Jan 17th, 12:00 AM

Just like you like it - The effects of transparency and decision outcome on the evaluation of human and algorithmic decision-making

Jessica Ochmann
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, jessica.ochmann@fau.de

Just like you like it - The effects of transparency and decision outcome on the evaluation of human and algorithmic decision-making

Jessica Ochmann

Friedrich-Alexander Universität Erlangen-Nürnberg, School of Business and Economics, Institute of Information Systems, Nürnberg, Germany
jessica.ochmann@fau.de

Abstract. Algorithms are increasingly offered for human decision-making processes to support individuals with sophisticated data-driven insights in a variety of decision domains. Despite this promising potential, prior findings on the perceptions of algorithmic decision-making are ambiguous. This has led to a vivid discussion regarding the comprehensibility and fairness perceptions associated with human and algorithmic decision agents that also emphasized the role that decision-related factors can play in determining the evaluation of a certain decision. In a preregistered online experiment with 400 participants, we find that differences in the perceptions of decision transparency and fairness can be rather ascribed to the processes and rules applied to arrive at the decision and the decision outcome's sentiment than to the decision agent. However, being confronted with a negative decision outcome in a situation that is characterized by uncertainty, leads to a preference for human decision agents.

Keywords: Algorithms, decision-making, transparency, fairness perceptions.

1 Introduction

When making decisions, individuals are increasingly confronted with algorithms aiming to support their reasoning processes and guide them towards the 'right' choice. These algorithms are capable of analyzing large data sets autonomously and enable their users to recognize patterns and derive novel insights for a variety of tasks. Consumers, for example, can rely on algorithms to find a partner, receive product recommendations, and manage their finances [1]. Next to these applications in the end-user domain, organizations have also recognized the great potential of algorithm supported decision-making and included sophisticated forecasting models in their internal processes, such as hiring, promotion, and credit decisions [2, 3].

In light of their predictive power and ability to even outperform human decision-makers [4, 5], algorithms are often readily accepted by individuals to support them in both their professional and private lives, especially for numeric tasks with objectively correct answers [6]. However, there are some decision domains in which people are reluctant to use and rely on algorithmic decision agents. When it comes to medical decisions, for example, individuals hesitate to use algorithms [7] and expect medical staff to rely on their own judgments instead of consulting an algorithm [8]. In addition, other empirical evidence suggests that people prefer human recommendations over those given by algorithms [6], with regard to investment recommendations [3] or the ability to predict which joke a person prefers [5]. These ambiguous results indicate that the use of algorithms in human decision-making is not only complex from a technical, but also from a behavioral and sociotechnical perspective. Up until now, no scientific consensus has been reached regarding the general attitude of humans towards algorithm-supported decision-making – exemplarily reflected by the two seemingly contradicting research streams of algorithm aversion [6, 9] and algorithm appreciation [10].

In the present study, I aim to contribute to prior research on attitudes and perceptions associated with algorithm-supported decision-making. More precisely, I am interested in evaluating whether the simple fact that either a human or an algorithm made a decision leads to different perceptions among individuals or if other decision-related factors determine the evaluation of human and algorithmic decision processes. In line with prior literature, I propose such additional factors, namely the processes and rules applied to arrive at the decision and the decision outcome's sentiment [11]. Thus, making sense of a decision's outcome might be - next to the general perception of the decision agent - a key determinant for its evaluation. This assumption aligns with prior research showing that individuals might be suspicious of algorithmic decision agents as they are not able to relate to their underlying processes and believe that human decision outcomes are easier to understand [5]. Based on this reasoning, I propose that increasing the comprehensibility and transparency of algorithmic decision-making processes might enable individuals to better understand the respective decision outcomes. Indeed, aligning the amount of information provided for algorithmic decision-making processes with the amount of information individuals would expect from a human could lead to similar decision transparency of human and algorithmic decision agents. To verify the proposed approach, I aim to answer the following research question:

RQ$_1$: *How do individuals' decision transparency perceptions of human and algorithmic decision agents differ when receiving additional information on the decision outcome's underlying processes?*

Recently, a vivid debate has emerged among scientists that research on algorithm-supported decision-making requires the consideration of fairness aspects as the use of algorithms in decision-making might cause unintended consequences, such as showing discriminatory tendencies [12–15]. To comply with guidelines for research ethics, the evaluation of decision transparency regarding human and algorithmic decision agents, thus, cannot be made without discussing the implications for the individual's subjective fairness perceptions. While prior research on algorithmic fairness has predominantly focused on the technical perspective to develop unbiased algorithms [12, 16], research on subjective fairness perceptions is still scarce and characterized by inconclusive findings regarding the preference for human and algorithmic decision agents [13]. Therefore, I aim to answer the call for a thorough empirical understanding of individual's fairness perceptions. To do so, I investigate one possibility to increase fairness perceptions of algorithm-supported decision-making: transparently communicating the decision process. Prior research has shown that perceptions of unfairness often result from the feeling that individuals are not able to understand how algorithmic decision agents evolve a certain decision [5, 17]. Indeed, transparent decisions should lead to similar evaluations of the decision in regards to fairness perceptions - regardless of whether the decision was made by a human or an algorithm. In addition, the results of prior scholarly work have shown that a decision's outcome affects its associated fairness perceptions, whereby positive outcomes are more likely to receive higher fairness ratings than negative outcomes [18]. Therefore, I aim to answer the following research question:

RQ$_2$: *How do individuals' fairness perceptions of human and algorithmic decision agents differ when receiving additional information on the decision outcome's underlying processes and being confronted with a positive decision outcome?*

To shed light on these research questions and to test the corresponding hypotheses, I conducted a preregistered survey-based online experiment with 400 participants. All participants read about either a positive or a negative credit decision that a financial institute makes. Depending on the experimental condition, participants further received additional information on the decision-making process. In addition, participants learned that the financial institute can either have an employee or have an algorithm decide whether a customer qualifies for a loan and expressed, in random order, their transparency and fairness perceptions of the respective decision-maker. This work, thereby, contributes to ongoing IS research by empirically investigating how individuals react to human and algorithmic decision agents in differing decision conditions. Thus, the suggested perspective extends prior research by investigating if decision-related factors can help to explain decision transparency and fairness perceptions.

This article proceeds as follows: First, I present related work on the interaction between human and algorithms and discuss the implications for fairness perceptions along with the hypotheses. This section is followed by a description of the experimental design and results. The article closes with a discussion of the findings and the resulting scholarly and practical implications.

2 Theoretical background and hypotheses development

2.1 Research on human-algorithm interaction

Integrating algorithmic support in human decision-making processes has the great potential to increase decision quality and ease the cognitive burden on individuals. For this reason, scholars have developed a great interest in the theoretical mechanisms underlying human decision-maker's acceptance of algorithms in various domains (e.g., finance, health, human resources) [10]. Recent evidence in this research stream suggests that individuals often prefer human over algorithmic advice – despite algorithms outperforming human reasoning in multiple forecasting tasks. This phenomenon has been coined algorithm aversion in prior work [4]. Önkal et al. [19], for example, have shown that individuals are more likely to adjust stock price forecasts when they stem from statistical forecasting methods compared to human reasoning. Dietvorst et al. [9] found a similiar tendency showing that individuals who see an algorithm err subsequently prefer human over algorithmic forecasts – even when being aware that humans also make mistakes.

This general aversion towards algorithms can result from algorithmic opacity [20–22], loss of control [15, 23, 24], human's overconfidence of their own decision-making [4], ethical concerns [25], and perceived lack of algorithm capabilities and ability to empathize [7, 26]. The exact mechanisms leading to algorithm aversion, however, are yet not fully understood [6]. While a considerable amount of empirical research has found evidence for the existence of algorithm aversion and identified approaches to overcome it, such as considering human decision-makers in the forecasting process and enable them to adjust the forecast [27] or modifying the algorithm [4], recent studies showed the opposite effect of algorithm appreciation for numeric tasks [6], social close human agents [10], and financial advice [28].

Summarizing, research on both, algorithm aversion and algorithm appreciation, compares the reactions towards human and algorithmic decision-makers. However, this research stream has particularly focused on the human decision maker that has to rely on algorithms in decision making processes [10]. While this work is relevant for the design of sustainable human-AI-interactions, it offers limited insights for individual's reactions towards algorithmic decision-making when they are exposed to algorithmic decision agents and do not use the technology actively [26]. Indeed, being an active user of algorithms and being exposed to an algorithm's outcome might result in a different set of behavioral patterns. For this reason, I concentrate on the evaluation of human and algorithm decision-making in situations were the rating individual is not the decision-maker, but potentially affected by the decision. In addition, I aim to examine if the different perceptions of human and algorithm decision-makers in prior literature are grounded in the nature of the decision-maker itself or in other variables, such as the decision-making process and the decision outcome.

