Taking Algorithmic (Vs. Human) Advice Reveals Different Goals to Others

Emir Efendić, Philippe P. F. M. Van de Calseyde, Štěpán Bahník & Marek A. Vranka

View supplementary material ⬀

Published online: 18 May 2023.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS   Check for updates

Taking Algorithmic (Vs. Human) Advice Reveals Different Goals to Others

Emir Efendić[a] , Philippe P. F. M. Van de Calseyde[b], Štěpán Bahník[c], and Marek A. Vranka[c]

[a]Department of Marketing and Supply-Chain Management, School of Business and Economics, Maastricht University, Maastricht, The Netherlands; [b]Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands; [c]Department of Management, Faculty of Business Administration, Prague University of Economics and Business, Prague, Czech Republic

ABSTRACT

People are increasingly likely to obtain advice from algorithms. But what does taking advice from an algorithm (as opposed to a human) reveal to others about the advice seekers' goals? In five studies (total $N = 1927$), we find that observers attribute the primary goal that an algorithm is designed to pursue in a situation to advice seekers. As a result, when explaining advice seekers' subsequent behaviors and decisions, primarily this goal is taken into account, leaving less room for other possible motives that could account for people's actions. Such secondary goals are, however, more readily taken into account when (the same) advice comes from human advisors, leading to different judgments about advice seekers' motives. Specifically, advice seekers' goals were perceived differently in terms of fairness, profit-seeking, and prosociality depending on whether the obtained advice came from an algorithm or another human. We find that these differences are in part guided by the different expectations people have of the type of information that algorithmic- vs. human advisors take into account when making their recommendations. The presented work has implications for (algorithmic) fairness perceptions and human-computer interaction.

Making accurate decisions typically hinges on having a profound understanding of all relevant information and decision-makers oftentimes seek out the advice of (presumably) knowledgeable parties. For instance, people are likely to solicit advice from experts in situations of high uncertainty (Hadar & Fischer, 2008), when the costs of advice solicitation are relatively low (Gino, 2008; Schrah et al., 2006), and when the underlying problem is complex (Schrah et al., 2006, Sniezek & Buckley, 1995). However, while in the past people primarily consulted with other humans, recent advances have paved the way for alternative advisors that people regularly call upon for advice: algorithms. For many people, taking advice from algorithms has become a common occurrence and this topic has generated much interest in recent years in exploring how people respond to algorithmic versus human advice (Alexander et al., 2018; Burton et al., 2020; Efendić et al., 2020; Lee, 2018). However, while much research has focused on understanding how people view algorithmic advice (Dawes, 1971; Dietvorst et al., 2015; Longoni et al., 2019), little is known about how people view others who take advice from algorithms. As such, the purpose of the current research is to provide more insights into how observers perceive and understand those who take

algorithmic advice. That is, we propose that soliciting and following (the same) advice generated by an algorithm, as opposed to a human advisor, leads observers to reach markedly different judgments about the goals that motivate advice seekers' decisions.

These inferred goals may largely differ in their morality and in the degree of their alignment to various moral values, such as fairness. Existing research has already addressed a variety of issues related to the fairness of algorithms and it consistently shows that both users and developers are highly concerned with an algorithm's fairness (Corbett-Davies & Goel, 2018; De Cremer & McGuire, 2022). In the literature, fairness is often discussed as a consequence of algorithm deployment. It has been argued that algorithms can reduce or overcome the role of human biases in advice giving as they have no agency and cannot be swayed by extraneous factors (e.g., emotion or cognitive biases; Kleinberg et al., 2018). However, there are also many examples where the deployment of algorithms has been perceived as unfair. For instance, algorithms have been found to perpetuate racial stereotypes in the area of justice, exclude citizens from food support, and mistakenly reduce disability benefits (Richardson, 2022). Algorithmic decisions were also

perceived to be less fair than human decisions, for tasks that people see as requiring a "human touch" (Lee, 2018) and a recent review on the perceived fairness of algorithmic systems concludes that the concept of algorithmic fairness tends to be highly context dependent. That is, fairness perceptions tend to be determined not only by the technical design of the algorithm, but also the area of application and the task at hand (Starke et al., 2022). Much research has thus focused on how people view the fairness of algorithmic advice. In this research, we move beyond the perception of algorithms and examine whether observing others taking advice from algorithms (vs. humans) is perceived differently in terms of fairness and other motivations.

To illustrate, imagine two individuals (Alice and Helen) investing in a company that, unrelated to its core business, has been praised in recent years for its positive impact on the environment. Alice invests in the company at the advice of an investment algorithm while Helen decides to invest in the company after consulting a close friend. Whose goal is it to maximize their profits with this investment: Alice, or Helen? We argue that, in answering this question, people attribute the goals that advisors are believed to pursue to those who follow their advice. In this instance, the investment algorithm is designed to pursue the primary goal of finding the most profitable investment opportunity, ignoring any information in its recommendation concerning the company's positive impact on the environment. Alice, who accepts and adopts the algorithm's suggestion, is therefore similarly seen as being primarily driven by maximizing profits with her investment. Human advisors, on the other hand, are able to take into account additional, secondary goals and information in their recommendation (e.g., the company's positive impact on the environment). Helen, who follows the friend's suggestion to invest in the pro-environment company, is therefore believed to have taken similar, additional considerations into account when investing in the company. As a result, Alice who invests in the company at the advice of an investment algorithm is expected to be perceived as being less honest and fair in her intentions than Helen, who invested after consulting a close friend. In sum, we propose that the primary goal that algorithmic advisors are designed to pursue is attributed to advice seekers when explaining their decisions, leaving less room for other possible motives that could account for their actions. Such secondary motives are, however, considered to influence people's decisions when the advice comes from human advisors. In what follows, we will first introduce prior work on how people use and view algorithmic advice, followed by a discussion on how people are expected to view *others* taking advice from algorithms (vs. human advisors).

