@gmail.com`

Appendix

Arnor: A Method to Model Social Intelligence

$RQ_{\text{Social intelligence}}$: How can we model social intelligence in a SIPA to help it deliver a satisfactory experience to its stakeholders?

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

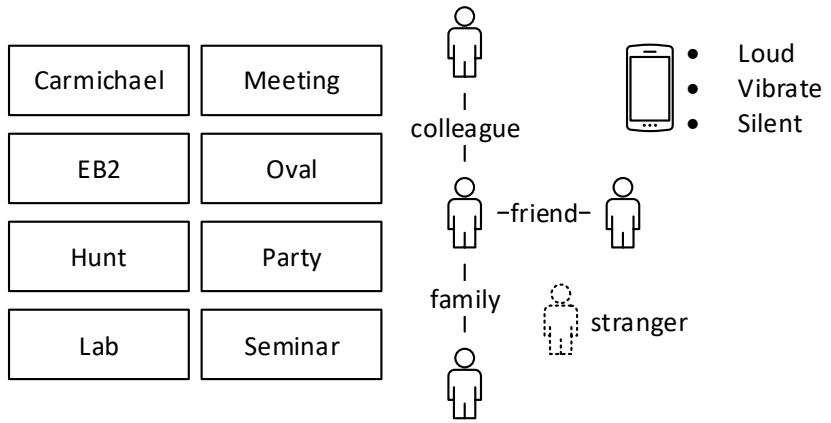
Mechanics: One factor; two alternatives

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

Metrics:

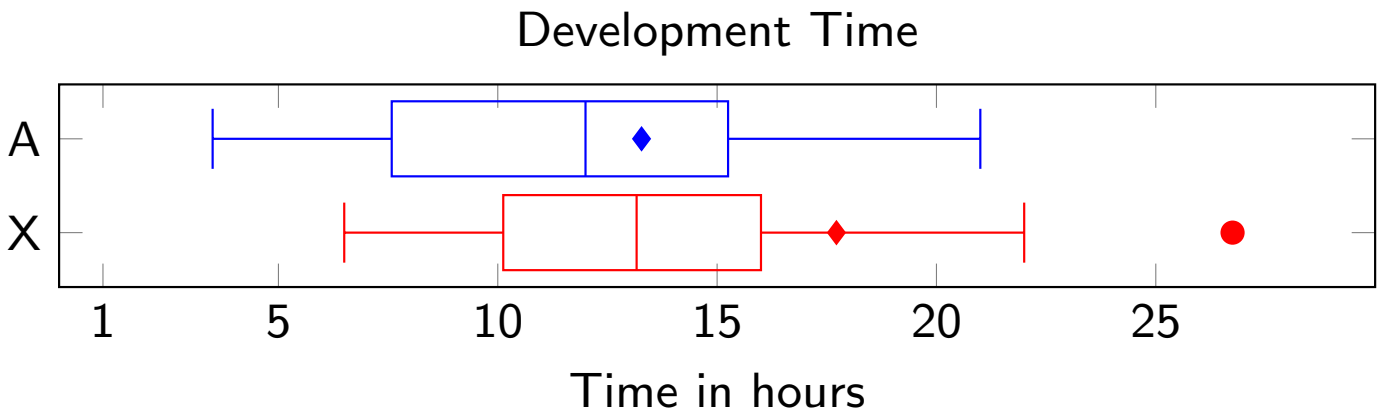
- Coverage and correctness
- Time and difficulty to develop