Prior research suggests, for example, that the evaluation of a certain decision might result from both, the consistent application of procedures across people and time and the decision outcome [11, 29, 30]. Following this argumentation, the decision-maker does not necessarily determine the evaluation of a certain decision if the underlying decision-making process is clearly communicated and the decision outcome is comprehensible. Building on this line of reasoning, the identical amount of information on the decision-making process should lead to equal transparency perceptions of a human's vs. an algorithm's decision given that the rating individual receives identical information on the decision-making underlying processes. In particular, the degree of perceived decision transparency should be predominantly driven by transparently communicating the underlying decision-making processes. Therefore, I hypothesize:

$\mathbf{H_{1a}}$: *Individuals will voice higher levels of decision transparency when subjected to decision-making processes with additional information.*

In contrast, there are reasons to believe that decisions made by algorithms may generate greater concerns than human decisions when no further information on the decision-making process is provided. Recent scholarly research has shown that individuals expect algorithms to decide differently than human decision-makers. While human decision-makers are associated with the ability to take individual circumstances into account and weight decision outcomes with regard to moral rules [9], algorithms are perceived to be reductionistic [26]. In addition, individuals assume algorithms to be more likely to violate moral rules as they are usually generated to maximize a predefined objective [1]. As a result, the decision procedures of algorithms will be reduced to numerical representations and lead to the assumption to be less accurate and barely comprehensible. Therefore, without any additional information provided on the decision-making process, individuals should rate human decisions as more comprehensible, such that I assume:

H_{1b}: *Individuals will voice lower levels of decision transparency for an algorithm's decision compared to the same decision made by a human when subjected to decision-making processes with no additional information.*

Directly related to transparency perceptions of decision-making processes are fairness considerations associated with the decision. Especially individuals who are not directly included in the decision-making process might have different perceptions regarding the decision fairness depending on the agent who made the decision. For organizations, these perceptions can be critical in determining their external perception by consumers and other important stakeholders. Therefore, it is important to determine how human and algorithmic decision agents differ in how fair and ethical they are perceived.

2.2 Fairness concerns regarding human-algorithm interactions

Fairness, in general decision situations, is defined as a social construct that attributes subjective perceptions of equity, equality, needs, underlying procedures, and interpersonal treatment to objective aspects of a certain decision [11]. With the increasing use of algorithms in human decision-making, fairness has also become a key determinant for the development of algorithmic decision-making [31]. Indeed, algorithm fairness is listed as a main requirement for the development of reliable artificial intelligence (AI) [12]. While prior research predominately focuses on approaches to advance algorithm fairness from a technical perspective to generate unbiased decision outcomes [21, 22], there is a relatively small body of literature that is concerned with individual's subjective fairness perceptions of human vs. algorithm decision-making (for a review see [13]). However, as the use of algorithms has also societal implications for the humans that are exposed to a particular decision outcome, it is crucial to investigate individual's fairness perceptions of algorithm decisions.

Recent research has recognized the importance of this demand and investigated algorithm fairness perceptions in the field of human resources [2, 26, 32], criminal justice [33], and resource allocation decisions [34]. In these studies, scholars provided several evidence that individuals rate decisions made by humans significantly fairer than those made by algorithms. The reason for this directive might be, for example, that individuals attribute algorithms to be reductionistic and not able to consider individual circumstances in a certain decision. As a result, individuals might assume that algorithms consider less accurate information and are unfairer than humans [26]. In contrast, other scholars have shown that a large majority of participants voice higher fairness levels for algorithms when they are directly confronted with the question if a human or an algorithm makes fairer decisions [35, 36]. The preference for algorithmic decision agents could be the result of the expectation that algorithms - provided that they fulfill fairness requirements from a technical perspective - can remove human bias in decision-making processes and, thereby, increase the objectivity of the decision outcome [21, 22, 37]. As the elimination of bias is one of the key elements for fairness perceptions [11], algorithmic decisions that are characterized by mathematical logic, might be perceived as fairer compared to a decision made by a human. To conclude, individuals who are affected by a human or an algorithmic decision agent might be exposed to both, the concern of reductionistic algorithms and the awareness of unbiased algorithms that suppress human biases by relying on objective mathematical rules. Based on this reasoning, I assume that individuals will balance these two assumptions such that fairness ratings of humans and algorithm should be similar.

In addition, I theorize that - similar to decision transparency - individuals' fairness perceptions are not necessarily related to the decision agent. In prior research, numerous scholars found evidence that individual fairness perceptions result merely from the allocation rules and the procedures applied, and the decision outcome [38]. Following this reasoning, the individual's rating of decision fairness should be predominately determined by decision-related characteristics such as the decision processes and the decision outcome and not by the decision agent. As favorable decision outcomes are more likely to receive higher fairness ratings [26, 39], I hypothesize:

H_{2a}: *Additional information on the decision process and a positive decision outcome positively influence the individuals' fairness perceptions.*

As additional information is expeected to have a positive effect on fairness perceptions and it was previously theorized that information positively influence individuals' perceptions of decision transparency, I consequently assume that:

H_{2b}: *Decision transparency positively influence the individuals' fairness perceptions.*

3 Methodology

To empirically test the hypotheses, I employed a survey-based online experiment in which participants evaluated the credit decision of a financial institute. The study was preregistered and complied ethical requirements as approved by the publication ethics commissioner of the university department.