1.1. Algorithmic advice

Algorithms can provide high-quality advice—often better than humans (Beck et al., 2011; Carroll et al., 1982; Meehl, 1954). Taking advice from algorithms is on the rise (O'Neil, 2016) and since advice-seekers aim to reduce the costs of achieving a desired outcome (Gall, 1985), soliciting advice

from algorithms can be an optimal solution to many decision problems. Indeed, people are hiring robo-lawyers to help with legal advice (Markovic, 2019) and robo-investors for advice on financial decisions (D'Acunto et al., 2019) while large organizations, as well as governments, are implementing algorithms on an even much wider scale. That is, algorithms are providing advice on who gets hired or fired, what medication to prescribe, or even who is at risk of criminal re-offence (Arkes et al., 2007; Fildes & Goodwin, 2007; Niszczota & Kaszás, 2020; Porter, 2018; Stacey et al., 2017).

However, people's judgments of algorithm-generated advice are multifaceted. For one, although algorithmic advice is generally of high quality, people tend to be reluctant to follow it. A large body of work shows that people do not seem to trust algorithm-generated advice, especially after seeing it making a mistake. For instance, people tend to prefer human to algorithmic predictions (Diab et al., 2011) and they weigh human input more strongly than algorithmic input (Önkal et al., 2009). This type of behavior has been dubbed *algorithm aversion* (Dietvorst et al., 2015) with people being particularly unwilling to accept algorithmic involvement in moral domains (Bigman & Gray, 2018).

Prior research has shown that people view algorithms as fundamentally different from human advisors (Johnson-Laird, 2013), in particular, because of the different ways in which algorithms approach problems and decisions (Bigman & Gray, 2018; Malle et al., 2016). More specifically, algorithms are believed to be single-minded agents, exclusively focussing on pursuing a situation's primary goal via rule-based instructions when finding solutions. Such an exclusive focus also implies an algorithm's disregard of information deemed irrelevant to the goal it is designed to attain. Indeed, the very definition of an algorithm confirms it is "a set of predefined rules a machine or computer follows to achieve a particular goal" (Merriam-Webster.com, 2022). Human advisors, on the other hand, are able to consider additional, secondary goals (and related information) when examining possible solutions (Gray et al., 2007; Longoni et al., 2019). This supposedly allows them to better adjust their advice to the situation at hand. We argue that people rely on these differences in how algorithmic- versus human advisors approach problems and decisions when assessing the goals that advice seekers strive to accomplish when following their advice.

1.2. Attribution theory, goal inferences and taking advice

Sometimes the goals that others pursue are readily available because they are explicitly communicated (Aarts et al., 2004). Oftentimes, however, goals are not explicitly conveyed so people need to use different strategies to make sense of others' motives. One of the most common approaches to accomplish this task is to infer goals from people's overt behaviors (Baker et al., 2009; Hauser et al., 2014). More specifically, following attribution theory, research consistently indicates that observers readily rely on people's behaviors to make sense of the goals that others

pursue, especially when their behaviors are believed to occur voluntarily (Jones & Davis, 1965). For example, while an act of trust typically signals to others a person's willingness to cooperate (Rousseau et al., 1998), this signal becomes rather vague when the person has no choice but to trust the other party (McCabe et al., 2003). These overall inferences following the actions of others seem to come rather effortlessly, even when observing inanimate actors. For example, in their seminal study, Heider and Simmel (1944) asked participants to watch a movie clip where three geometrical figures were shown moving in various directions, asking them to "write down what happened in the picture" (p. 245). Almost all participants attributed motivational states to the pictured animations when explaining their movements (e.g., the larger triangle is attacking the smaller circle because it's angry). In sum, people often go "beyond the information given" (Bruner, 1957) when making sense of the goals that others pursue and here, we argue that similar processes are at work when observing others taking advice from an algorithm versus another individual.

Advice refers to the process of soliciting and obtaining recommendations from other parties on what to do, say, or think about a problem (Guntzviller et al., 2020) and often occurs in a social context (with advice seekers interacting with advisors and third parties observing these interactions; Bonacio & Dalal, 2006). Yet, although social in nature, the interpersonal dynamics of advice taking have only recently been getting more attention (Ache et al., 2020; Blunden et al., 2019). Of relevance is the notion that observing an advice seeker acting upon an advisor's recommendation signals to others that the person accepts and adopts the reasoning that led the advisor to favor a certain option (Guntzviller et al., 2017, 2020; Liljenquist, 2010). Here we argue that the reasons that led an *algorithmic* advisor to favor a certain option are perceived to be fundamentally different than when the same option is favored by a human advisor. As such, a person who accepts and adopts the suggestion of an algorithm (vs. a human advisor) is therefore expected to be judged differently in terms of their reasoning that led them to favor a certain option.

In what follows, we first examine, in three initial experiments (Studies 1A − 1C), how people view others taking algorithmic vs. human advice in various areas where the issue of algorithmic fairness is critical and relevant (i.e., financial investment, healthcare, and public policy). Subsequently, we tap into the suggested process and demonstrate that making the human advisor more focused on attaining the situation's primary goal (and thus more algorithm-like) leads to similar judgments of the advice seeker when they follow algorithmic advice (Study 2). We find that these differences in judgments are in part guided by the different expectations people have of the type of information that algorithmic advisors take into account when making their recommendations. In our final experiment, we use these insights to ultimately explain when algorithmic advice-takers are viewed to be motivated by fairness or not. We show that one can be perceived as both more or less guided

by fairness when taking algorithmic vs. human advice, depending on the nature of the (primary vs. secondary) information in the situation.

