# Leveraging Actionable Explanations to Improve People's Reactions to AI-based Decisions

Markus Langer[1][0000-0002-8165-1803] and Isabel Valera [2][0000-0002-6440-4376]

[1] University of Freiburg, Department of Psychology, Freiburg im Breisgau, Germany
[2] Saarland University, Department of Computer Science, Saarbrücken, Germany

**Abstract.** This paper explores the role of explanations in mitigating negative reactions among people affected by AI-based decisions. While existing research focuses primarily on user perspectives, this study addresses the unique needs of people affected by AI-based decisions. Drawing on justice theory and the algorithmic recourse literature, we propose that actionability is a primary need of people affected by AI-based decisions. Thus, we expected that more actionable explanations – that is, explanations that guide people on how to address negative outcomes – would elicit more favorable reactions than feature relevance explanations or no explanations. In a within-participants experiment, participants ($N$ = 138) imagined being loan applicants and were informed that their loan application had been rejected by AI-based systems at five different banks. Participants received either no explanation, feature relevance explanations, or actionable explanations for this decision. Additionally, we varied the degree of actionability of the features mentioned in the explanations to explore whether features that are more actionable (i.e., reduce the amount of loan) lead to additional positive effects on people's reactions compared to less actionable features (i.e., increase your income). We found that providing any explanation led to more favorable reactions, and that actionable explanations led to more favorable reactions than feature relevance explanations. However, focusing on the supposedly more actionable feature led to comparably more negative effects possibly due to our specific context of application. We discuss the crucial role that perceived actionability may play for people affected by AI-based decisions as well as the nuanced effects that focusing on different features in explanations may have.

**Keywords:** explainability, actionability, consequential decision-making, acceptance, affected people

## 1 Introduction and Related Work

AI-based decisions affect the fate and future of individuals. This is true in high-stakes contexts such as medical diagnosis and treatment [1], hiring [2], and loan contexts [3]. Whereas users of AI-based decision-support tools (e.g., doctors, hiring managers, loan officers) can decide whether and to what extent they want to use such tools, people affected by AI-based decisions (e.g., patients, job and loan applicants) are in a very different position [4]. Although regulation such as the European GDPR in principle requires the possibility to opt out of a fully automated evaluation by AI-based tools, people affected by AI-based decisions often have less control over whether and to what extent their health, job suitability, or creditworthiness is evaluated by AI-based tools. Particularly in high-stakes situations and when decisions are highly automated (i.e., there is little human influence on the decision), research has shown that people affected by AI-based decisions tend to react negatively to the use of AI-based tools [4], [5]. This critical view of AI-based decisions by the public can undermine the potential benefits of AI-based tools for society and can also be detrimental to the image of organizations that use such tools for high-stakes decisions.

Providing explanations for AI-based decisions has been proposed as a promising way to counteract unfavorable reactions to AI-based decisions [6]. The basic idea is simple: by default, many AI-based tools do not provide any insights into their decision-making. Providing explanations could help to better understand the reasons for AI-based decisions, and thus lead to other positive effects such as a higher perceived contestability, higher perceived justice, or a general higher acceptance [7]. However, empirical evidence supporting the expected positive effects of explanations on people affected by AI-based decisions is surprisingly rare.

In fact, most research on the effects of explanations by AI-based systems has focused on users of AI-based systems [7], [8]. It is questionable whether findings from this area can be generalized to people affected by AI-based decisions simply because the position of users in the decision context differs substantially from that of people affected by AI-based systems. For example, the former can ignore AI-based outputs, whereas the latter are subject to these outputs. Additionally, the former may have more experience with and insight into these systems, whereas the latter may have little experience and no way of gaining insights into the systems decision rationale. Consequently, the needs with respect to explanations of people affected by AI-based decisions may also be substantially different [7]. The few studies that have examined the perspective of people affected by AI-based decisions have produced mixed results [4], [5]. Whereas some studies found positive effects of explanations on important outcomes such as perceived informational justice, procedural justice and perceived overall fairness [9], others found no effect [10], ambivalent effects [6], [11] or even negative effects of explanations [12]. One possible reason for this heterogeneity is that research has often compared different explainability approaches without a clear rationale for why one would be better suited to the needs of people affected by AI-based decisions than others [6].

