

# ELESSAR: Ethics in Norm-Aware Agents

Nirav Ajmeri<sup>1</sup>

Hui Guo<sup>1</sup>

Pradeep K. Murukannaiah<sup>2</sup>

Munindar P. Singh<sup>1</sup>

<sup>1</sup>North Carolina State University

<sup>2</sup>Delft University of Technology

May 2020

# Examples of Ethical Concerns

Audio leaking: Intrusion of solitude and disclosure of music taste



Source: <https://twitter.com/akokitamura/status/728521725172846592>



Source: <https://twitter.com/TheSimpsons/status/441000198995582976>

## Tradeoffs: Values of Power, Pleasure, and Benevolence

# Socially Intelligent Personal Agent (SIPA)

A SIPA adapts to social context and supports meeting social norms

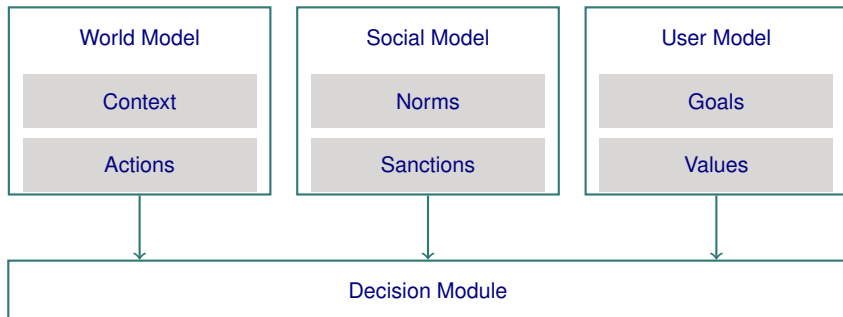
- Ethical: Seeks to balance needs of
  - Primary user, who directly interacts with the agent
  - Other stakeholders, who are affected by the agent's actions

## ELESSAR

Framework that enables ethical-decision making in light of users having distinct value preferences

- adapts a multi-criteria decision-making approach (VIKOR) [Opricovic, 2004] to identify a consensus action
- addresses decision making by an individual agent but in a social context

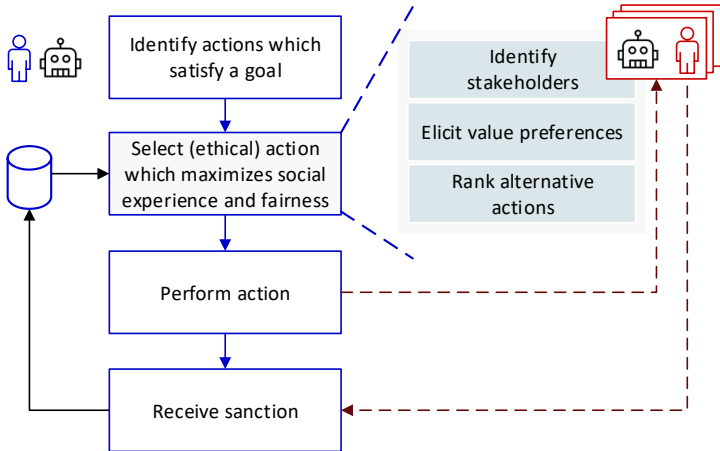
# A SIPA: Schematically



# Interaction in ELESSAR

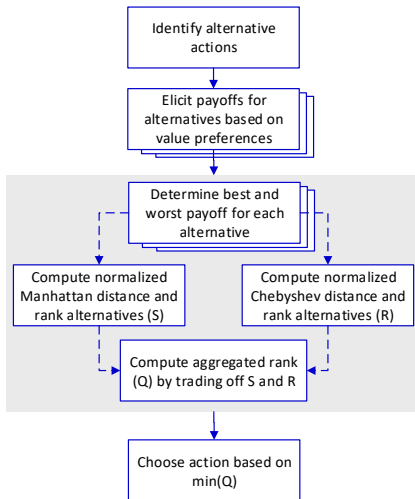
How can SIPAs aggregate value preferences of their stakeholders?