2. Present research

The present research contributes to the literature on fairness, advice seeking, and human-algorithm interaction by taking a closer look at the social consequences of taking advice from algorithms as opposed to humans. We delve deeper into the complex perceptions and expectations that occur when people observe others interacting with algorithms, highlighting the role that these new technologies have on interpersonal judgments, in particular goal inferences and fairness perceptions. Note that all studies' (except Study 2) hypotheses and analysis plans were pre-registered (see Appendix for direct links to pre-registrations). We report how we determined the sample size, all data exclusions (if any), all manipulations, and all measures in the study. Materials, data, and analysis code can be accessed at https://osf.io/63ua2/

3. Studies 1A, 1B, and 1C

We first conducted three studies that tested whether people judge the goals of advice seekers differently depending on whether they took advice from an algorithm or another individual. We focused on three areas of interest where algorithms are increasingly likely to act as advisors: investing (1A), healthcare (1B), and public policy (1C).

3.1. Method

3.1.1. Participants

Participants were recruited from Prolific and were from the US (Studies 1A and 1B) and UK (Study 1C). In all three studies, we aimed to recruit about 300 people in total. Following our pre-registration plans, we excluded participants who failed an attention check shown at the end of the survey. See Table 1 for more details about the samples.

3.1.2. Procedure

Participants read a description of a situation where the advice provider was either a human or an algorithm (between-subjects design). For Study 1A, the scenario read:

> Mr. Jones recently had a large financial windfall and considers investing it in a private company. Out of all companies that he takes into consideration, he can only invest in one.

Table 1. Sample details of Study 1A, 1B, and 1C.

	Study 1A	Study 1B	Study 1C
Final total N	295	290	298
Excluded n	5	8	2
Mean Age (SD)	33.5 (10.8)	32.3 (11.0)	36.1 (11.4)
Percentage female	53%	50%	72%

Being out of his depth, Mr. Jones solicits the advice of a close friend [an algorithm] who suggests investing in TOMRA™, a company providing sensor-based sorting in the waste and metal industries. This company has been praised repeatedly in recent years for its positive impact on the environment. Mr. Jones decides to follow his friend's [the algorithm's] advice and invests in TOMRA™.

DV: How much do you think Mr. Jones was guided by maximizing profit when investing in TOMRA™?

For Study 1B, the scenario read:

Dr. Jones recently received a patient who presented with severe headaches and is considering which medication to prescribe. Out of all the medications that he takes into consideration, he can only prescribe one.

Before he makes his decision, Dr. Jones solicits the advice of a close colleague [medical algorithm] who suggests prescribing the patient TOMRA™, a drug based on acetaminophen and tromethamine. The pharmaceutical company making the medication has been criticized repeatedly in recent years for actively lobbying and paying doctors. Dr. Jones decides to follow the colleague's [algorithm's] advice and prescribes TOMRA™.

DV: "How much do you think Dr. Jones was guided by maximizing the patient's welfare when prescribing TOMRA™?

Finally, for Study 1C, the scenario read:

Mr. Jones is a local politician and city planner who needs to determine which neighborhood should receive funding for infrastructure development. Money for development is limited so out of all the neighborhoods, the council can only choose one.

Before he makes his decision, Mr. Jones solicits the advice of a party member [planning algorithm] who suggests that the optimal investment location is the Southwick neighborhood. In the previous election, the Southwick neighborhood was most likely to vote for Mr. Jones' party. Mr. Jones decides to follow the party member's [algorithm's] advice and proposes to invest in the Southwick neighborhood.

DV: How much do you think Mr. Jones was guided by finding the most optimal investment location when proposing to invest in the Southwick neighborhood?

All dependent variable (DV) questions were measured on a 1 (*not at all*) to 9 (*completely*) scale. Following our reasoning, for Study 1A, we predicted that the person who solicited and followed an algorithm's (as opposed to a friend's) advice to invest in the company will be seen as more guided by maximizing profit. We expected this as the algorithm's primary goal in the situation is to find the most profitable investment opportunity, ignoring an auxiliary information related to the company's environmental friendliness in its advice. In turn, observing an investor acting upon the algorithm's advice signals to observers that the person accepts and adopts the reasoning that led the algorithm to favor the investment option, including its inattention or disregard for the company's environmental impact. Human advisors, on the other hand, are believed to take the additional information into account in their recommendation and the person accepting the recommendation is likewise seen as someone who considered it when making their investment decision.

For Study 1B, we predicted that the doctor who solicited and followed an algorithm's (as opposed to a colleague's) advice to prescribe medication from the drug company will be seen as more guided by maximizing a patient's welfare. That is, in this scenario, the algorithm's primary goal is to find the most suitable drug for the patient, ignoring in its recommendation the financial benefits that prescribing this drug may bring. Observing the doctor acting upon the algorithm's recommendation, in turn, signals to observers that he accepts and adopts the reasoning that led the algorithm to favor the suggested drug, including its disregard for the benefits it may bring him.

Finally, for Study 1C, we predicted that the politician who solicited and followed an algorithm's (as opposed to another party member's) advice to invest in the suggested neighborhood will be seen as being more guided by finding the optimal investment location. We expected this as the algorithm's primary goal in this situation is to find the optimal funding location, ignoring the voting history of the particular neighborhood in its recommendation. Observing the politician acting upon the algorithm's recommendation, in turn, signals to observers that she accepts and adopts the reasoning that led the algorithm to favor the suggested location, including its disregard for the neighborhood's voting history.

4. Results

In Study 1A, we found the predicted effect of the advice provider, $t(292) = 3.19$, $p = 0.002$, $d = .37$. More specifically, Mr. Jones was judged as being more guided by maximizing profit when investing in the environmentally friendly company following the advice of an investment algorithm ($M = 6.18$; $SD = 1.95$) as opposed to a close friend ($M = 5.45$; $SD = 1.97$). In Study 1B, we similarly found the predicted effect, $t(287) = 2.10$, $p = 0.04$, $d = .25$. Dr. Jones was judged as being more guided by maximizing his patient's welfare when the prescribed drug that would benefit him came at the suggestion of a medical algorithm ($M = 5.07$; $SD = 2.27$) as opposed to a close colleague ($M = 4.52$; $SD = 2.17$). Finally, in Study 1 C, we again found the expected effect of the advice provider, $t(261) = 6.76$, $p < 0.001$, $d = .79$. Mr. Jones was judged as being more guided by finding the most optimal investment location when he solicited and followed the advice of the planning algorithm ($M = 6.45$; $SD = 1.63$) as opposed to a party member ($M = 4.88$; $SD = 2.29$) (see Figure 1).