Study Unit: RINGER SIPAs

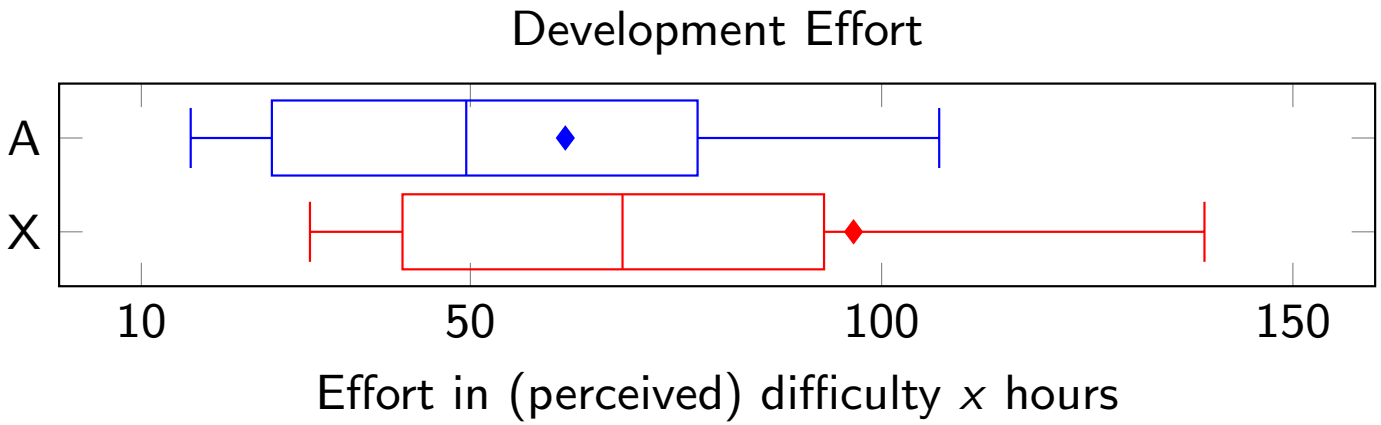


Developer Study

Result: Developers who follow Arnor spend less time and less effort to develop a SIPA, than those who follow Xipho (a previous approach)



$\mu_A < \mu_X,\ p = 0.39$



$\mu_A < \mu_X,\ p = 0.31$

Evaluation: User Study (Simulations)

Developed RINGER SIPAs simulated in varying adaptation scenarios:

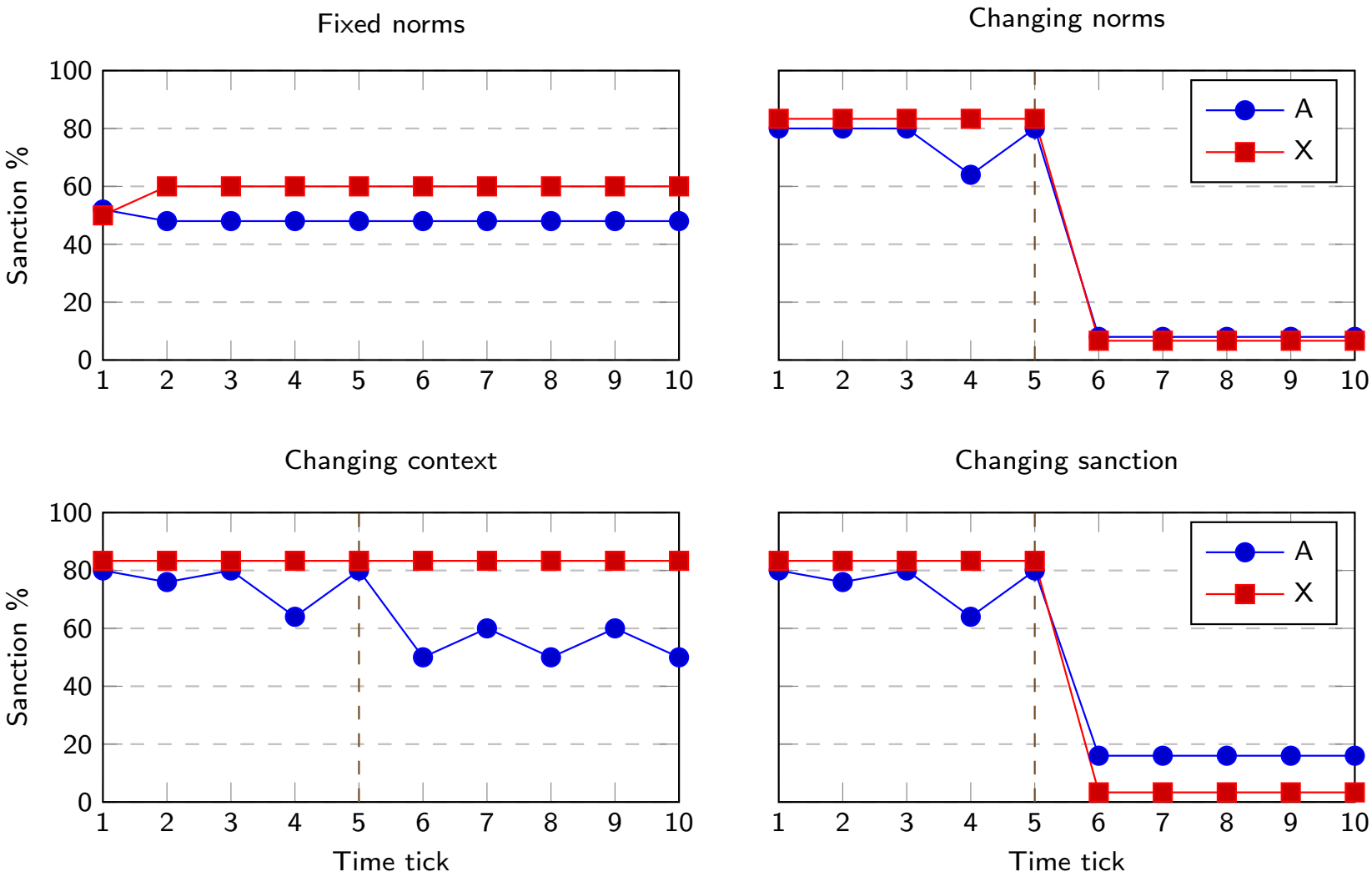
- Fixed norms
- Changing norms
- Changing context
- Changing sanction

