Beyond Algorithm Aversion
The Impact of Conventionality on Error Evaluations

Hamza Tariq, Jonathan A. Fugelsang, Derek J. Koehler
Corresponding Author: h33tariq@uwaterloo.ca

University of Waterloo
Department of Psychology

NSERC CRSNG

Introduction

- People tend to judge and penalize algorithmic errors more harshly than identical human mistakes—a bias known as **algorithm aversion**.
- This aversion can be **irrational**, as it often leads to a preference for inferior human forecasters even when more accurate and reliable algorithmic decision aids are available.
- It is speculated that **expectations for algorithms** to be flawless, along with concerns about their black box nature, potential for systematic errors, lack of qualitative judgment, and ethical implications, may contribute to this aversion.
- Our goal has been to move beyond the properties of the algorithm itself and explore **the role of context** —such as the status quo or conventionality—an aspect we believe existing research has not sufficiently addressed.

Current Research

In judgments comparing human and algorithm mistakes, do people always exhibit a higher bias against the algorithm, or is the bias affected by which option—human or algorithm—is the convention and which is the alternate?

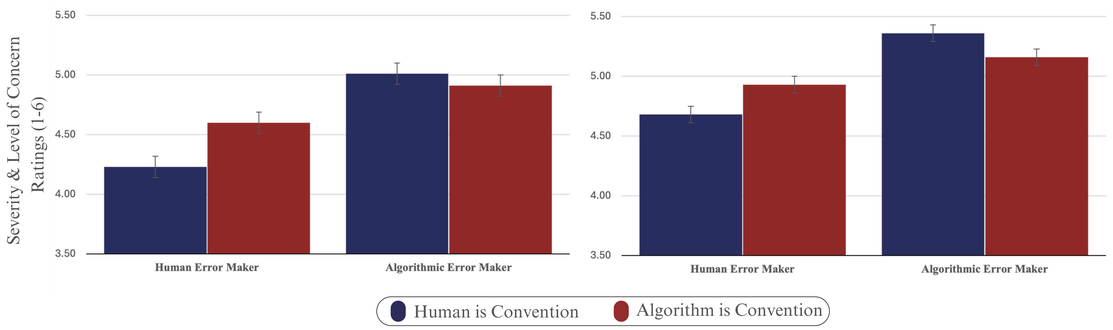
MAIN HYPOTHESIS: We propose that **alternate aversion** occurs when the presence of a conventional option leads to stronger aversion against the non-conventional option. This results in harsher judgment and penalties for identical errors by the non-conventional option. In algorithm aversion, it may be the algorithm's typical status as the non-conventional or alternative option in the human-algorithm comparison that drives this bias, and not just the algorithm itself.

Results

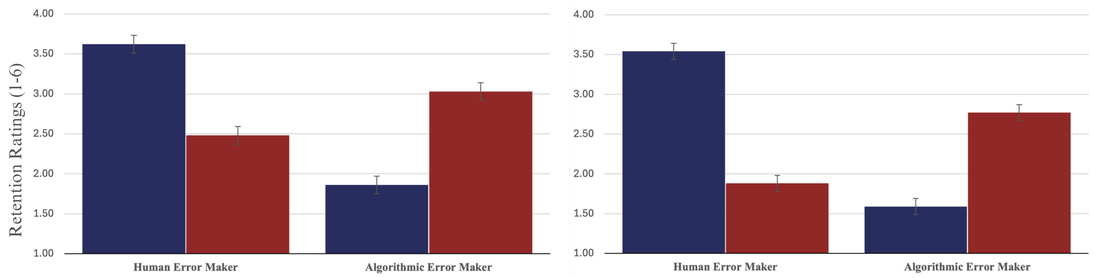
STUDY 1
Admissions Scenario
(N = 594)

STUDY 2
Speakers Scenario
(N = 605)

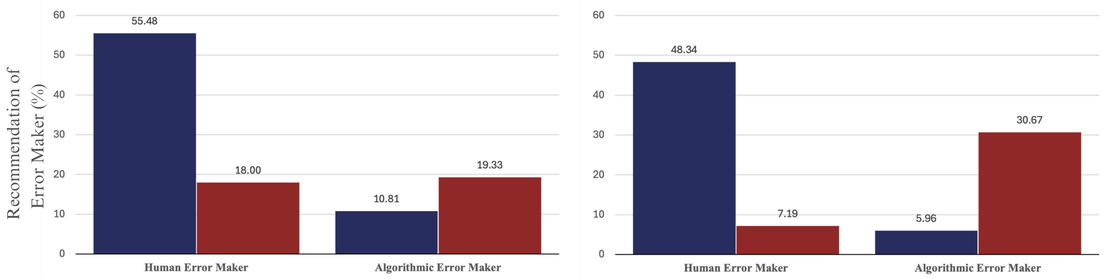
Judgments of Mistake Severity and Level of Concern After Error



Preference to Retain Error Maker After Error



Recommendations for Error Maker's Future Use



Methods

Two between-subject studies (N = 1,199) where participants were presented with a hypothetical scenario where either a human or an algorithmic agent is the conventional decision maker in that domain.

STUDY 1
Conventional or alternative agent makes an error while screening college applicants.

STUDY 2
Conventional or alternative agent makes an error while quality testing faulty sound speakers.

Participants asked to judge the severity of the mistake, their concern for the future use of the error maker, their preference to retain the error maker, and to recommend its future use in that domain.

Summary

- When told decisions were conventionally made by humans, participants judged algorithmic mistakes more harshly, confirming algorithm aversion.
- Framing the algorithm as the conventional option **reduced, eliminated or even reversed algorithm aversion**, showing that conventionality impacts error judgments.
- There is **evidence for alternate aversion** — suggesting that people are averse to non-conventional decision-makers, whether human or algorithmic.
- As our relationship with technology continues to evolve, a human **preference for the status quo** could be key to understanding human interactions with modern algorithmic tools like AI.

References

- Bonezzi, A., Ostinelli, M., & Melzner, J. (2022). The human black-box: The illusion of understanding human better than algorithmic decision-making. *Journal of Experimental Psychology: General*, 151(9), 2250–2258.
- Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. Cognition, 181, 21–34.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668–1674.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126.
- Logg, J. M. (2022). The psychology of Big Data: Developing a "theory of machine" to examine perceptions of algorithms. In The Psychology of Technology (pp. 349–378). American Psychological Association
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7–59.