5. Study 2

As predicted, people who took advice from an algorithm (as opposed to another human) were judged differently in terms of their goals, even though they solicited and followed the same advice. We propose that the primary goals that algorithmic advisors are designed to attain are attributed to advice seekers when explaining their decisions, leaving less room for other, secondary motives that could account for their actions. Such secondary goals are likely to affect advice seekers' decisions when the advice comes from human advisors. More specifically, in all three of the previous scenarios,
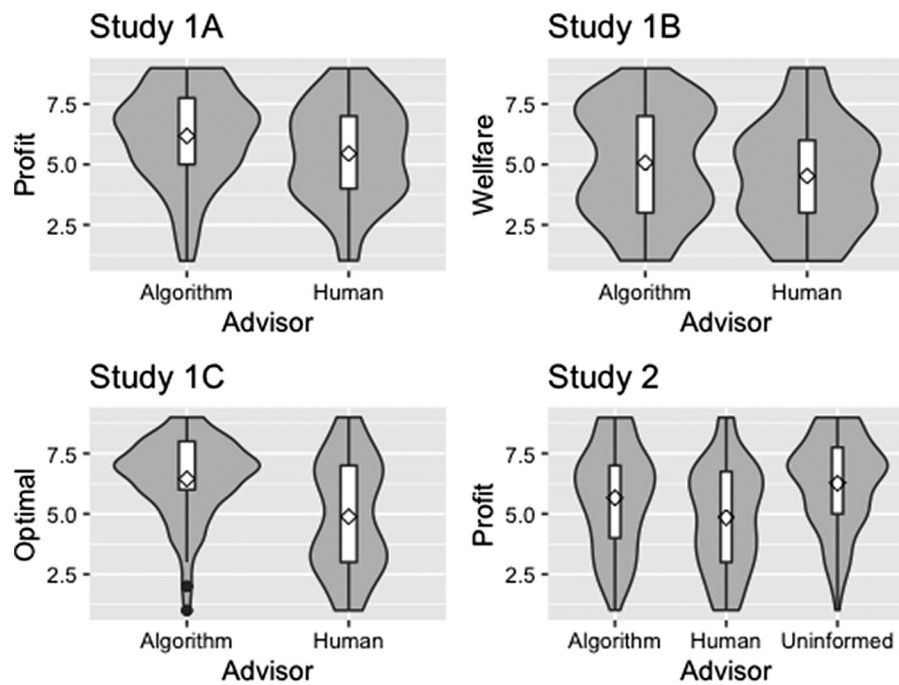
Figure 1. Violin and box plots of participants' ratings by condition for Studies 1A, 1B, 1C and 2. Note that the diamond within the boxplots represents the mean.

there was always information available regarding a secondary goal that could account for the human advisor's recommendation (e.g., environmental friendliness of the company). Crucially, we propose that algorithmic advisors are thought to disregard this additional information, predominantly focusing on attaining the situation's primary goal (e.g., finding profitable investment opportunities). In the current experiment, we tested this notion directly. If our reasoning is accurate, when a human advisor's situation is similar to how an algorithm is believed to approach the situation, an observer's judgment of these two advisors should be similar. That is, both ought to be perceived as pursuing the situation's primary goal, disregarding the additional secondary information.

More specifically, in Study 2, we used a similar design as in Study 1A, but added a third experimental condition where a human advisor is not aware that the company is an environmentally friendly company when suggesting this option as an investment opportunity. This manipulation should make the human advisor more similar to the investment algorithm in what information they considered in their recommendation.

5.1. Method

5.1.2. Participants and procedure
UK participants were recruited from Prolific and were randomly assigned to one of three experimental conditions (the same two experimental conditions as in experiment 1A, plus one additional condition in which the human advisor was explicitly unaware of the company's impact on the environment). The procedure was the same as in the previous studies and after exclusions we were left with 445 participants (71% female; $M_{Age} = 34.8$, $SD_{Age} = 12.7$). The only difference in the additional condition was that participants were

told: "Unbeknownst to the close friend, this company has been praised repeatedly in recent years for its positive impact on the environment." Additionally, after responding to the main DV about profit maximization, on a separate screen, participants in all three conditions were asked an additional question: "To what extent do you think the close friend [algorithm] took the positive environmental impact into account when suggesting TOMRA^{TM}?" using the same scale as for the main DV question.

6. Results

We ran a one-way-ANOVA on the measure of profit maximization. There was an overall effect at, $F(2, 442) = 18.21$, $p < 0.001$, $\eta_p^2 = .04$. Post-hoc tests with Bonferroni corrections confirmed that, as expected, Mr. Jones was judged as more profit maximizing when he solicited and followed the advice of the algorithm ($M = 5.66$; $SD = 2.09$) than the human advisor who (presumably) knew about the company's environmental impact ($M = 4.85$; $SD = 2.12$) ($p = 0.002$). Mr. Jones was judged as most profit maximizing when he solicited and followed the human advisor who (explicitly) did not know the environmental impact of the company ($M = 6.27$; $SD = 1.85$)—somewhat more than the algorithm ($p = 0.03$) and much more than the condition where the human advisor (presumably) knew the company's impact on the environment ($p < 0.001$) (see Figure 1).

Crucially, when we looked at the question to what extent the environmental friendliness was taken into account by the advisor, we found patterns consistent with our proposed account. A one-way ANOVA showed a significant overall effect, $F(2, 442) = 35.70$, $p < 0.001$, $\eta_p^2 = .07$. Post-hoc tests confirmed that, as expected, the human advisor who (presumably) knew about the company's environmental impact was judged as taking it into account in their advice the most

($M = 6.53$; $SD = 1.89$) as compared to the algorithm ($M = 5.55$; $SD = 2.31$). Finally, the human advisor who explicitly did not know the company's impact on the environment was judged as taking it into account even less ($M = 4.27$; $SD = 2.65$) (all three conditions differed significantly from each other, all $ps < 0.001$).