Metrics:

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

Simulation Experiments

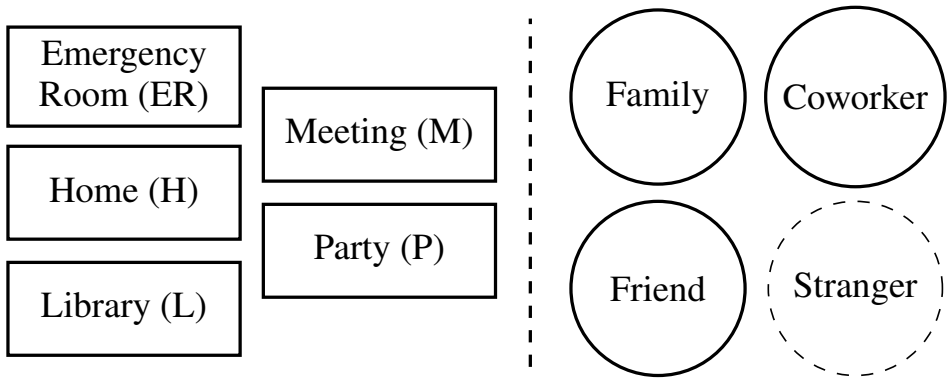
Result: SIPAs developed using Arnor yield lower sanction proportions than SIPAs developed using Xipho (a previous approach)