A SIPA's secondary stakeholders can change with the context



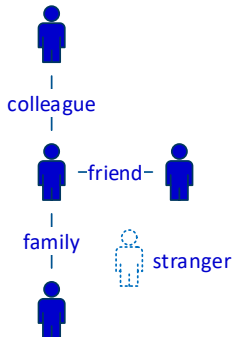
# Choosing an Ethical Action

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



# Setting: Context Sharing

Places, companion, sharing policy



Safe	¬Sensitive	Safe	¬Sensitive
Attending graduation ceremony		Presenting a conference paper	
Safe	¬Sensitive	Safe	¬Sensitive
Studying in a library		Visiting an airport	
¬Safe	¬Sensitive	¬Safe	¬Sensitive
Hiking at night		Being stuck in a hurricane	
Safe?	Sensitive	Safe?	Sensitive
Visiting a bar with fake ID		Visiting a drug rehab center	



- Share with all
- Share with common friends
- Share with companions
- Share with no one



# Evaluation: Crowdsourcing Study

Schnorff et al.'s privacy attitude survey: Level of comfort in sharing personal information

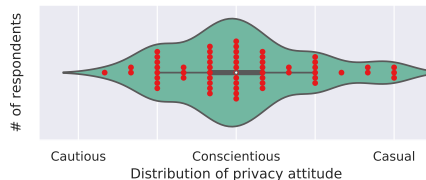
**Level of comfort** in setting context sharing policy

- Context includes place, activity, and social relationship with companions
- Places provided by us but not their safety and sensitivity ratings

**Priming** Based only on context to prime the users

**Survey** Based on context and value preferences (pleasure, privacy, recognition, safety)

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

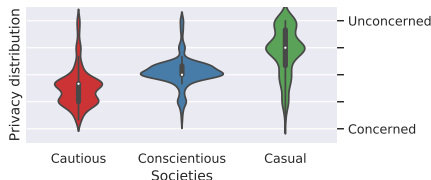


# Evaluation: Simulation

## Simulated societies:

- Mixed
- Cautious
- Conscientious
- Casual

## Privacy attitude:



## Decision-making strategies:

$S_{\text{ELESSAR}}$ : Policy based on VIKOR

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

$S_{\text{conservative}}$ : Least privacy-violating

$S_{\text{majority}}$ : Most common

# Measures of Ethicality

For each interaction, ...

**Social experience** is the utility obtained by a society as a whole divided by the number of stakeholders

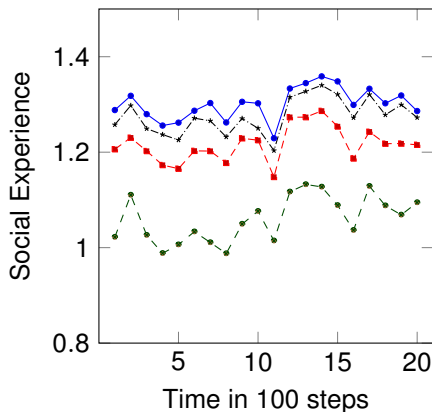
**Best individual experience** is the maximum utility obtained across the SIPA's stakeholders during a single interaction

**Worst individual experience** is the minimum utility obtained across the SIPA's stakeholders during a single interaction

**Fairness** reciprocal of the disparity between the best and worst accumulative individual experiences obtained by users

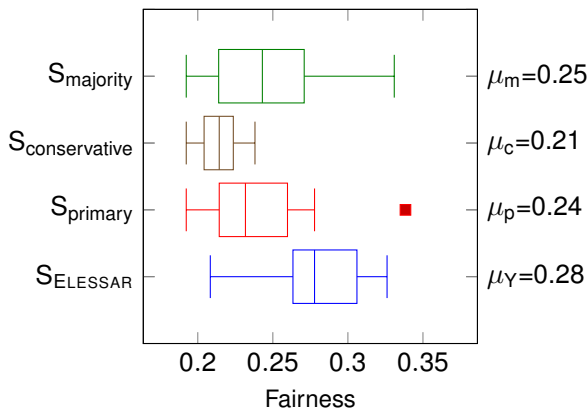
# Experience: Experiment with Mixed Privacy Attitudes

Result: ELESSAR yields higher social experience ( $p < 0.01$ ; Glass'  $\Delta > 0.8$  indicating large effect size) than the baselines.