7. Study 3

So far, we focused on situations that stride the tension between self-serving and other-serving goals (e.g., investing purely for profit vs. taking into account environmental concerns or doctors caring about a patient's welfare vs. prescribing the patient a drug for personal gain). An important consequence of pursuing self-serving vs. other serving goals is how fair an advice seeker is viewed by others. More specifically, in Study 3, we aimed to demonstrate that judgments on whether advice seekers are guided by fairness concerns can differ depending on the nature of the potential secondary motive that can account for people's decisions. That is, the results of our prior studies suggest that algorithmic advisors are believed to discount any secondary information in their advice (and by association, those who follow the algorithm's advice are similarly seen as ignoring this information in their decisions). Human advisors, on the other hand, are expected to be guided in their advice by the presented secondary information and those who follow their advice are therefore similarly seen as being guided in their decisions by this information. These dynamics imply that changing the nature of the situation's secondary information will also change people's responses to those who follow algorithmic- vs. human advice. That is, when the secondary information implies some insincere, self-serving motivation (e.g., investing in a city area where most of a politician's constituents reside), those who follow an algorithmic advisor are therefore expected to be seen as *more fair* than those who follow a human advisor suggesting the option. Conversely, when the secondary information implies a sincere motivation to help others (e.g., investing in a city area that has been repeatedly overlooked in recent years by former city councils), those who follow an algorithmic advisor are therefore expected to be seen as *less fair* than those who follow a human advisor suggesting the option.

7.1. Method

7.2. Participants and procedure

Participants from the UK were recruited from Prolific. The study builds on the scenario used in Study 1C. We aimed to recruit 600 participants. After exclusions, we were left with 599 participants (50% female; $M_{Age} = 40.3$, $SD_{Age} = 13.9$). Participants were randomly assigned to a 2 (Advisor: human vs. algorithm) × 2 (Scenario: unfair proposal vs. fair proposal) between-subject design.

Overall, the procedure was the same as in Study 1C. In the "unfair proposal" scenario, participants were asked to read the same scenario as in Study 1C (i.e., where the politician chose to invest in an area where many of their constituents reside at the advice of either a planning algorithm vs. a party member). In the "fair proposal" condition, participants read the same scenario, except participants were now told: "The Southwick neighborhood is the most racially diverse neighborhood in the county and has been repeatedly overlooked in recent years by former city councils." After reading the scenario, participants were asked to provide their response to two dependent variables on a 1 (*not at all*) to 9 (*completely*) scale. The two questions were presented in a random order. Specifically: "How much do you think Mr. Jones was guided by fairness principles when investing in the Southwick neighborhood?" and "How much do you think Mr. Jones was guided by equality principles when investing in the Southwick neighborhood?". We computed a "perceived fairness" composite score by averaging the scores on the two DV's as they were highly correlated $r(597) = .79$, $p < 0.001$.[1]

8. Results

We conducted a 2 (Advisor: human vs. algorithm) × 2 (Scenario: unfair proposal vs. fair proposal) between-subject ANOVA. There was a main effect of advisor, $F(1, 595) = 7.12$, $p = 0.008$, $\eta_p^2 = .01$, and a main effect of scenario, $F(1, 595) = 255.70$, $p < 0.001$, $\eta_p^2 = .30$. Crucially, as predicted, there was a significant interaction, $F(1, 595) = 62.17$, $p < 0.001$, $\eta_p^2 = .09$. Looking at the "unfair proposal" scenario, there was a significant effect of the advisor, $F(1, 292) = 52.68$, $p < 0.001$, $\eta_p^2 = .15$. Specifically, Mr. Jones was perceived as more fair when investing in a "pro-party neighborhood" at the advice of a planning algorithm ($M = 4.63$; $SD = 1.89$) as opposed to at the advice of a party member ($M = 3.08$; $SD = 1.78$). Conversely, looking at the "fair proposal" scenario, there was also a significant effect of the advisor, but in the opposite direction, $F(1, 292) = 14.39$, $p < 0.001$, $\eta_p^2 = .05$. That is, now Mr. Jones was perceived as more fair when the investment in the disadvantaged area
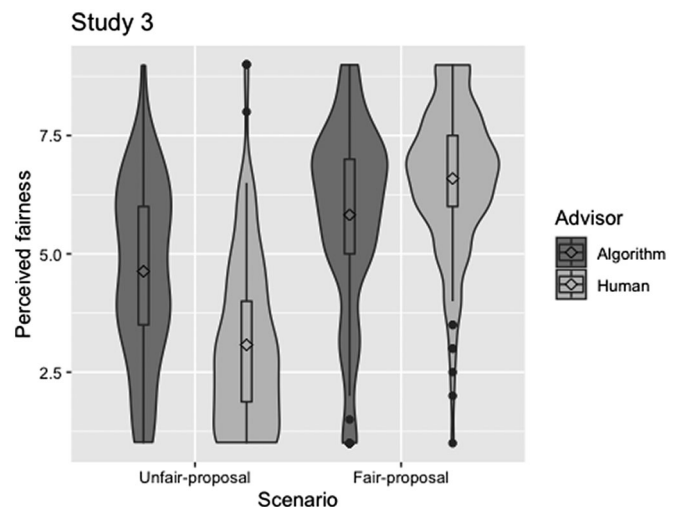
Figure 2. Violin and box plots of participants' ratings on perceived fairness as a function of scenario and advisor conditions for Study 3. Note that the diamond within the boxplots represents the mean.

followed at the suggestion of a party member ($M = 6.59$; $SD = 1.53$) as opposed to a planning algorithm to ($M = 5.82$; $SD = 1.97$) (see Figure 2).