In line with claims from research on algorithmic recourse [13], [14], we propose that one crucial need of people affected by AI-based decisions is the actionability of an explanation accompanying AI-based decisions. In other words, they want to be able to act on such explanations, they want to know ways forward, especially in cases where they experience a negative outcome in an AI-based decision situation (e.g., a loan application is rejected). For example, instead of simply telling loan applicants that their application was denied because their income was too low, it may be more actionable to tell them that they need a certain percentage of additional income. This intuition from research on algorithmic recourse is supported by justice theory [15]. In fact, particularly in the case of a negative outcome, the perceived justice of a decision becomes important in determining people's overall reactions to the decision context [16]. In such cases, explanations need to be accurate, timely, and insightful to increase people's perceived informational justice of the decision situation [15]. This line of reasoning supports the intuition that any explanation should be better than no explanation and that actionable explanations may indeed be what people affected by decisions desire in the case of negative outcomes. Further support for the role of actionability comes from research suggesting that the design of explainability approaches in a given context needs to be informed by people's goals and needs [7], [17]. In the case of a negative outcome, people affected by a decision may want to know what to do next. This requires an explanation that is actionable.

In line with these considerations, we thus propose that providing actionable explanations should lead to more favorable reactions to an AI-based decision situation than providing no explanation or than providing a feature relevance explanation. Whereas feature relevance explanations should also be beneficial in terms of the perceived actionability relative to receiving no explanation at all, they typically only focus on giving people insights into important features that were important for the AI-based system's outputs. In contrast, actionable explanations aim at telling people what to do in order to achieve a better outcome in the future [13], [14].

To date there is little empirical evidence to support the claim that actionable explanations can help to foster acceptance, let alone evidence that actionable explanations are better than other explanations at doing so. For example, Schoeffer et al. [9] introduced their participants to a third-person perspective (seeing others being affected by an AI-based decision) and found that more detailed explanations led to more favorable reactions. Additional qualitative findings showed that their participants emphasized that the actionability of explanations and the actionability of highlighted features were important for them to find explanations helpful. Binns et al. also introduced their participants to a third-person perspective and found that explanations had ambivalent effects: case-based explanations (highlighting cases similar to the affected person that may lead to insights into why a system has produced a respective output) led to comparatively more negative reactions than sensitivity-based explanations (highlighting what would have needed to be different for an output to be different; note that in other research, this kind of explanation was called a counterfactual explanation [13]). Additional qualitative insights may help to understand these findings because some participants mentioned that sensitivity-based explanations were perceived as more actionable. Singh et al. [14] found that participants in the role of a user (e.g., in the role of a loan officer) preferred more actionable explanations as the explanations that they would communicate to people affected by AI-based decisions. Additional qualitative findings indicate that explanations that focused on features that may be perceived as little actionable led to negative reactions. For example, their participants said that they found explanations that told people to increase their income by changing their job as impolite. One reason for this finding may be that this kind of explanation appears to be of limited actionability for many people who may not be able to simply change their job.

To shed light on the role of actionability of explanations on people's reactions to AI-based decisions, our study employed a within-participant design where participants received decisions with either no explanation, feature relevance explanations, or actionable explanations. Additionally, we tested whether highlighting features with different degrees of actionability (i.e., where people could have the impression that it is easier to act upon the respective feature) affect people's reactions to the AI-based decision situations differentially. For example, we expected that most people would find it easier to apply for a slightly lower loan amount than to increase their income. If this actionability is important to people affected by AI-based decisions, providing more actionable explanations and focusing on more actionable features should lead to more favorable perceptions. This leads to the following hypotheses that we test in our study:

Hypothesis 1:[1] Receiving any explanation will lead to more favorable reactions to the AI-based decision situation (i.e., perceived attractiveness of the bank as a place to apply for a home loan, fairness, informational justice, procedural justice).

Hypothesis 2: Receiving actionable explanations will lead to more favorable reactions than receiving a feature relevance explanation.

Hypothesis 3: Focusing in an actionable explanation on a more actionable feature (i.e., reduce the amount of loan) will lead to more favorable perceptions than focusing on a less actionable feature (i.e., increase your income).

## 2 Methods

### 2.1 Sample

In our preregistration, we stated that we wanted to collect data from at least $N$ = 120 participants. We decided to collect the data via the university's participant pool and via Prolific. We ended up with $N$ = 156 participants before data cleaning. In line with our preregistered exclusion criteria, we excluded ten participants because they stated that their data should not be used (e.g., due to being inattentive), four participants who failed either one of two attention checks included in the questionnaire, and one participant who took less than 3 minutes to respond, indicating inattentive responding. We also excluded three participants who indicated their age to be 15 although we informed participants in the beginning that only participants above the age of 18 are allowed to participate. On average, the study took about 10 minutes to complete ($SD$ = 3). Student participants were compensated with course credit and Prolific participants received £1.80. The final sample consisted of $N$ = 138 participants. Of those, 68% were from Prolific. There were 67% participants who indicated their gender to be female, 32% male, and 1% diverse. The mean age was 39 ($SD$ = 14). Regarding their education, the majority indicated that they had finished school (25%), had a Bachelor's degree (37%),

[1] We preregistered this study on https://aspredicted.org/5gq53.pdf. There, we included an additional hypothesis that proposed that there would be an interaction effect, i.e., a stronger difference regarding the favorability of perceptions between the feature relevance and the actionable explanation condition for the more actionable feature. However, due to an error in the study design, participants saw an old version of one of the feature relevance conditions. Specifically, one of the feature relevance conditions did not mention the amount of loan as the decisive feature but mentioned the "bank balance." This made it impossible to test this interaction hypothesis.

or a Master's degree (25%). About half (55%) of participants had already applied for a bank loan before.

