Ethics, Values, and Personal Agents

NC STATE UNIVERSITY

Nirav Ajmeri[†], Hui Guo[†], Pradeep K. Murukannaiah[‡], and Munindar P. Singh[†]

HOTSOS 2018

Motivation

Socially intelligent personal agent (SIPA)

- understands the social norms governing its user's interaction in a society,
- adapts according to the social context,
- and acts on behalf of or assists a user to bring about his or her goals [1].

SIPAs ought respect their user's values and act ethically.

Norms, Values, & Ethics

Values: what is right or good according to an individual

- terminal: security, privacy, happiness, recognition
- non-terminal: honesty, helpfulness, forgiveness

Ethics: system of values

Social norm: characterizes interactions between autonomous parties

Social norms indicate whether a SIPA should perform or not perform an actions. *Values* provide a SIPA a reason to pursue or not to pursue those actions [2].

Research Question

RQ. How can we engineer an ethical SIPA such that it

1. understands its user's preferences among values and
2. reasons about values to make ethical policy decisions?

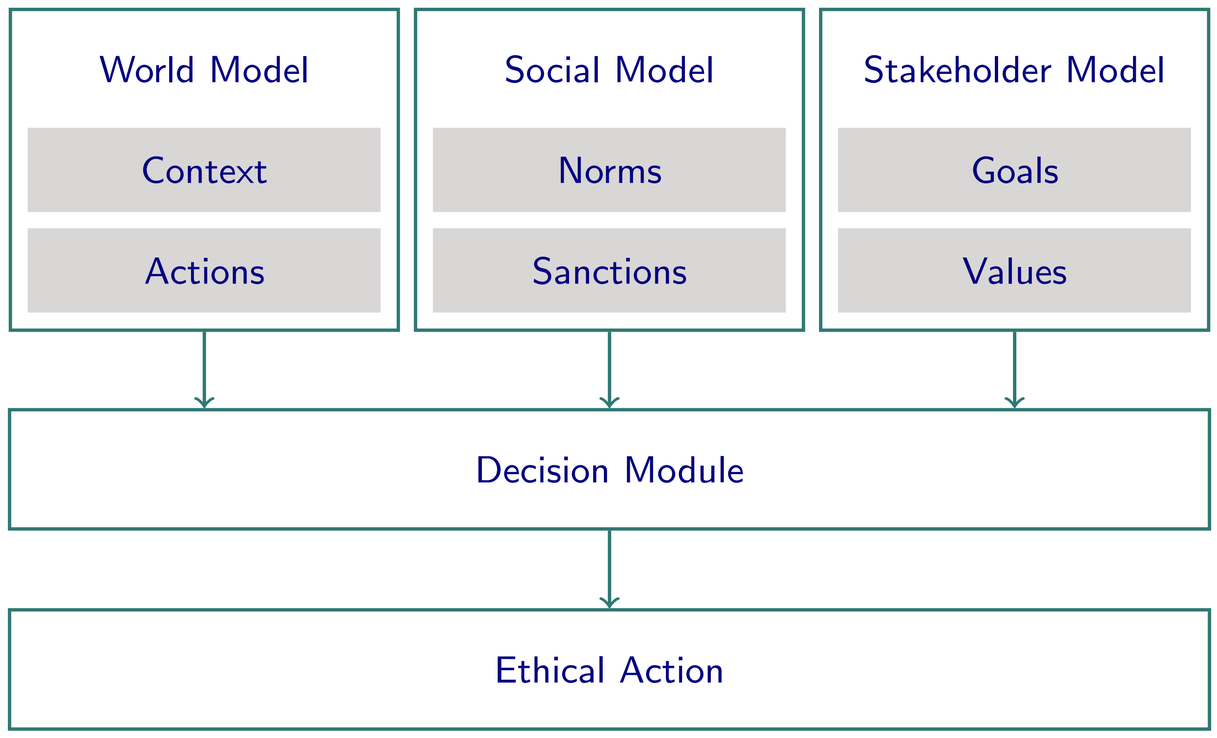
Challenges

- A SIPA's action may simultaneously promote and demote several values. What values are relevant for choosing an ethical action?
- A SIPA's decision affects other users as users may have conflicting value preferences. How can a SIPA understand and reason about value preference of others?

Contributions

- Ainur, a framework for engineering value-driven, ethical SIPAs
- Such a SIPA can make value-promoting ethical decisions, especially, in scenarios where the applicable norms conflict

Ainur SIPA Schematically



Empirical Study

- 33 computer science graduate and undergraduate students
- one factor design with two alternatives:
  - *control*: not aware of values;
  - *treatment*: aware of values promoted and demoted in each context

Survey

- imagine you are in a given context
  - a combination of a place, time of the day of visit, and companions (alone, a colleague, crowd, a family member, or a friend)
  - tagged safe, unsafe, sensitive, or sensitive
- select check-in policy for the context

Metrics

- *Check-in* policy on privacy preservation scale: share with *none*, *companions*, common friends, and all
- *Confidence* in policy on a Likert scale of 1 (very low) to 5 (very high)