9. General discussion

While in the past people primarily consulted with others for advice, nowadays it is increasingly likely to obtain advice from algorithms. The present research is an initial step towards understanding the social consequences of taking algorithmic advice by examining the goals that advice seekers are believed to pursue when they follow algorithmic (vs. human) advice. More specifically, in five studies, we consistently find that the primary goal that an algorithm is designed to attain is attributed to advice seekers when explaining their behaviors, leaving less room for other possible motives that could account for their actions. Such secondary goals are, however, taken into account when (the same) advice comes from human advisors, leading to different judgments about advice seekers' underlying motives for their behaviors. Importantly, these inferences, in turn, influenced perceived moral characteristics, such as advice seekers' fairness motivations. In what follows, we discuss the implications, limitations, and future research directions of these findings.

The present set of studies adds to a growing stream of research showing how people's judgements of algorithms and algorithmic decision-making are fundamentally different from that of humans (Dietvorst et al., 2015; Logg et al., 2019; Önkal et al., 2009). These perceived differences also contribute to the observed findings in the current research, where algorithms are viewed as mainly pursuing the primary goal for which they are designed and deployed. Likewise, previous findings indicate that people expect machines to exhibit 'uniqueness neglect'—a tendency to treat situations in the same way, and to overlook distinctive circumstances (Longoni et al., 2019). Similar to "uniqueness neglect," people observing others taking advice from an algorithm also assumed that advice seekers did not take into consideration additional, secondary information such as the environmentally friendliness of a company.

The presented results thus not only speak to the growing literature on how algorithmic advice is judged, but also on how people taking algorithmic advice are viewed by others. This has important consequences for advice seekers. As the results of our studies show, people soliciting algorithmic- or human advice could reveal motives to others that they may not necessarily hold. For example, one could be seen as being primarily guided by profit (Study 1A), more guided by improving patient welfare (Study 1B), or less guided by fairness (Study 3) when taking advice from an algorithm as opposed to a human. These assessments by others may not match with the true motives of advice seekers. That is, are people taking algorithmic advice genuinely guided by the primary goal that the algorithm is designed to attain? Likewise, are secondary goals indeed taken into account more often by advice seekers when they take human advice? Since we did not assess the actual intentions of advice

seekers in our research, the current findings cannot speak to the issue of whether taking algorithmic- or human advice distorts observers' beliefs or not. Given the importance of holding accurate beliefs of others' intentions and motivations, an important future research direction is to explore whether observers' assessments, in fact, correspond with the actual motives that advice seekers pursue when taking algorithmic vs. human advice.

Our results also demonstrate how advice seekers who follow algorithmic (vs. human) advice are perceived in terms of a dimension highly important for the adoption of algorithms: fairness. Algorithms are increasingly being implemented in sensitive areas and algorithmic fairness has held the interest of many in recent years (both in public- and academic discourse). Although important, in the present research, we were concerned with how taking algorithmic advice affects the perceived fairness of advice seekers. More specifically, advice seekers following algorithmic (vs. human) advice were perceived as more or less fair depending on the nature of the secondary information in the situation (Study 3). This suggests that following the recommendation of an algorithm does not necessarily imply either fairness or unfairness of the advice seeker. Rather, it seems to depend (in part at least) on the source of advice (i.e., whether the advice came from an algorithm- or another person), a notion consistent with the conjecture that people's fairness perceptions are not only influenced by the outcome of other's decisions, but also by the procedures a decision-maker followed when coming to a decision (Brockner & Wiesenfeld, 1996; Lind & Tyler, 1988; Thibaut & Walker, 1975).

It is important to note that advice seekers in our scenarios took algorithmic- or human advice at their own choice. Oftentimes, however, formal procedures (e.g., mandates in various institutions) require or strongly encourage people to seek out the advice of (algorithmic) decision support systems or other individuals before arriving at a decision (Haesevoets et al., 2021). This raises an important question about whether the intentions of advice seekers are perceived differently when learning that they were forced to solicit the advice of an algorithm or another individual. That is, following attribution theory, behaviors that are freely chosen by a person are seen by others to originate from internal factors such as a person's goals. However, when a person's behavior is constrained by the situation (e.g., a doctor is mandated to solicit and follow the advice of a medical algorithm in their prescriptions), observers are more likely to infer that the person acted to comply with the external requirements, leaving less room to explain their behaviors as arising from internal factors (Jones & Davis, 1965; Ross, 1977). Future research should attempt to disentangle in more detail whether following (algorithmic or human) advice either voluntary or not impacts observers' perceptions and judgments.

Likewise, it is worth reflecting on the notion that in the current research, we assessed the judgments of participants who were in the role of third-party observers (Langer & Landers, 2021). Although third parties often play a crucial

role in many institutions (e.g., jurors in legal systems), they are typically unaffected by an actor's actions. As such, an interesting question is whether second parties (as decision receivers) would respond in a similar fashion when advice seekers make decisions on their behalf following algorithmic- vs. human advice. That is, given their involvement, second parties typically respond with more intense emotions and judgements to the actions of others (Fehr & Fischbacher, 2004) and broadening our scope to second parties would allow us to test whether this is also true when such a party views an advice seeker to take algorithmic- vs. human advice.