### 2.2 Procedure

The experimental procedure was approved by the IRB of the first author's previous institution. The study was conducted in English and the experiment followed a within-person design with five conditions: no explanation, a feature relevance explanation focusing on the loan applicant's insufficient bank balance, a feature relevance explanation focusing on the applicant's insufficient monthly income, an actionable explanation focusing on the applicant's monthly income, and an actionable explanation focusing on the loan amount.

After being directed to the online survey, participants were welcomed and received information about data privacy and about the study. After giving their consent, participants were informed that they had to imagine that they wanted to buy a house. They had applied for a home-loan at five different banks. They were then told that in all those banks, an AI-based system decides about their home-loan. They were then informed that their application was rejected by all of the banks but that the reasons that those banks have provided differ. They were then told that they will see the rejection letters from the banks, each followed by a set of statements regarding their reactions to the respective decision that they are asked to respond to. The rejection letters included the experimental manipulations and were presented to participants in a randomized order to prevent order effects on our results.

Every rejection letter included the following information:

> "*Dear applicant, we have deployed an AI-based evaluation software that helps us process applicant documents faster and with more precision. The data you have provided us with was*
>
> - *Your age, Tenure, Income, Bank Balance*
>
> *We regret to inform you that the AI-based evaluation software has rejected your application for a home-loan -EXPERIMENTAL MANIPULATION-.*
> *We thank you for choosing us and look forward to seeing you sometime again in the future*."

The EXPERIMENTAL MANIPULATION was then filled with the following information

a) No explanation condition: No additional text
b) Feature relevance condition focusing on the bank balance: *due to insufficient bank balance.*
c) Feature relevance condition focusing on the monthly income: *due to insufficcient monthly income.*
d) Actionable explanation focusing on monthly income: *However, you would have a higher chance of approval for future loan applications if you increased your monthly income by 10%.*

e) Actionable explanation focusing on amount of loan: *However, you would have a higher chance of approval for future loan applications if you reduced the amount of the loan you are asking for by 10%.*

For the actionable explanations, we decided to use the monthly income and the amount of loan anticipating that one would be perceived as more actionable than the other as indicated by the qualitative insights from [6], [9], [14]. We then decided for a 10% increase or decrease to keep the 10% consistent.

After each scenario, participants responded to measures capturing their perceived attractiveness of the bank as a place to apply for a home loan, perceived fairness, perceived informational justice, perceived actionability, and perceived procedural justice (i.e., they responded to these items five times). After the final scenario, we asked about participants' affinity for technology interaction, for their level of education, their socioeconomic status, their demographic information, whether they had ever applied for a bank loan, if we could use their data for our analyses, and whether they have any additional remarks. In the end, participants were debriefed about the purpose of the study.

### 2.3 Measures

All items were measured on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) unless otherwise stated.

To check if our experimental manipulation increased the perceived actionability, we captured perceived actionability with three self-developed items. Those items were "I found the explanation to be actionable", "With the explanation I received, I now would know what to do differently the next time I apply for a loan", and "It would be easy for me to successfully apply for a loan in the future given the information that I have received from the bank."

Regarding measures to capture participants' reactions to the AI-based decision situation, we measured perceived attractiveness of the bank as a place to apply for a home-loan because this is a practically relevant reaction for organizations using AI-based tools to inform their decisions. Additionally, we focused on established measures related to the perceived justice [15] and overall fairness [18], [19] of the decision situation that were also used in prior research on the effects of explanations on people affected by AI-based decisions [9].

We measured perceived attractiveness of the bank as a place to apply for a home-loan with three items adapted from the organizational attractiveness measure by [20]. A sample item was "I would recommend others to apply for a home-loan at this bank."

We captured perceived fairness of the decision with two items by [18]. A sample item was "I think that the decision itself was fair."

We measured perceived informational justice with four items by [15]. A sample item was "Were explanations regarding the decision reasonable?"

We measured perceived procedural justice with four items by [15]. A sample item was "Have those procedures been free of bias?"

We captured participants' affinity for technology interactions with the four items of the scale by [21]. Here, we used the original six-point response scale. A sample item was "I like to occupy myself in greater detail with technical systems."