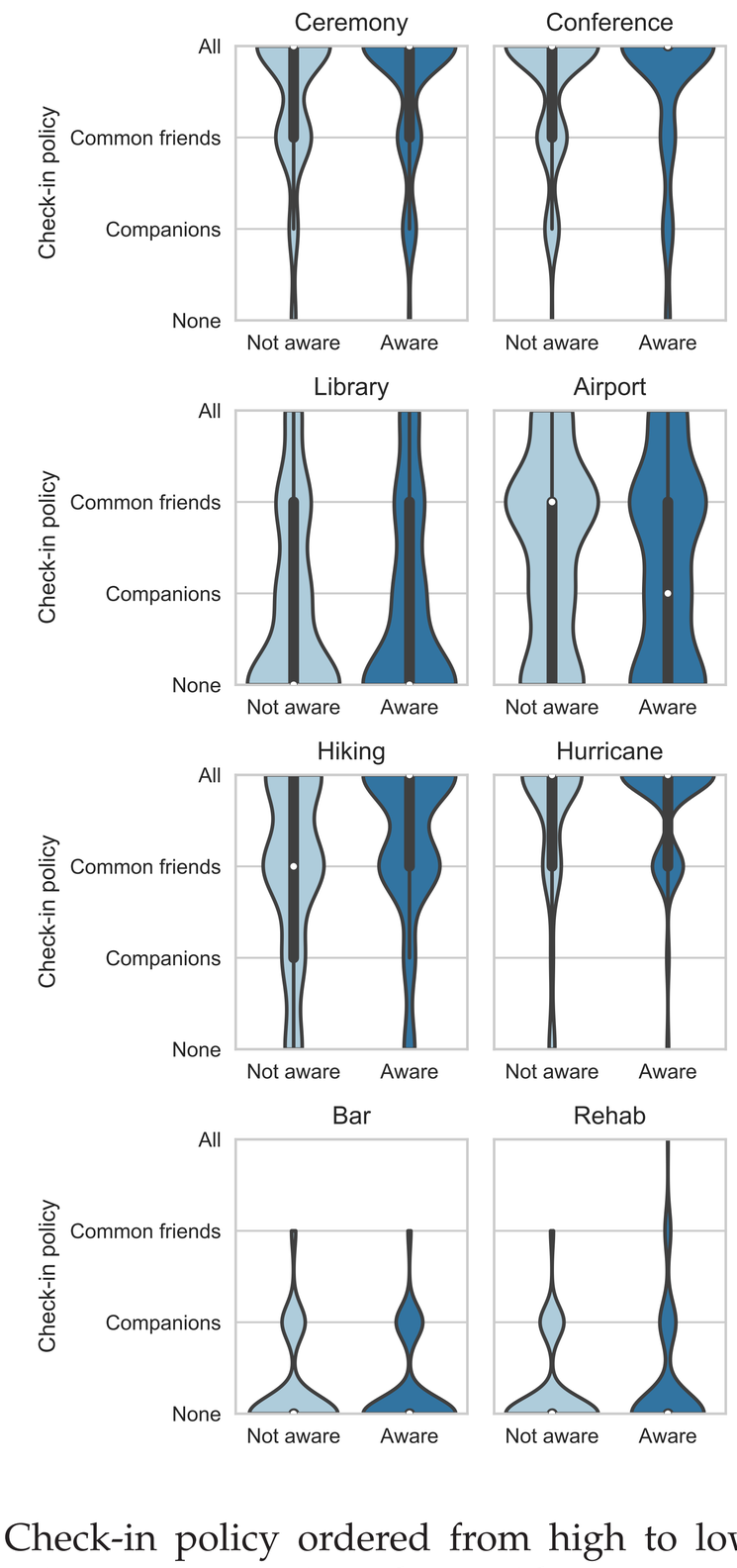
Hypotheses

- Making an informed decision
- Making a confident decision

Future Works

- Evaluate the effectiveness of Ainur via a developer study
- Crowdsource data about values and decision making about sharing policies on a larger scale
- Employ machine learning to assist SIPAs learn their user's value preferences

Survey Results



Check-in policy ordered from high to low privacy preservation: share with *none*, *companions*, *common friends* (of companions), and *all*

Acknowledgements

We thank the US Department of Defense for support through the Science of Security Lablet grant to NC State University.

References

[1] N. Ajmeri, P. K. Murukannaiah, H. Guo, and M. P. Singh. Arnor: Engineering personal agents to deliver a social experience. In *Proc. 16th International Conf. Autonomous Agents and Multiagent Systems*, pages 230–238, 2017.

[2] F. Dechesne, G. D. Tosto, V. Dignum, and F. Dignum. No smoking here: values, norms and culture in multi-agent systems. *Artificial Intelligence and Law*, 21(1):79–107, Mar. 2013.

Hot Topics in the Science of Security
April 10–11, 2018 | Raleigh, North Carolina

http://research.csc.ncsu.edu/mas/
[†]{najmeri,hguo5,mpsingh}@ncsu.edu, [‡]pkmvse@rit.edu