Similarly, it is worth noting that most human advisors in our studies were introduced as someone "close" to the advice seeker (e.g., close friend or close colleague). This raises the questions as to whether the observed effects can be explained by the nature of the advice seeker's relationship with the advisor (i.e., a close friend vs. algorithm). That is, one could argue that the differences in the observed results are due to the human advisor having a closer relationship to the advice seeker as opposed to the algorithm (and not so much due to a difference in how human- and algorithmic advisors process information). There are, however, several reasons to question the validity of an explanation based on "closeness." First, not all human advisors were explicitly introduced in our studies as someone close to the advice seeker (e.g., the political party member in Studies 1C and 3). Second, in Study 2, we included a condition where a close friend (who had not been aware of the company's environmental impact) advised the investor to invest in the respective company. The results in this condition mimicked the results of the condition where the person is advised to invest in the company at the suggestion of an algorithm. If the perceived 'closeness' of the advisor was a factor, it would follow that this condition would mimic the results of the condition where a similar close friend (who had been aware of the company's environmental impact) suggested the company as an investment opportunity. This is not to say that the identity or relationship of the advisor is, in fact, irrelevant. It could be that the reported effects may differ depending on the reputation of the human- or algorithmic advisor. For example, a trusted algorithm that one has used on prior occasions or a rather unreliable human advisor with opposing views and motives (e.g., a member of an opposing political party) could lead observers to reach different conclusions about the nature of the advice seeker' goals. Future research could disentangle further how the identity or reputation of a human- and algorithmic advisor impacts observers' judgments of those who follow their advice.

A strength of the present studies is that they involved randomized experiments, allowing to infer causal effects. However, our use of experiments is not without limitations as we mostly relied on hypothetical scenarios, limiting the potential generalizability of our findings. To move beyond scenarios, future research could test these effects in settings where people actually observe others taking algorithmic (vs. human) advice. In a similar vein, although the observed differences when taking algorithmic (vs. human) advice are consistent with a goal attribution process, it is possible that other explanations could account for these findings.[2] For example, the fact that a person seeks out an algorithm as opposed to a human for advice may provide hints at their knowledge or expertise about the situation (e.g., a lay investor seeking out the advice of an investment algorithm may already signal their expertise as an investor). Finally, we rely on scenarios that describe domains where algorithms are currently being used. Future research should extend these investigations to novel, cutting-edge domains, where algorithms are being piloted or trialed (e.g., predictive maintenance; Ton et al., 2020; Van Oudenhoven et al., 2022).

9.1. Conclusion

The present research takes the initial steps towards understanding the social consequences of following algorithmic- vs. human advice. To understand how people judge advice seekers, it is important to consider the goals that algorithmic- vs. human advisors are expected to pursue. Taking advice from an algorithm (vs. a human) led to markedly different judgments about the advice seekers' motives, highlighting the importance of considering these consequences of human-algorithm interactions.

Notes

1. In accordance with our pre-registration we also look at the results for each DV separately. The results are the same as for the combined variable and are reported in the supplementary materials.
2. Similarly, other moderators could affect the strength of the observed relationship. For instance, in an ongoing project out of the scope of the current paper, we employ some scenarios where the difference in attribution of motives between algorithmic and human advisors disappears.

Author contributions

EE and PvdC share first authorship. EE and PvdC conceived of the idea, designed the studies, collected the data, wrote the paper and reviewed the final version before submission. EE analyzed and reported the data. ŠB and MV designed the studies, wrote the paper and reviewed the final version before submission.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Emir Efendić 🔟 http://orcid.org/0000-0002-2365-0247

References

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology*, 87(1), 23–37. https://doi.org/10.1037/0022-3514.87.1.23

Ache, F., Rader, C., & Hütter, M. (2020). Advisors want their advice to be used–but not too much: An interpersonal perspective on advice taking. *Journal of Experimental Social Psychology*, 89, 103979. https://doi.org/10.1038/s41586-018-0637-6

Alexander, V., Blinder, C., & Zak, P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology. *Computers in Human Behavior*, 89, 279–288. https://doi.org/10.1016/j.chb.2018.07.026

Arkes, H. R., Shaffer, V. A., & Medow, M. A. (2007). Patients derogate physicians who use a computer-assisted diagnostic aid. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, 27(2), 189–202. https://doi.org/10.1177/0272989X06297391

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, 113(3), 329–349. https://doi.org/10.1016/j.cognition.2018.08.003

Beck, A. H., Sangoi, A. R., Leung, S., Marinelli, R. J., Nielsen, T. O., van de Vijver, M. J., West, R. B., van de Rijn, M., & Koller, D. (2011). Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science Translational Medicine*, 3(108), 108ra113. https://doi.org/10.1126/scitranslmed.3002564

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Blunden, H., Logg, J. M., Brooks, A. W., John, L. K., & Gino, F. (2019). Seeker beware: The interpersonal costs of ignoring advice. *Organizational Behavior and Human Decision Processes*, 150, 83–100. https://doi.org/10.1089/big.2016.0047

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Brockner, J., & Wiesenfeld, B. M. (1996). An integrative framework for explaining reactions to decisions: Interactive effects of outcomes and procedures. *Psychological Bulletin*, 120(2), 189–208. https://doi.org/10.1037/0033-2909.120.2.189

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, 64(2), 123–152. https://doi.org/10.1037/h0043805

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239. https://doi.org/10.1002/bdm.2155

Carroll, J. S., Wiener, R. L., Coates, D., Galegher, J., & Alibrio, J. J. (1982). Evaulation, diagnosis, and prediction in parole decision making. *Law & Society Review*, 17(1), 199. https://doi.org/10.2307/3053536

Corbett-Davies, S., & Goel, S. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*. https://doi.org/10.48550/arXiv.1808.00023

D'Acunto, F., Prabhala, N., & Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5), 1983–2020. https://doi.org/10.1093/rfs/hhz014

Davies, M., & Stone, M. J. (Eds.). (1995). *Mental simulation: Evaluations and applications*. Blackwell. https://doi.org/10.1037/xge0000033

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist*, 26(2), 180–188. https://doi.org/10.1037/h0030868

De Cremer, D., & McGuire, J. (2022). Human–algorithm collaboration works best if humans lead (because it is fair!). *Social Justice Research*, 35(1), 33–55. https://doi.org/10.1007/s11211-021-00382-z

Diab, D. L., Pui, S.-Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non-US samples. *International Journal of Selection and Assessment*, 19(2), 209–216. https://doi.org/10.1111/j.1468-2389.2011.00548.x

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. https://doi.org/10.1037/xge0000033