# 3 Results

Table 1 includes overall mean values, standard deviations, and intercorrelations between the study variables. For this table, we calculated the mean over all scenarios to get insights into how the variables are correlated to each other. This table shows the uncorrected correlations and indicates that all study variables were positively and strongly correlated to each other as can be expected with measures on perceived justice and fairness [19]. The overall mean values for the dependent variables were all below the mean of the scale (i.e., below 3), indicating that the overall reaction to the decision context was rather negative, which can be expected in the case of a negative decision outcome (i.e., a loan being rejected).

Table 1.
*Means and standard deviation between the mean of study variables over all scenarios.*

| Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| 1. Actionability | 3.09 | 0.62 | | | | | | |
| 2. Attractiveness of the bank | 2.74 | 0.57 | .61** | | | | | |
| 3. Fairness | 2.92 | 0.73 | .57** | .66** | | | | |
| 4. Informational Justice | 2.83 | 0.63 | .71** | .60** | .73** | | | |
| 5. Procedural Justice | 2.99 | 0.60 | .67** | .63** | .73** | .74** | | |
| 6. Affinity for technology interactions | 3.49 | 1.07 | .11 | .11 | .00 | -.01 | .10 | |

*Notes.* $N = 138$
$^{**}p < .01$

## 3.1 Testing of Hypotheses

Figure 1 shows the mean values of the different scenarios for the dependent variables.

For all hypotheses, we calculated several hierarchical linear models with the participant as a random factor to account for the nested nature of the data. Specifically, scenarios are nested within participants because all participants saw all scenarios.

Hypothesis 1 stated that receiving any explanation will lead to more favorable reactions (i.e., perceived attractiveness of the bank as a place to apply for a home loan, fairness, informational justice, procedural justice). To test this hypothesis, we compared the no explanation condition to all the other conditions. Receiving any explanation led to more perceived actionability indicating that our manipulation worked as intended. Additionally, for all dependent variables, we found that receiving any explanation led to more favorable reactions. This supports hypothesis 1.

Hypothesis 2 proposed that receiving actionable explanations will lead to more favorable reactions than receiving a feature relevance explanation. To test this hypothesis, we compared the feature relevance conditions to the actionable explanation conditions. The actionable explanations led to a higher perceived actionability indicating that our manipulation worked as intended. Additionally, we found that receiving an actionable explanation led to more favorable reactions for all dependent variables. This supports hypothesis 2.

Hypothesis 3 stated that focusing in an actionable explanation on a more actionable feature (i.e., reduce the amount of loan) will lead to more favorable reactions than focusing on a less actionable feature (i.e., increase income). To test this hypothesis, we compared the actionable explanation focusing on the amount of loan condition to the actionable explanation focusing on the increase of income condition. Focusing on reducing the amount of loan led to more perceived actionability indicating that our manipulation worked. However, we found no significant difference for fairness, informational justice, or for procedural justice. Focusing in the explanation on reducing the amount of loan even led to a lower perceived attractiveness of the bank.