3.1 Experimental design

Participants who passed the attention check read about the decision of a financial institute whether or not a customer qualifies for a loan. Participants further learned that the financial institute can either have an employee or an algorithm making this credit decision. In the survey, participants completed two tasks, where they evaluated the same credit decision of a human and algorithmic decision maker in randomized order. Depending on the condition, participants received additional information on the employee's and the algorithm's decision-making process [The employee/algorithm made this decision by weighting objective criteria such as age, gender, and income and using prior data of similar customers to calculate the creditworthiness.] and read about either a positive [the customer qualifies for a loan] or a negative [the customer doesn't qualify for a loan] credit decision. Next, participant rated how transparent and fair the decision of the employee and the algorithm is.

The experiment consisted of a two-factor between-subjects design. The two factors were whether the the decision-making process was transparently communicated (*Information*; yes vs. no) and whether the credit decision's outcome was positive or negative (*Outcome*; positive vs. negative). The *Outcome* condition was included to measure the robustness of the effects and control whether outcome favorability changes the perception of decision transparency as suggested by prior research [26]. This design led to four experimental groups, namely "NEG_I" (negative x information), "POS_I" (positive x information), "NEG_NOI" (negative x no information), "POS_NOI" (positive x no information) to which the participants were randomly assigned by the survey tool. In addition, a within-subject design was implemented to address the decision-maker (*Decision-maker*; Algorithm vs. Employee) in a randomized order.

3.2 Measures and scales

To cover the participants' decision transparency perceptions I adapted the transparency scale by Zhou et al. [40]. In addition, I constructed a 6-items decision fairness scale by adapting the organizational justice scale from Conlon [41]. All scales showed a very good internal consistency (Cronbach's α >= .92) and were answered on a 7-point Likert-Scale reaching from 1 = "Strongly disagree" to 7 = "Strongly agree" (see Table 1).

Table 1. Measurement scales and items used in the experiment.

Construct	Items
Decision transparency	(1) I could fully understand the way the [algorithm] [employee] determined the decision of whether the customer qualifies for a loan.
	(2) I have a clear idea about the way the [algorithm] [employee] determined the decision of whether the customer qualifies for a loan.
	(3) I have a clear understanding of the [algorithm's] [employee's] decision.
	(4) I am able to comprehend the [algorithm's] [employee's] decision very well.
	(5) The way the [algorithm] [employee] determined the decision of whether the customer qualifies for a loan is transparent to me.
Fairness	(1) The way the [algorithm] [employee] determined that the customer [doesn't qualify] [qualifies] for a loan seems fair.
	(2) The [algorithm's] [employee's] process for deciding that the customer [doesn't qualify] [qualifies] for a loan was fair.
	(3) The decision that the customer [doesn't qualify] [qualifies] for a loan was fair.
	(4) The outcome of the decision that the customer [doesn't qualify] [qualifies] for a loan was fair.
	(5) The [algorithm] [employee] made this decision in an unbiased and neutral manner.
	(6) The [algorithm] [employee] treated the customer with dignity and respect in making this decision.

3.3 Participants

I preregistered that I would recruit 400 participants from Prolific. After posting the study, 438 participants clicked on the study link, 38 participants failed the attention check and were not allowed to begin the study, and 400 participants completed all depended measures. Participants were randomly assigned to the four experimental conditions at the beginning of the experiment. Failed attention checks lead to slightly unbalanced randomization, however, there were no significant differences across the experimental conditions regarding demographic characteristics. All participants who completed the survey received a financial compensation.

Table 2. Demographic summary of participants in experimental groups

Characteristics		NEG_I	POS_I	NEG_NOI	POS_NOI
n		101	102	96	101
	Female	77 (76%)	79 (77%)	68 (71%)	68 (67%)
Sex	Male	23 (23%)	23 (23%)	28 (29%)	32 (32%)
	Other	1 (1%)	0	0	1 (1%)
Age mean		32.5	32.6	33.3	30.6

4 Results

4.1 Manipulation check

To measure whether the transparency manipulation successfully increased the perceived decision transparency, I adapted the data use transparency scale by Martin et al. [42]. Participants evaluated the way the algorithm respectively the employee determined the decision on a 7-point semantic differential using the items "Unclear to me - Clear to me", "Confusing - Straightforward", "Difficult to understand - Easy to understand", and "Vague - Transparent". Participants in the information conditions (M = 5.51, SD = 1.23) and in the non information conditions (M = 3.58, SD = 1.51) differed significantly regarding their perception of process transparency ($F(1,398) = 196.453$, $p < .001$). This result indicates that the transparency manipulation was successful.

