Indecision Modeling

Duncan C McElfresh[1], Lok Chan[2], Kenzie Doyle[3], Walter Sinnott-Armstrong[2],
Vincent Conitzer[2], Jana Schaich Borg[2], John P Dickerson[1]

1: University of Maryland, College Park
2: Duke University
3: University of Oregon

AAAI 2021
Room BC-D3-R1
Position E5

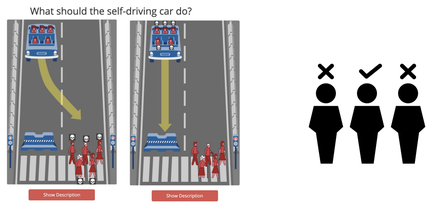
Preprint: https://arxiv.org/abs/2012.08485

Motivation: Preference Modeling & AI

Many AI systems used to **make** or **contribute to important decisions** are guided by **mathematical models of stakeholder preferences**.

Our main motivation: Preference aggregation for policy design
Stakeholder preferences are *learned* and *aggregated* to make important decisions. The literature has considered applications of...

- Self-driving cars [Noothigattu et al., 2018]
- Kidney exchange [Freedman et al., 2020]



https://www.moralmachine.net/

Background: Preference Modeling & Indecision

Decision-maker ("Agent") **Alternatives**

Assumption: Agent decisions are consistent with a preference relation:
for each pair of alternatives (i, j) either…

1. $i > j$ " i is more-preferred than j ", or
2. $i < j$ " i is less-preferred than j ", or
3. $i \approx j$ " neither $i > j$ nor $i < j$ "

Assumption: Preferences are…

- Complete (defined for all (i, j) pairs)
- Transitive:
  - $i > j$ and $j > k$ imply $i > k$
  - $i \approx j$ and $j \approx k$ imply $i \approx k$

There exists a *utility function* representation for preferences, s.t.

- $u(i) > u(j) \Leftrightarrow i > j$
- $u(i) = u(j) \Leftrightarrow i \approx j$

However, agents may sometimes be *indecisive...*

What if i and j are **not comparable?** [Pini'11]

What if the agent doesn't want to **"play God"** [Gangemi'13]

Our Contributions

We *test* & *reject* the assumption:

"indecision" ⇔ "alternatives are equally preferred"

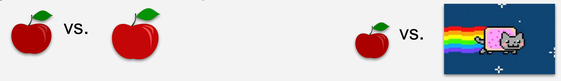
(1) We explore possible causes of indecision, and develop a modeling framework for hypothesis testing.

(2)

Causes for Indecision

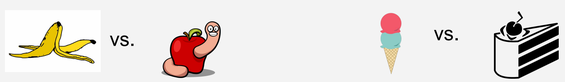
Difference-Based Indecision

...when the *perceived difference* between X and Y is too small (or too large) to arrive at a strict preference.



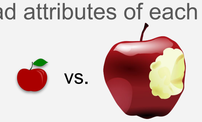
Desirability-Based Indecision

...when both X and Y are "too bad" (or "too good"). [Zakay'84] [Luce'98]



Conflict-based Indecision

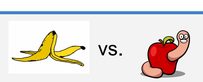
...when there are both good and bad attributes of each alternative. [Tversky'92]



Indecision Modeling Framework

Example

Desirability-Based Indecision



… indecisive when both items are *below* utility threshold λ

(A) Response function

$$R(i,j) \equiv \begin{cases} > & \text{if } u(i) \geq \max\{u(j), \lambda\} \\ < & \text{if } u(j) \geq \max\{u(i), \lambda\} \\ \approx & \text{if } \lambda \geq \max\{u(i), u(j)\} \end{cases}$$

(B) Equivalent "score-based" response function

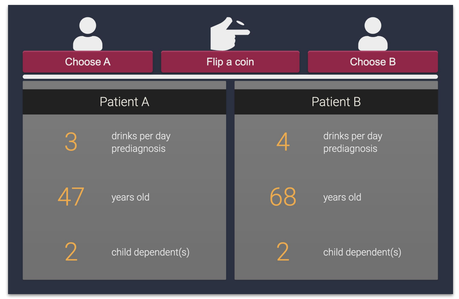
$$S_{\succ}(i,j) \equiv u(i)$$
$$S_{\prec}(i,j) \equiv u(j)$$
$$S_{\approx}(i,j) \equiv \lambda$$
$$R(i,j) = \underset{r \in \{\prec, \succ, \approx\}}{\arg\max} S_r(i,j)$$

Study 1: indecision ≠ indifference

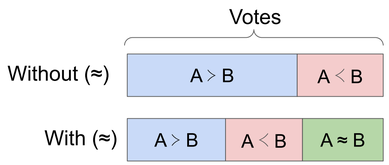
Hypothetical Decision Scenario:
Participants choose to give a donor kidney to one of two patients.

Participants **"vote"** for a patient….

- some are *forced to choose*
- some can be indecisive (≈)



Example question used in our surveys.



We **reject** ($p<0.01$) the hypothesis:

Indecisive voters prefer both alternatives equally.

For our participants, **"indecision"** does not mean
"both alternatives are equally preferred"

Questions for Future Work

Q: How do we *aggregate* indecisive stakeholders' preferences, when indecision has several meanings?
Are traditional social choice methods appropriate?

Q: What if we *ask* people why they are indecisive?
Will their responses match any of our models?

Q: How do we elicit different causes of indecision without a communication bottleneck?

References

- Freedman, Rachel, et al. "Adapting a kidney exchange algorithm to align with human values." (AIJ, 2020)
- Noothigattu, Ritesh, et al. "A voting-based system for ethical decision making." (AAAI'18)
- Pini et al' "Incompleteness and incomparability in preference aggregation: Complexity results" (AIJ, 2011)
- A Gangemi, F Mancini "Moral choices: the influence of the do not play god principle" (CogSci'13)

Acknowledgments

Dickerson and McElfresh were supported in part by NSF CAREER IIS-1846237, NSF CCF-1852352, NSF D-ISN #2039862, NIST MSE #20126334, NIH R01 NLM-013039-01, DARPA GARD #HR00112020007, DoD WHS #HQ003420F0035, DARPA Disruptioneering (SI3-CMD) #S4761, and a Google Faculty Research Award. Conitzer was supported in part by NSF IIS-1814056. This publication was made possible through the support of a grant (TWCF0321) from Templeton World Charity Foundation, Inc. to Conitzer, Schaich Borg, and Sinnott-Armstrong. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of Templeton World Charity Foundation, Inc.