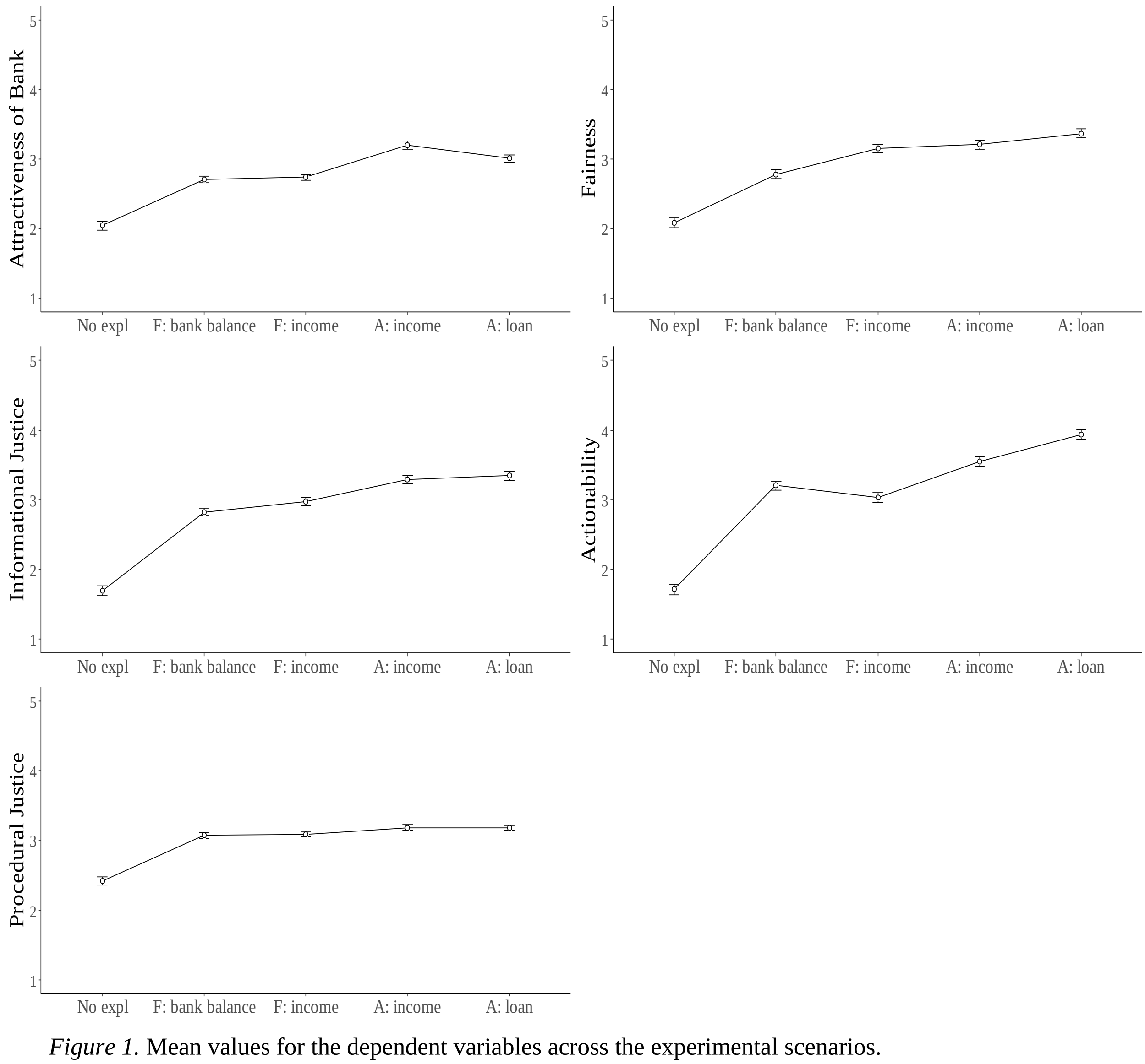


*Figure 1.* Mean values for the dependent variables across the experimental scenarios.
*Notes:* Expl. = explanation, F = feature relevance explanation, A = actionable explanation. The error bars display standard errors.

Table 2.
*Results of the hierarchical linear models to test the hypotheses*

| | Actionability | | | Attractiveness | | | Fairness | | | Informational Justice | | | Procedural Justice | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **HLM analysis for Hypothesis 1** | | | | | | | | | | | | | | | |
| Predictors | *b* | *CI* | *p* | *b* | *CI* | *p* | *b* | *CI* | *p* | *b* | *CI* | *p* | *b* | *CI* | *p* |
| Intercept | 1.71 | 1.55, 1.88 | <.01 | 2.04 | 1.91, 2.18 | <.01 | 2.08 | 1.91, 2.25 | <.01 | 1.69 | 1.54, 1.85 | <.01 | 2.42 | 2.29, 2.54 | <.01 |
| No Explanation vs. Explanation | 1.71 | 1.55, 1.88 | <.01 | 0.87 | 0.75, 0.99 | <.01 | 1.05 | 0.90, 1.20 | <.01 | 1.42 | 1.28, 1.56 | <.01 | 0.71 | 0.62, 0.81 | <.01 |
| Marginal $R^2$ | .31 | | | .15 | | | .14 | | | .28 | | | .13 | | |
| Observations | 690 | | | 690 | | | 690 | | | 690 | | | 690 | | |
| **HLM analysis for Hypothesis 2** | | | | | | | | | | | | | | | |
| Intercept | 3.12 | 2.99, 3.24 | <.01 | 2.72 | 2.61, 2.83 | <.01 | 2.97 | 2.83, 3.11 | <.01 | 2.90 | 2.77, 3.03 | <.01 | 3.08 | 2.97, 3.18 | <.01 |
| Feature vs. Actionability | 0.63 | 0.49, 0.76 | <.01 | 0.38 | 0.28, 0.48 | <.01 | 0.32 | 0.19, 0.45 | <.01 | 0.42 | 0.31, 0.53 | <.01 | 0.10 | 0.03, 0.18 | .01 |
| Marginal $R^2$ | .10 | | | .06 | | | .03 | | | .05 | | | .01 | | |
| Observations | 552 | | | 552 | | | 552 | | | 552 | | | 552 | | |
| **HLM analysis for Hypothesis 3** | | | | | | | | | | | | | | | |
| Intercept | 3.55 | 3.40, 3.70 | <.01 | 3.20 | 3.07, 3.34 | <.01 | 3.21 | 3.05, 3.37 | <.01 | 3.29 | 3.14, 3.44 | <.01 | 3.18 | 3.06, 3.30 | <.01 |
| Income vs. Loan | 0.39 | 0.22, 0.57 | <.01 | -0.20 | -0.34, -0.05 | .01 | 0.16 | -0.01, 0.32 | .06 | 0.06 | -0.08, 0.20 | .43 | -0.00 | -0.10, 0.10 | .94 |
| Marginal $R^2$ | .05 | | | .02 | | | .01 | | | .00 | | | .00 | | |
| Observations | 276 | | | 276 | | | 276 | | | 276 | | | 276 | | |

*Note.* CI = 95% Confidence Interval $N$ = 138.

