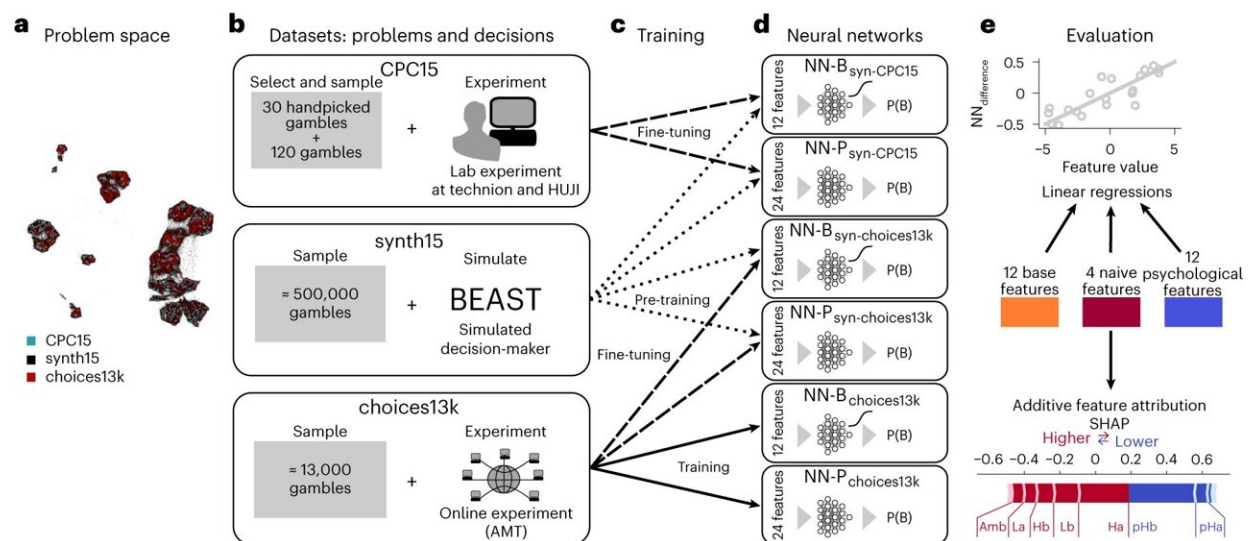


Investigating dataset bias in machine-learned theories of economic decisions

January 12 2024, by Bettina Bastian



Schematic of analyses of the relationship between datasets and models. Credit: *Nature Human Behaviour* (2024). DOI:10.1038/s41562-023-01784-6

Researchers at the Center for Cognitive Science at TU Darmstadt and hessian.AI have investigated the properties of behavioral economic theories automatically learned by AI.

In our daily lives, we are constantly confronted with risky choices. Economists, psychologists and [cognitive scientists](#) have long studied people's risky choices in the laboratory with equivalent gambles because gambles have the three components that define risky choices: there is a

choice between alternative actions, the outcomes of choices have a certain probability, and the outcomes have payoffs.

Would you rather choose \$100 with certainty or get a [lottery ticket](#), where the chances of winning \$150 are 75%, but in the remaining 25% you get nothing?

The results of such experiments have shown consistently that people systematically deviate from the mathematical optimum choice, i.e., they lose money. In the above-mentioned game of chance, for example, the second option has a higher expected value, but when asked, many people prefer the first option.

Due to the serious impact of these decision errors on the lives of individuals and the economy as a whole, predicting how humans make suboptimal decisions is of ongoing scientific interest. The work of Daniel Kahneman and Amos Tversky, who stated they were studying "natural stupidity instead of [artificial intelligence](#)," led to a better description of human decision-making through Cumulative Prospect Theory and the awarding of the Nobel Prize in Economic Sciences in 2002.

However, there are still many anomalies and contexts in which decisions cannot be predicted well. Beyond predicting people's decisions, it is even more difficult to find explanations for them. This includes prominent ideas such as Gerd Gigerenzer's that human decisions are based on cognitive shortcuts, so-called heuristics.

In a [recent study](#) published in *Science*, a research team from Princeton University, U.S., used artificial intelligence to better understand human decisions in risky gambles.

The idea is if [deep neural networks](#) become increasingly successful at

predicting data, they could perhaps also predict human decisions better than behavioral economic theories. However, such [neural networks](#) require huge amounts of data in order to be trained. For this reason, a data set containing human decisions on more than 13,000 bets was collected as part of the study.

When predicting people's gambling decisions based on this data set, the neural networks that were least constrained by theoretical assumptions performed best. Peterson et al. then derived a "machine-learned theory of economic decision-making"—an interpretable summary of the neural networks' behavior.

In a [new publication](#) in the journal *Nature Human Behavior*, researchers from the LOEWE project "WhiteBox" at TU Darmstadt have systematically investigated the predictions that result from combining different machine learning models with different decision data sets.

The publication, which originated in a course project in the Master's program in cognitive science, found striking differences in the prediction of human decisions: While some neural networks were most accurate at predicting decisions from the 2021 study dataset, they failed to predict human behavior from smaller psychological experiments. This is a typical example of how biases in data sets can lead to interaction effects between models and data sets, so that findings from one data set do not transfer to another.

From these observations, the researchers were able to develop a cognitive generative model that quantitatively explains the differences between the actual decisions from the data sets and the predictions of the AI models with classical behavioral economic results of human decision uncertainty.

"Neural networks may outperform all proposals by human theorists in

terms of prediction errors on a dataset, but that does not guarantee that this transfers to other datasets of human gambles, or even more naturalistic, everyday decisions," explains Professor Constantin Rothkopf from the Center for Cognitive Science at TU.

The study underlines that [cognitive science](#) still cannot easily be automated by artificial intelligence and that a careful combination of theoretical reasoning, machine learning and data analysis are required to understand and explain why [human](#) decisions deviate from the mathematical optimum.

More information: Thomas, T et al, Modelling dataset bias in machine-learned theories of economic decision-making. *Nature Human Behaviour* (2024). [DOI: 10.1038/s41562-023-01784-6](https://doi.org/10.1038/s41562-023-01784-6)

Provided by Technische Universität Darmstadt

Citation: Investigating dataset bias in machine-learned theories of economic decisions (2024, January 12) retrieved 8 May 2024 from <https://techxplore.com/news/2024-01-dataset-bias-machine-theories-economic.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