Efendić, E., Van de Calseyde, P. P. F. M., & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes*, 157, 103–114. https://doi.org/10.1016/j.obhdp.2020.01.008

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), 570–576. https://doi.org/10.1287/inte.1070.0309

Gall, S. N. L. (1985). Chapter 2: Help-seeking behavior in learning. *Review of Research in Education*, 12(1), 55–90. https://doi.org/10.3102/0091732X012001055

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes*, 107(2), 234–245. https://doi.org/10.1016/j.obhdp.2008.03.001

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science (New York, N.Y.)*, 315(5812), 619–619. https://doi.org/10.1126/science.1134475

Guntzviller, L. M., Liao, D., Pulido, M. D., Butkowski, C. P., & Campbell, A. D. (2020). Extending advice response theory to the advisor: Similarities, differences, and partner-effects in advisor and recipient advice evaluations. *Communication Monographs*, 87(1), 114–135. https://doi.org/10.1080/03637751.2019.1643060

Guntzviller, L. M., MacGeorge, E. L., & Brinker, D. L., Jr. (2017). Dyadic perspectives on advice between friends: Relational influence, advice quality, and conversation satisfaction. *Communication Monographs*, 84(4), 488–509. https://doi.org/10.1080/03637751.2017.1352099

Hadar, L., & Fischer, I. (2008). Giving advice under uncertainty: What you do, what you should do, and what others think you do. *Journal of Economic Psychology*, 29(5), 667–683. https://doi.org/10.1016/j.joep.2007.12.007

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, 119, 106730. https://doi.org/10.1016/j.chb.2021.106730

Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*, 511(7508), 220–223. https://doi.org/10.1038/nature13530

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243–259. https://doi.org/10.2307/1416950

Johnson-Laird, P. N. (2013). Mental models and cognitive change. *Journal of Cognitive Psychology*, 25(2), 131–138. https://doi.org/10.1080/20445911.2012.759935

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220–266). Academic Press.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10, 113–174. https://doi.org/10.1093/jla/laz001

Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior, 123*, 106878. https://doi.org/10.1016/j.chb.2021.106878

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Liljenquist, K. A. (2010). *Resolving the impression management dilemma: The strategic benefits of soliciting others for advice* [Doctoral dissertation]. Northwestern University.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science & Business Media.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior & Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Malle, B. F., Scheutz, M., Forlizzi, J., & Voiklis, J. (2016, March). Which robot am I thinking about? The impact of action and appearance on people's evaluations of a moral robot. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 125–132). IEEE. https://doi.org/10.1109/HRI.2016.7451743

Markovic, M. (2019). Rise of the robot lawyers. *The Arizona Law Review, 61*(2), 325.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization, 52*(2), 267–275. https://doi.org/10.1016/S0167-2681(03)00003-9

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. https://doi.org/10.1037/11281-000

Niszczota, P., & Kaszás, D. (2020). Robo-investment aversion. *PloS One, 15*(9), e0239277. https://doi.org/10.1371/journal.pone.0239277

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown/Archetype.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making, 22*(4), 390–409. https://doi.org/10.1002/bdm.637

Porter, J. (2018, October 10). Robot lawyer DoNotPay now lets you 'sue anyone' via an app. *The Verge*. https://www.theverge.com/2018/10/10/17959874/donotpay-do-not-pay-robot-lawyer-ios-app-joshua-browder

Richardson, S. (2022). Exposing the many biases in machine learning. *Business Information Review, 39*(3), 82–89. https://doi.org/10.1177/02663821221121024

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Academic.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making, 19*(1), 43–60. https://doi.org/10.1002/bdm.514

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes, 62*(2), 159–174. https://doi.org/10.1006/obhd.1995.1040

Spiegel, J. S. (2012). Open-mindedness and intellectual humility. *Theory and Research in Education, 10*(1), 27–38. https://doi.org/10.1177/1477878512437472

Stacey, D., Légaré, F., Lewis, K., Barry, M. J., Bennett, C. L., Eden, K. B., Holmes-Rovner, M., Llewellyn-Thomas, H., Lyddiatt, A., Thomson, R., & Trevena, L. (2017). Decision aids for people facing health treatment or screening decisions. *The Cochrane Database of Systematic Reviews, 4*(4), CD001431. https://doi.org/10.1002/14651858.CD001431.pub5

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society, 9*(2), 205395172211151. https://doi.org/10.1177/20539517221115189

Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Lawrence Erlbaum.

Ton, B., Basten, R., Bolte, J., Braaksma, J., Di Bucchianico, A., van de Calseyde, P., Grooteman, F., Heskes, T., Jansen, N., Teeuw, W., Tinga, T., & Stoelinga, M. (2020). PrimaVera: Synergising predictive maintenance. *Applied Sciences, 10*(23), 8348. https://doi.org/10.3390/app10238348

Van Oudenhoven, B., Van de Calseyde, P., Basten, R., & Demerouti, E. (2022). Predictive maintenance for industry 5.0: Behavioural inquiries from a work system perspective. *International Journal of Production Research, 0*(0), 1–20. https://doi.org/10.1080/00207543.2022.2154403

About the authors

Emir Efendić is an assistant professor at the School of Business and Economics Maastricht University. He studies how people make judgments and decisions, focusing on human-computer interaction, risk, and methods.

Philippe P. F. M. Van de Calseyde is an assistant professor at the Eindhoven University of Technology. His background and field of interest is mainly in the area of human judgment and decision-making.

Štěpán Bahník is an assistant professor at the Faculty of Business Administration of the Prague University of Economics and Business. His research focuses on judgment and decision-making, especially in the domain of morality.

Marek A. Vranka is a researcher at the Faculty of Business Administration of the Prague University of Economics and Business. His research focuses on moral judgment and social norms nudging.

Appendix

Links to study pre-registrations:
   Study 1a: https://osf.io/uba6e
   Study 1b: https://osf.io/v65sh
   Study 1c: https://osf.io/tmf6y
   Study 3: https://osf.io/qfc45