# 4 Discussion

The goal of this study was to assess whether actionable explanations are better suited than less actionable ones at mitigating negative reactions in a consequential decision scenario with a negative outcome for people affected by an AI-based decision. The main findings of this study were that a) any explanation for the unfavorable outcome was perceived as more actionable and led to more favorable reactions, b) what we defined to be more actionable explanations were also perceived as more actionable and led to more favorable reactions than feature relevance explanations, and c) focusing in an actionable explanation on a presumably more actionable feature (i.e., the loan amount) was perceived as more actionable, but led to less favorable reactions in the context of a loan application scenario.

Consequently, one main takeaway of our study is that providing explanations – particularly actionable ones – can alleviate negative reactions to unfavorable decision outcomes. In cases where AI-based tools make high-stakes decisions that affect human lives, it is inevitable that some people will not receive the outcome that they had hoped for. In particular, when AI-based systems make high-stakes decisions, this can lead to unfavorable reactions of people affected by AI-based decisions [4]. To mitigate such negative reactions, research and practice see potential in providing explanations for AI-based decisions. However, research has produced mixed results in this regard [6], [10]. We found that any explanation is better than no explanation for a negative outcome. Participants expressed higher levels of perceived informational justice, procedural justice, fairness, and most importantly, had less negative reactions toward the banks that provided them with an explanation.

Consistent with the intuition of research on the importance of the actionability of explanations for people affected by AI-based decisions [13], and consistent with qualitative findings from prior research [6], [9] indicating that actionability may be what people desire when they are affected by AI-based decisions, we found empirical evidence for the positive effect of actionable explanations. In contrast to the specific feature relevance explanations that we chose, the actionable explanations suggested to participants how to improve their chances of getting a loan next time. In line with justice theory [15], this improved our participants' perceived informational and procedural justice, as well as the perceived overall fairness of the decision processes [9]. Furthermore, in line with the propositions by [7] and [17], actionable explanations seem to have been better at providing our participants with information that would help them to work toward desired positive decision outcomes. This may have contributed to the overall more positive reactions when receiving actionable explanations.

While we propose that our actionable explanation led to stronger positive effects than the feature relevance explanation because of its "actionability," there are also alternative possible explanations. For example, highlighting what needs to be done to get a better result also provides additional insight into how the model works. This means that our actionable explanation is not only more actionable, but also provides more details for people to understand the system. Thus, the current results could be driven by a "perceived understanding" effect rather than a "perceived actionability" effect. Disentangling these effects will be a task for future research. Another task for future research will be to understand the inherent subjectivity of the actionability of explanations. By definition, actionability is a subjective aspect of explanations rather than something that can be described objectively. In other words, what is actionable for one person may not be actionable for another, and what is actionable for one person at one time may not be actionable for the same person in the future. In our study, our manipulations all led to the expected effects on perceived actionability, but predicting whether an explanation will actually be *perceived* as actionable by people affected by AI-based decisions, and whether the respective explanations will have other expected downstream effects (e.g., on perceptions of justice) may be more complex in other contexts.

Contrary to our expectations, focusing on the amount of loan as a supposedly more actionable feature in an actionable explanation did not have any additional positive effects. In fact, it negatively affected the perceived attractiveness of the bank. In line with previous research [6], [9], [14], we proposed this hypothesis because we expected that participants would think that it would require less effort on their part to reduce the loan amount instead of increasing their monthly income. In line with this reasoning, we found that reducing the loan amount was perceived as more actionable. However, this did not translate into positive reactions for the other dependent variables – although we want to highlight the slightly positive, but not significant effect on perceived fairness ($b$ = 0.16, $p$ = .06). Instead, providing the actionable explanation to reduce the loan amount led to a lower perceived attractiveness of the bank. In hindsight, this result makes sense. We can see that our participants imagined that they were asking for this amount of money for a reason. A bank that says that you should ask for less money may indeed sound unattractive. However, given that focusing on the features income versus loan amount led to different perceived actionability, given that there was a slightly positive (but not significant) effect of focusing on the more actionable feature on the perceived fairness of the decision, and given that perceived actionability was significantly positively correlated with, for instance, the attractiveness of the bank, we still think that there is reason to believe that focusing on more actionable features in explanations can lead to more favorable reactions to AI-based decisions. Thus, future research could investigate the effects of focusing on other features in actionable explanations that may seem less problematic than a bank telling you to ask for less money (e.g., in the context of loans: repayment duration, interest rate).