4.2 Hypotheses testing

Decision transparency To test hypotheses H_{1a} and H_{1b} and compare the group means with regard to the perceived decision transparency, I ran a two-way mixed ANOVA. The model included the four experimental conditions as between-subject factors and the transparency evaluation of the decision agent (algorithm vs. employee) as a within-subject factor.

The two-way mixed ANOVA revealed significant main effects of the experimental conditions, ($F(3, 369) = 54.992$, $p < .01$), and the decision agent, ($F\ (1, 396) = 16.735$, $p < .01$) on the decision transparency expressed by the participants. In a subsequent post-hoc analysis, pairwise comparison shows that the mean transparency score was significantly different in conditions with additional information vs. no additional information on the decision-making process, thus supporting hypothesis H_{1a} and H_{1b} (see Table 3).

Table 3. Post-hoc analysis of the group differences in the experimental conditions

Group1		Group 2		*p (adj.)*
	n		*n*	
NEG_I	202	POS_I	204	n.s.
NEG_I	202	NEG_NOI	192	<.001
POS_I	204	NEG_NOI	192	<.001
NEG_I	202	POS_NOI	202	<.001
POS_I	204	POS_NOI	202	<.001
NEG_NOI	192	POS_NOI	202	n.s.

The post-hoc analysis further revealed that the only significant difference in transparency perceptions between algorithmic and human decision makers was in the NEG_NOI condition ($F(1{,}190) = 4.102$, $p < .01$) where participants experienced a negative credit decision and didn't receive further information on the decision-making process, suggesting that this significant main effect was mostly driven by this condition. This indicates that in situations under uncertainty, individuals seems to rely more on human decision agents. In contrast, in situations that are transparent and linked to a rather positive decision outcome, individuals do not seem to have any preferences for either a human or an algorithm. Figure 1 summarizes the participants' decision transparency ratings per condition.

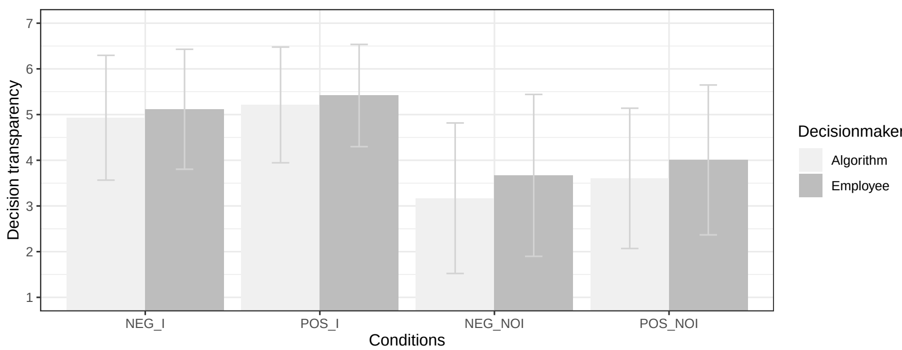

Figure 1. Transparency perceptions of algorithmic and human decision-maker per experimental group.

Fairness To determine whether the four experimental groups differed regarding the fairness perceptions and and to test hypotheses H_{2a} and H_{2b}, I lastly calculated a two-way mixed ANOVA containing the four experimental conditions as between-subject factors and the fairness perceptions of the decision-maker (algorithm vs. employee) as a within-subject factor .

The two-way mixed ANOVA revealed significant differences between the group means of the conditions with regard to the fairness perceptions ($F(3, 396) = 21.998$, $p < .01$). However, even though the descriptive statistics shows slightly higher fairness levels for the human decision agent (see Figure 2) the ratings for the algorithm and the employee were, in general, the same ($F(1, 396) = 2.555$, $p = .134$). In the post-hoc pairwise comparison, the Bonferroni adjusted p-value shows that the mean fairness score was significantly different in the conditions NEG_I vs. POS_I ($p < .001$), POS_I vs. NEG_NOI ($p < .001$), POS_I vs. POS_NOI ($p < .001$). These results indicate that the participants' rating whether they perceive the decision to be fair is rather driven by the decision outcome and additional information on the decision process then by the decision-maker.

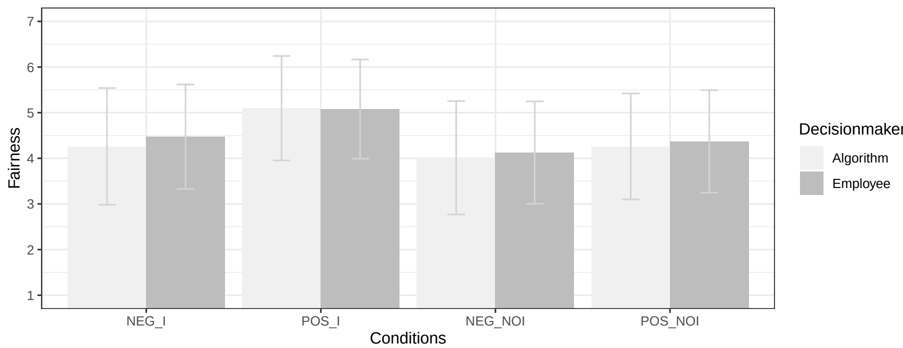

Figure 2. Fairness perceptions of algorithmic and human decision-maker per experimental group.