Interaction and Learning in Poros

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

Evaluation: The Ringer Environment



Agent Societies

- Pragmatic
- Considerate
- Selfish

Agent Types

- Fixed
- Sanctioning
- Poros

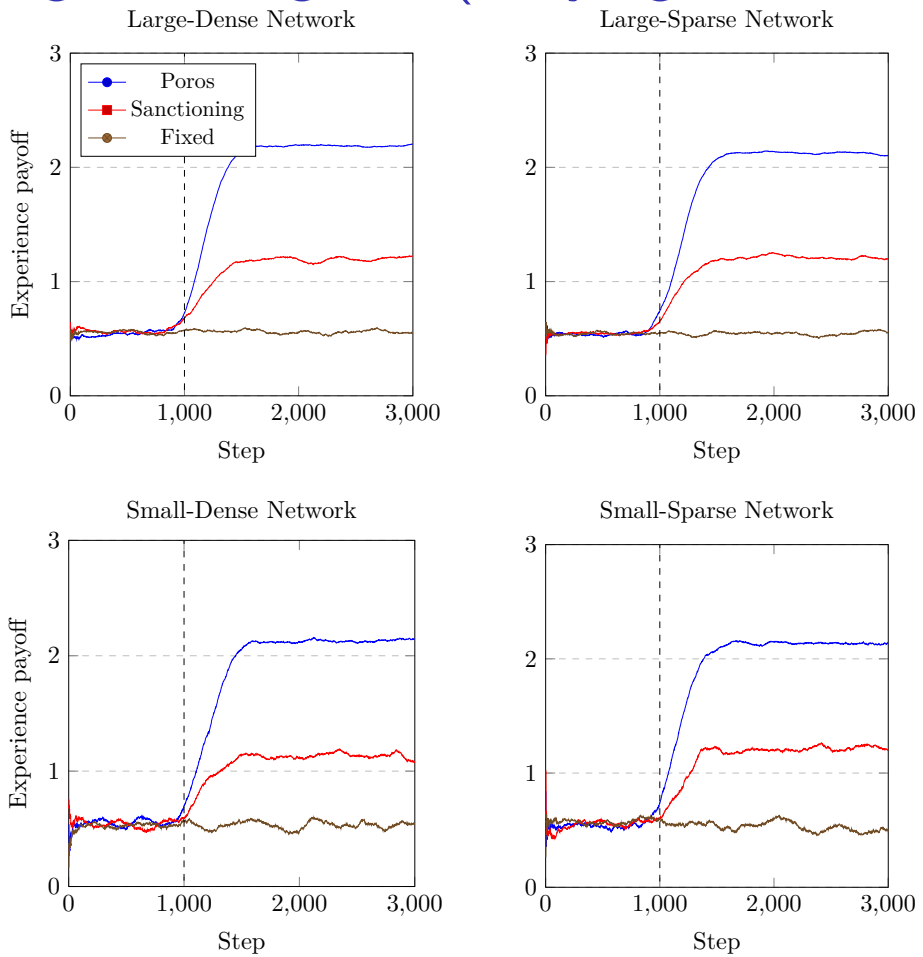
Evaluation: Social Simulations

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

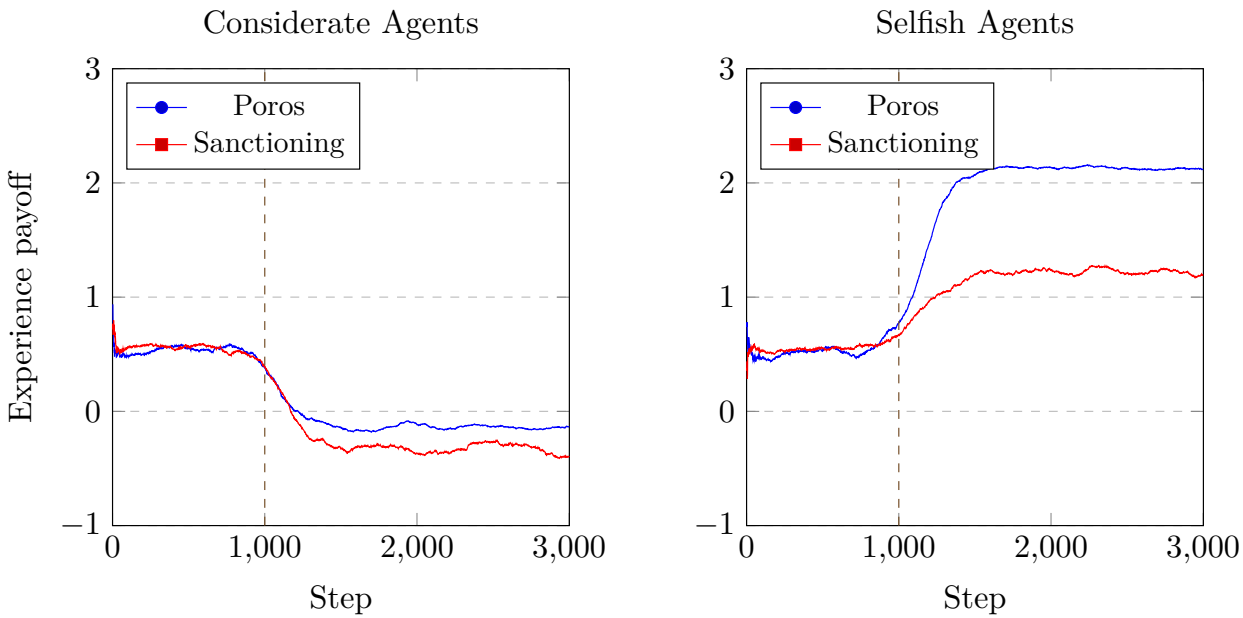
Social experience measures the goal satisfaction delivered by an agent (computed by aggregating payoffs for all stakeholders)

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

Interaction and Decision Making in Elessar

RQ_{Ethics}: How can we enable SIPAs to select actions which are *just*, with respect to the applicable social norms and value preferences of their stakeholders?

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			

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

Simulated Places in the Simulation with Attributes Safe and Sensitive

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

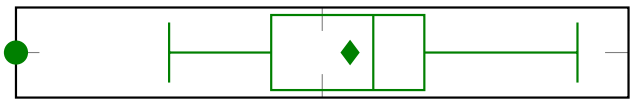
Example Numeric Utility Matrix for a Stakeholder

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

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 [Schnorff et al., 2014]



Casual Conscientious Cautious

Privacy Attitude

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)

Metrics in Society with Mixed Privacy Attitudes

Strategy	Social	Best	Worst	Fairness	p
S_{Elessar}	**1.31**	3.07	**–0.57**	**0.28**	–
S_{primary}	1.23	3.01	–1.14	0.25	<0.01
$S_{\text{conservative}}$	1.07	3.07	–1.55	0.22	<0.01
S_{majority}	1.28	**3.08**	–1.15	0.24	<0.01

Metrics in Society with Majority Privacy Attitudes

Strategy \ Attitude	Cautious				Conscientious				Casual			
	Social	Best	Worst	Fairness	Social	Best	Worst	Fairness	Social	Best	Worst	Fairness
$S_{Elessar}$	**1.25**	2.90	**–0.70**	**0.28**	**1.30**	2.93	**–0.46**	**0.30**	1.38	3.12	**–0.67**	**0.27**
$S_{primary}$	1.15	2.86	–1.07	0.26	1.22	2.91	–1.21	0.25	1.33	3.13	–1.03	0.24
$S_{conservative}$	0.93	2.89	–1.79	0.22	1.09	2.93	–1.42	0.23	1.23	3.13	–1.38	0.23
$S_{majority}$	1.20	**2.92**	–1.27	0.24	1.28	**2.94**	–0.86	0.27	**1.39**	3.13	–0.92	0.25

Location Sharing Survey: Policy Selection

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