### 4.1 Practical Implications

When using an AI-based system in a consequential context, it is worth considering to provide individuals who are confronted with a negative decision outcome with an explanation. Although overall perceptions are still likely to be rather negative due to the strong effect that negative decision outcomes have, providing explanations can at least buffer some of the negative reactions [15]. Eventually, explanations may even help to maintain some degree of a positive organizational image. It is particularly advisable to provide actionable explanations. However, it is important to consider whether providing explanations may conflict with other goals. For example, providing people with an actionable explanation could lead to making decision processes too transparent, enabling people to game the system [7]. Likewise, imagine providing an actionable explanation and having a person actually follow that explanation. If that person still does not get the desired outcome the next time they try, this can lead to particularly negative reactions.

Another implication of our study is to be aware of which feature is highlighted in an actionable explanation. In our specific case, highlighting that the applicant should ask for a lower amount of loan led to less favorable reactions – most likely because a bank that tells you to ask for less money does not sound like a bank you would ask for a loan again. However, even this explanation led to more positive reactions than not providing an explanation at all. Nevertheless, it may be advisable for system designers to enable decision recipients to provide information to the AI-based decision tool in order to personalize the explanation process [14]. For example, if loan applicants consider the loan amount to be unchangeable, it makes little sense for an explanation to suggest reducing the amount of loan requested. Instead, the explanation could focus on other features that may also provide ways forward for applicants. Alternatively, systems could also provide a variety of explanations with different ways forward [14].

### 4.2 Limitations

There are at least three limitations to our study that readers need to consider. First, this was a scenario-based study, there were no real consequences for the participants. Although more than half of our participants had experience applying for a loan and could therefore probably imagine being in the situation described, being denied a loan in reality can strongly impact people's lives.[2] Thus, the effects of explanations may be different from what we found in this study. This may be particularly true for actionable explanations. For people affected by AI-based decisions, the importance of receiving actionable explanations may only be fully realized in actual decision situations. Thus, we could hypothesize that the effects of actionable explanations are stronger for actual decision situations. Second, we focused only on the loan context. Although there are similarities in other consequential situations such as hiring (e.g., individuals apply for different positions, they receive rejections, they may receive explanations focusing on different features, these features may differ in their actionability), the current findings need to be replicated in other contexts in order to generalize our insights. Third, for the sake of simplicity, we chose to compare a 10% increase in income to a 10% decrease in the loan amount. This 10% itself could feel differently actionable for increasing income versus for decreasing the loan amount. Perhaps a 10% increase in monthly income would feel similar to a 30% decrease in loan amount. Nuances like these need to be tested in future studies.

[2] Note that we tested whether including experience in applying for loan was a significant predictor in our analyses and whether it changed our results. It was not a significant predictor in any of the regressions. The only effect that was affected by including experience with applying for loan was that receiving an actionable explanation did not lead to significantly greater perceived procedural justice compared to receiving a feature relevance explanation.

### 4.3 Conclusion

Actionable explanations can be effective in mitigating unfavorable reactions to negative decision outcomes. Particularly in the case of consequential decisions, people want to know how they may be able to achieve a better outcome in future [13]. Our study showed that providing an actionable explanation led to better reactions to the decision process and to a better image for the company using the AI-based system to make its decision. Future research could further examine the specific needs that people have in these kind of decision situations with respect to the explanations they receive. Actionability seems to be one need, but we can also see others such as a personalization of the explanation process where people could inform the system about their specific situation and about what features would be more or less actionable for them.

**Acknowledgments.** This work was partially funded by the DFG grant 389792660 as part of TRR248 Center for Perspicuous Computing, project A6, and by the project "EIS – Explainable Intelligent Systems" funded by the VolkswagenStiftung as part of the grant 98513.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

[1] S. Gaube *et al.*, "Do as AI say: susceptibility in deployment of clinical decision-aids," *npj Digit. Med.*, vol. 4, no. 1, p. 31, Dec. 2021, doi: 10.1038/s41746-021-00385-9.

[2] M. Langer, C. J. König, and M. Papathanasiou, "Highly-automated job interviews: Acceptance under the influence of stakes," *International Journal of Selection and Assessment*, vol. 27, no. 3, pp. 217–234, 2019, doi: 10.1111/ijsa.12246.