To shed further light on these findings, I ran a linear regression model. The model included the following independent variables as predictors of the fairness perceptions: *Information* (dummy: 1 if participant received additional information on the decision process), *Outcome* (dummy: 1 if participant was in a positive outcome group), *Decision transparency*, and the interaction term of *Information*Outcome*.

The regression model (see Table 4) for the participants' fairness perceptions revealed a significant positive main effect of *Decision transparency*, suggesting that participants who perceive a higher degree of decision transparency experience the decision to be more fair. This result is in line with H_{2b}. Interestingly, the model further revealed a significant negative main effect of additional *information* and no significant main effect of the *outcome* variable. However, the significant positive interaction effect between *Information* and *Outcome* suggests that fairness perceptions are complex and determined by the interplay of decision-related factors, rather then by the decision-maker. In particular, the interaction effect suggests that fairness perceptions are higher for decisions that are accompanied by the transparent communication of decision processes and with a positive decision outcome. This result is in line with H_{2a}.

Table 4. Regression results.

	B	*Std. Error*	β	*t*	*p*
Constant	2.516	0.128		19.575	<.001 ***
Decision transparency	0.461	0.031	0.671	15.024	<.001 ***
Information	-0.454	0.118	-0.225	-3.853	<.001 ***
Outcome	0.079	0.107	0.039	0.736	0.462
Information*Outcome	0.511	0.150	0.221	3.409	<.001 ***

*, **, *** indicate significance at 90%, 95%, and 99% level

5 Discussion

The ubiquitous opportunity for humans to use algorithms in their decision-making processes offers them the chance to autonomously analyze large data sets and derive valuable insights in various decision domains. Despite this promising potential, the ambiguous findings regarding the perceptions of algorithmic decision-making that might hinder its acceptance among individuals illustrate the need for further research [2, 26, 35, 36]. I argue, that the differing findings in prior literature might rather be grounded in decision-related factors than in the human or algorithmic nature of the respective decision agent. The present study, therefore, extends prior research by challenging the overall notion that differences in the perceptions of decision transparency and fairness result solely from the distinct evaluations of human and algorithmic decision agents [6, 9]. As prior studies predominantly concentrated on the decision-makers that apply algorithms in their own decision-making processes, this work focuses on individuals that are not actively involved in the decision-making process to shed further on individuals' adoption of algorithms.

Following this directive, I conducted a preregistered survey-based online experiment in which a sample of 400 participants was asked to rate credit decisions of a financial institute. In line with the hypotheses, I found that participants expressed - regardless of the decision agent - a higher degree of decision transparency when they received additional information on the decision outcome's underlying processes. When participants were facing a negative decision outcome and received no additional information on the decision processes, however, I found that decision transparency perceptions of human and algorithmic decision agents differ significantly. The findings further indicate that decision-related factors such as the transparent communication of the underlying processes and the decision outcome significantly shape individual's subjective fairness perceptions - regardless of whether the decision was made by a human or an algorithm. In addition, I showed that perceptions of decision transparency determine the individual's fairness ratings.

These results contribute to ongoing IS research in multiple ways. First, prior studies on the effects of algorithmic decision-making provided inconclusive results with regard to the preference for human and algorithmic decision-agents. The present paper, offers an alternative explanatory approach by highlighting the importance of the individual's ability to making sense of a decision's outcome. I further suggest that decision-related factors, namely decision transparency and decision outcome are key determinants of decision evaluation, thus illustrating that the evaluation of a decision does not solely depend on the decision agent.

Second, I developed and tested the hypotheses drawing from prior findings on decision-making processes that were not necessarily related to algorithms (e.g., [11]). Thereby, I show that findings of prior research on behavioral IS are adaptable to the context of algorithm decision-making. Thus, the results can serve as a starting point to encourage scholars to rely on existing and well-established theories to explain new phenomenons such as the perceptions of human and algorithmic decision agents and their decisions.

