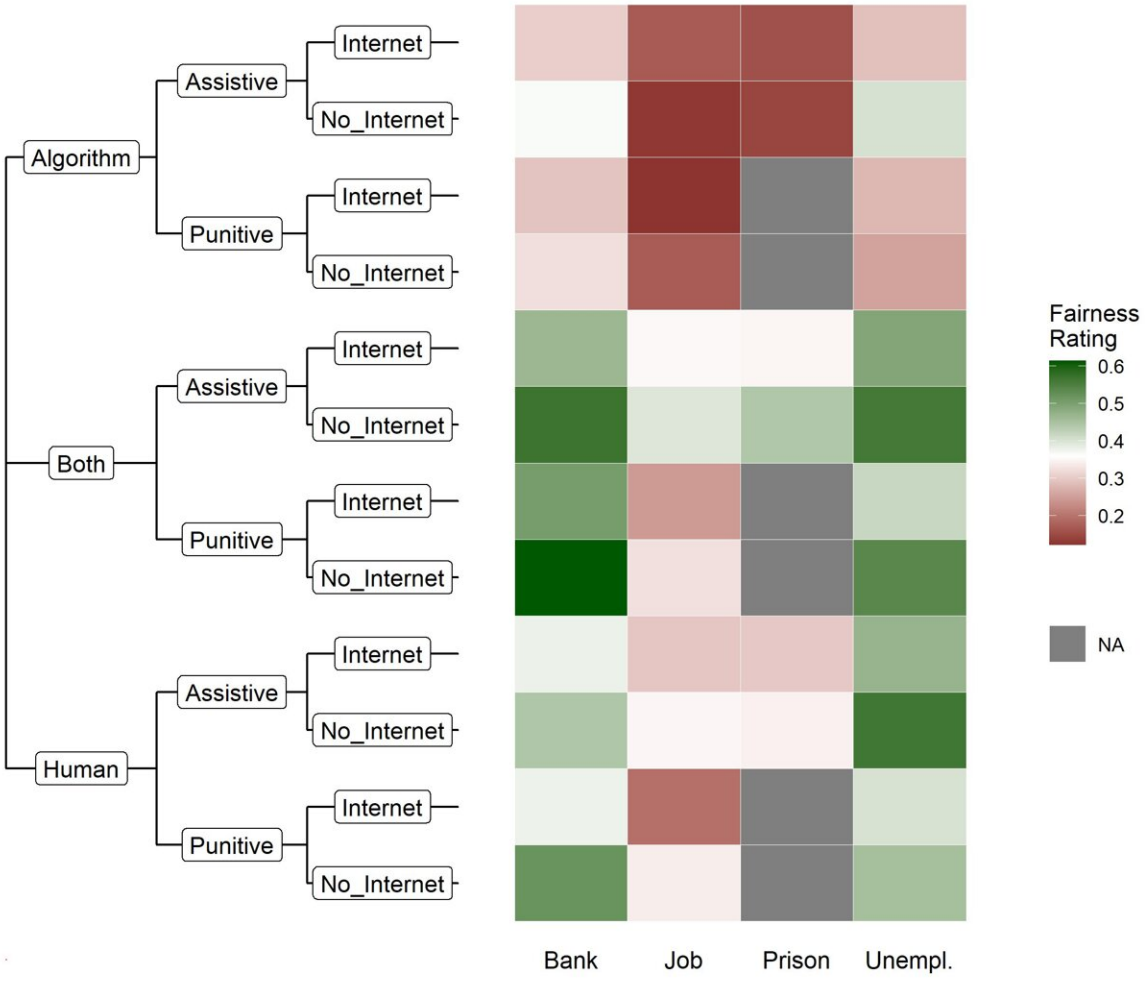
Do humans think computers make fair decisions?

September 29 2022



The heatmap shows relative frequencies of respondents that rated a scenario as "Fair" (i.e., either "Somewhat fair" or "Very fair"). The color scale is centered at the average fairness rating over all experiments. Credit: Patterns/Gordon and

Today, machine learning helps determine the loan we qualify for, the job we get, and even who goes to jail. But when it comes to these potentially life-altering decisions, can computers make a fair call? In a study published September 29 in the journal *Patterns*, researchers from Germany showed that with human supervision, people think a computer's decision can be as fair as a decision primarily made by humans.

"A lot of the discussion on fairness in machine learning has focused on technical solutions, like how to fix unfair algorithms and how to make the systems fair," says computational social scientist and co-author Ruben Bach of the University of Mannheim, Germany. "But our question is, what do people think is fair? It's not just about developing algorithms. They need to be accepted by society and meet normative beliefs in the real world."

Automated decision-making, where a conclusion is made solely by a computer, excels at analyzing large datasets to detect patterns. Computers are often considered objective and neutral compared with humans, whose biases can cloud judgments. Yet, bias can creep into computer systems as they learn from data that reflects discriminatory patterns in our world. Understanding fairness in computer and human decisions is crucial to building a more equitable society.

To understand what people consider fair on automated decision-making, the researchers surveyed 3,930 individuals in Germany. The researchers gave them hypothetical scenarios related to the bank, job, prison, and unemployment systems. Within the scenarios, they further compared different situations, including whether the decision leads to a positive or

negative outcome, where the data for evaluation comes from, and who makes the final decision—human, computer, or both.

"As expected, we saw that completely automated decision-making was not favored," says computational social scientist and co-first author Christoph Kern of the University of Mannheim. "But what was interesting is that when you have human supervision over the automated decision-making, the level of perceived fairness becomes similar to human-centered decision-making." The results showed that people perceive a decision as fairer when humans are involved.

People also had more concerns over fairness when decisions related to the criminal justice system or job prospects, where the stakes are higher. Possibly viewing the weight of losses greater than the weight of gains, the participants deemed decisions that can lead to positive outcomes fairer than negative ones. Compared with systems that only rely on scenario-related data, those that draw on additional unrelated data from the internet were considered less fair, confirming the importance of data transparency and privacy. Together, the results showed that context matters. Automated decision-making systems need to be carefully designed when concerns for fairness arise.

While hypothetical situations in the survey may not fully translate to the real world, the team is already brainstorming next steps to better understand fairness. They plan on taking the study further to understand how different people define fairness. They also want to use similar surveys to ask more questions about ideas such as distributive justice, the fairness of resource allocation among the community.

"In a way, we hope that people in the industry can take these results as food for thought and as things they should check before developing and deploying an automated decision-making system," says Bach. "We also need to ensure that people understand how the data is processed and how

decisions are made based on it."

More information: Christoph Kern and colleagues, Humans vs. Machines: Who is Perceived to Decide Fairer? Experimental Evidence on Attitudes Towards Automated Decision-Making, *Patterns* (2022). DOI: 10.1016/j.patter.2022.100591. www.cell.com/patterns/fulltext … 2666-3899(22)00209-4

Provided by Cell Press