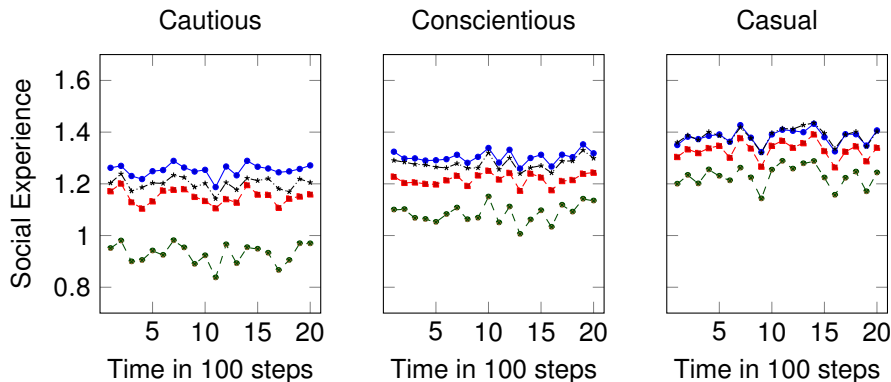
# Fairness: Experiment with Mixed Privacy Attitudes

Results: Fairness in a mixed society. ELESSAR gives significantly better ( $p < 0.01$ ) fairness with large effect size (Glass'  $\Delta > 0.8$ ) than the baseline methods.



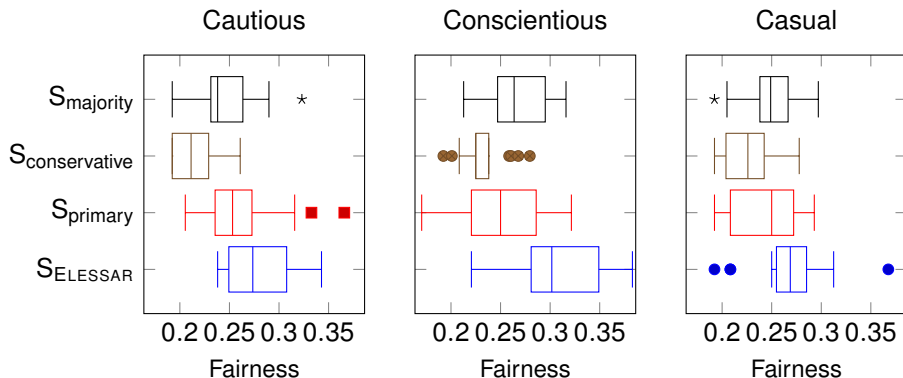
# Experience: Experiments with Majority Privacy Attitudes

Result: ELESSAR yields higher social experience than baselines ( $p < 0.01$ ; Glass'  $\Delta > 0.8$  indicating large effect size) than baselines.



# Fairness: Experiments with Majority Privacy Attitudes

Results: ELESSAR gives significantly better ( $p < 0.01$ ) fairness with large effect size (Glass'  $\Delta > 0.8$ ) than baselines.



# Conclusions

Ethics inherently involves looking beyond one's self interests

## ELESSAR

- advances science of security and privacy by tackling a nuanced notion of privacy—understood as an ethical human value
- considers users other than primary users
  - accommodates stakeholders' value preferences in its decision making
  - demonstrates the gains in fairness accruing



# Thank You

Multiagent Systems and Service-Oriented Computing Laboratory, NCSU  
<http://research.csc.ncsu.edu/mas/>

Acknowledgement:  
Science of Security Lablet at NCSU