Third, prior literature on algorithm fairness predominantly focuses on the technical perspective of algorithmic decision-making and examines approaches to develop unbiased and fair algorithms [13, 43]. While these attempts are of particular importance for the design of sustainable human-algorithm interactions, research on the individuals' subjective fairness perceptions is limited [13]. By investigating fairness perceptions in the experiment, I extend prior research and give guidance how providers of algorithmic decision-making should communicate the underlying processes to increase fairness ratings. In addition, I show that fairness perceptions are mostly driven by decision-related factors and are not necessarily resulting from the fact that a human or an algorithm made a certain decision. This finding highlights the need for future research that investigates not only perceptions of human and algorithms, but also factors that are associated with human and algorithmic decision-making and might affect the evaluation of a certain decision outcome.

Next to these theoretical contributions, this paper provides implications for managers and practitioners. The findings on the importance of transparency and decision outcomes in human and algorithmic decision-making give organizations guidance of how decision processes should be communicated and how they can avoid reluctant behavior in situations with negative decision outcomes. Lastly, with the transparent communication of decision processes, the present study identifies an easily implementable intervention to shape individuals' perceptions of algorithmic decision making, illustrating that encouraging individuals to rely on algorithms is neither costly nor complex.

5.1 Limitations and future research

Although this research provides valuable results, I want to point out potential limitations of the present study. First, participants in the study evaluated a hypothetical credit decision that does not have real consequences for their own finances. This limitation, however, gives scholars guidance regarding the individuals' baseline perceptions towards human and algorithmic decision agents and shows that individuals do not reject algorithm decision-making per se. This baseline, thus, suggests that implementing organizations have many possibilities to shape the individuals' perceptions of algorithms. Second, the present study was conducted with participants from the UK with a sample size of 400. Participants from other countries and with a differing cultural background might express contrasting decision transparency and fairness perceptions.

5.2 Conclusion

The present study aimed at identifying whether decision transparency and fairness perceptions differ when individuals are confronted with human or algorithmic decision-making. With respect to the research questions, I found evidence that individuals prefer human over algorithms when they are confronted with a undesirable decision outcome and are not aware of the processes that were applied to derive a certain decision outcome. However, individuals show no significant preference for either a human or an algorithm when they receive additional information on the underlying decision-making processes. In line with these findings, I found that fairness perceptions are not necessarily related to the decision agent. Instead, I showed that fairness perceptions are mostly affected by decision-related factors (i.e., information and outcome) and decision transparency. To conclude, the findings give guidance on how to communicate algorithmic decision-making and suggest that decision-related factors should also be taken into account when relying on human or algorithmic decision agents.

References

1. Dietvorst, B.J., Bartels, D.M.: Consumers Object to Algorithms Making Morally Relevant Tradeoffs Because of Algorithms' Consequentialist Decision Strategies. Journal of Consumer Psychology p. jcpy.1266 (2021)
2. Acikgoz, Y., Davison, K.H., Compagnone, M., Laske, M.: Justice perceptions of artificial intelligence in selection. International Journal of Selection and Assessment 28(4), 399–416 (2020)
3. Niszczota, P., Kaszás, D.: Robo-investment aversion. PLOS ONE 15(9), e0239277 (2020)
4. Dietvorst, B.J., Simmons, J.P., Massey, C.: Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. Management Science 64(3), 1155–1170 (2018)
5. Yeomans, M., Shah, A., Mullainathan, S., Kleinberg, J.: Making sense of recommendations. Journal of Behavioral Decision Making 32(4), 403–414 (2019)
6. Castelo, N., Bos, M.W., Lehmann, D.R.: Task-Dependent Algorithm Aversion. Journal of Marketing Research 56(5), 809–825 (2019)
7. Longoni, C., Bonezzi, A., Morewedge, C.K.: Resistance to Medical Artificial Intelligence. Journal of Consumer Research 46(4), 629–650 (2019)
8. Shaffer, V.A., Probst, C.A., Merkle, E.C., Arkes, H.R., Medow, M.A.: Why Do Patients Derogate Physicians Who Use a Computer-Based Diagnostic Support System? Medical Decision Making 33(1), 108–118 (2013)
9. Dietvorst, B.J., Simmons, J.P., Massey, C.: Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General 144(1), 114–126 (2015)
10. Logg, J.M., Minson, J.A., Moore, D.A.: Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes 151, 90–103 (2019)
11. Colquitt, J.A.: On the dimensionality of organizational justice: A construct validation of a measure. Journal of Applied Psychology 86(3), 386–400 (2001)
12. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nature Machine Intelligence 1(9), 389–399 (2019)
13. Starke, C., Baleis, J., Keller, B., Marcinkowski, F.: Fairness Perceptions of Algorithmic Decision-Making: A Systematic Review of the Empirical Literature. arXiv:2103.12016 [cs] (2021), arXiv: 2103.12016
14. Gal, U., Jensen, T.B., Stein, M.K.: Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. Information and Organization 30(2), 100301 (2020), publisher: Elsevier
15. Kellogg, K.C., Valentine, M.A., Christin, A.: Algorithms at work: The new contested terrain of control. Academy of Management Annals 14(1), 366–410 (2020)
16. Friedler, S.A., Scheidegger, C., Venkatasubramanian, S.: The (Im)possibility of fairness: different value systems require different mechanisms for fair decision making. Communications of the ACM 64(4), 136–143 (2021)
17. Shin, D.: The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies 146, 102551 (2021)
18. Lind, E.A., Tyler, T.R.: The social psychology of procedural justice. Critical issues in social justice, Plenum Press, New York (1988)
19. Oenkal, D., Goodwin, P., Thomson, M., Gönül, S., Pollock, A.: The relative influence of advice from human experts and statistical methods on forecast adjustments. Journal of Behavioral Decision Making 22(4), 390–409 (2009)