[3] G. Yalcin, S. Lim, S. Puntoni, and S. M. J. van Osselaer, "Thumbs up or down: Consumer reactions to decisions by algorithms versus humans," *Journal of Marketing Research*, vol. 59, no. 4, pp. 696–717, Aug. 2022, doi: 10.1177/00222437211070016.

[4] M. Langer and R. N. Landers, "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*, vol. 123, p. Article 106878, Oct. 2021, doi: 10.1016/j.chb.2021.106878.

[5] C. Starke, J. Baleis, B. Keller, and F. Marcinkowski, "Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature," *Big Data & Society*, vol. 9, no. 2, p. 205395172211151, Jul. 2022, doi: 10.1177/20539517221115189.

[6] R. Binns, M. Van Kleek, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt, "'It's reducing a human being to a percentage'; Perceptions of justice in algorithmic decisions," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–14. doi: 10.1145/3173574.3173951.

[7] M. Langer *et al.*, "What do we want from Explainable artificial intelligence (XAI)? A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research.," *Artificial Intelligence*, vol. 296, p. 103473, 2021, doi: 10.1016/j.artint.2021.103473.

[8] V. Lai, C. Chen, A. Smith-Renner, Q. V. Liao, and C. Tan, “Towards a science of human-AI decision making: An overview of design space in empirical human-subject studies,” in *2023 ACM Conference on Fairness, Accountability, and Transparency*, Chicago IL USA: ACM, Jun. 2023, pp. 1369–1385. doi: 10.1145/3593013.3594087.

[9] J. Schoeffer, N. Kuehl, and Y. Machowski, “‘There is not enough information’: On the effects of explanations on perceptions of informational fairness and trustworthiness in automated decision-making,” in *2022 ACM Conference on Fairness, Accountability, and Transparency*, Jun. 2022, pp. 1616–1628. doi: 10.1145/3531146.3533218.

[10] N. Schlicker, M. Langer, S. K. Ötting, C. J. König, K. Baum, and D. Wallach, “What to expect from opening ‘Black Boxes’? Comparing perceptions of justice between human and automated agents,” *Computers in Human Behavior. Advance online publication.*, 2021, doi: 10.1016/j.chb.2021.106837.

[11] M. K. Lee, A. Jain, H. J. Cha, S. Ojha, and D. Kusbit, “Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–26, Nov. 2019, doi: 10.1145/3359284.

[12] M. Langer, C. J. König, and A. Fitili, “Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection,” *Computers in Human Behavior*, vol. 81, pp. 19–30, 2018, doi: 10.1016/j.chb.2017.11.036.

[13] A.-H. Karimi, B. Schölkopf, and I. Valera, “Algorithmic recourse: From counterfactual explanations to interventions,” in *Proceedings of the 2021 FAccT Conference on Fairness, Accountability, and Transparency*, Virtual Event Canada: ACM, Mar. 2021, pp. 353–362. doi: 10.1145/3442188.3445899.

[14] R. Singh *et al.*, “Directive explanations for actionable explainability in machine learning applications,” *ACM Trans. Interact. Intell. Syst.*, vol. 13, no. 4, pp. 1–26, Dec. 2023, doi: 10.1145/3579363.

[15] J. A. Colquitt, “On the dimensionality of organizational justice: A construct validation of a measure,” *Journal of Applied Psychology*, vol. 86, no. 3, pp. 386–400, 2001, doi: 10.1037/0021-9010.86.3.386.

[16] E. A. Lind and K. van den Bos, “When fairness works: Toward a general theory of uncertainty management,” *Research in Organizational Behavior*, vol. 24, pp. 181–223, Jan. 2002, doi: 10.1016/S0191-3085(02)24006-X.

[17] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, “Metrics for explainable AI: Challenges and prospects,” *arXiv*, 2018, doi: 10.48550/arXiv.1812.04608.

[18] T. N. Bauer, D. M. Truxillo, R. J. Sanchez, J. M. Craig, P. Ferrara, and M. A. Campion, “Applicant reactions to selection: Development of the Selection Procedural Justice Scale (SPJS),” *Personnel Psychology*, vol. 54, pp. 387–419, 2001, doi: 10.1111/j.1744-6570.2001.tb00097.x.

[19] J. A. Colquitt and J. B. Rodell, “Measuring justice and fairness 8,” *The Oxford handbook of justice in the workplace*, vol. 1, p. 187, 2015.

[20] S. Highhouse, F. Lievens, and E. F. Sinar, “Measuring attraction to organizations,” *Educational and Psychological Measurement*, vol. 63, pp. 986–1001, Dec. 2003, doi: 10.1177/0013164403258403.

[21] T. Franke, C. Attig, and D. Wessel, "A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale," *International Journal of Human–Computer Interaction*, vol. 35, no. 6, pp. 456–467, Apr. 2019, doi: 10.1080/10447318.2018.1456150.