20. Kumar, D., Taylor, G.W., Wong, A.: Opening the Black Box of Financial AI with CLEAR-Trade: A CLass-Enhanced Attentive Response Approach for Explaining and Visualizing Deep Learning-Driven Stock Market Prediction. arXiv:1709.01574 [cs] (2017), arXiv: 1709.01574
21. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115 (2020)
22. Adadi, A., Berrada, M.: Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 6, 52138–52160 (2018)
23. Ananny, M.: Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. Science, Technology, & Human Values 41(1), 93–117 (2016)
24. Burton, J.W., Stein, M., Jensen, T.B.: A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making 33(2), 220–239 (2020)
25. Eastwood, J., Snook, B., Luther, K.: What People Want From Their Professionals: Attitudes Toward Decision-making Strategies. Journal of Behavioral Decision Making 25(5), 458–468 (2012)
26. Newman, D.T., Fast, N.J., Harmon, D.J.: When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. Organizational Behavior and Human Decision Processes 160(April 2018), 149–167 (2020), publisher: Elsevier
27. Kawaguchi, K.: When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business. Management Science 67(3), 1670–1695 (2021)
28. Gunaratne, J., Zalmanson, L., Nov, O.: The Persuasive Power of Algorithmic and Crowdsourced Advice. Journal of Management Information Systems 35(4), 1092–1120 (2018)
29. Leventhal, G.S.: What Should Be Done with Equity Theory? In: Gergen, K.J., Greenberg, M.S., Willis, R.H. (eds.) Social Exchange, pp. 27–55. Springer US, Boston, MA (1980)
30. Konovsky, M.A.: Understanding Procedural Justice and Its Impact on Business Organizations. JOURNAL OF MANAGEMENT 26(3), 23 (2000)
31. Hutchinson, B., Mitchell, M.: 50 Years of Test (Un)fairness: Lessons for Machine Learning. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. pp. 49–58. ACM, Atlanta GA USA (2019)
32. Kaibel, C., Koch-Bayram, I., Biemann, T., Mühlenbock, M.: Applicant Perceptions of Hiring Algorithms - Uniqueness and Discrimination Experiences as Moderators. Academy of Management Proceedings 2019(1), 18172 (2019)
33. Wang, A.J.: Procedural Justice and Risk-Assessment Algorithms. SSRN Electronic Journal (2018)
34. Lee, M.K., Baykal, S.: Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. pp. 1035–1048. ACM, Portland Oregon USA (2017)
35. Helberger, N., Araujo, T., de Vreese, C.H.: Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. Computer Law & Security Review 39, 105456 (2020)
36. Marcinkowski, F., Kieslich, K., Starke, C., Lünich, M.: Implications of AI (un-)fairness in higher education admissions: the effects of perceived AI (un-)fairness on exit, voice and organizational reputation. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. pp. 122–130. ACM, Barcelona Spain (2020)
37. Lee, M.K.: Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society 5(1), 205395171875668 (2018)
38. Colquitt, J.A., Wesson, M.J., Porter, C.O., Conlon, D.E., Ng, K.Y.: Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. Journal of Applied Psychology 86(3), 425–445 (2001)

39. Lind, E.A., Tyler, T.R.: The social psychology of procedural justice. Critical issues in social justice, Plenum Press, New York (1988)
40. Zhou, L., Wang, W., Xu, J.D., Liu, T., Gu, J.: Perceived information transparency in B2C e-commerce: An empirical investigation. Information & Management 55(7), 912–927 (2018)
41. Conlon, D.E., Porter, C.O., Parks, J.M.: The Fairness of Decision Rules. Journal of Management 30(3), 329–349 (2004)
42. Martin, K.D., Borah, A., Palmatier, R.W.: Data Privacy: Effects on Customer and Firm Performance. Journal of Marketing 81(1), 36–58 (2017)
43. Friedler, S.A., Scheidegger, C., Venkatasubramanian, S.: On the (im)possibility of fairness. arXiv:1609.07236 [cs, stat] (2016), arXiv: 1609.07236