Choosing Between Human and Algorithmic Advisors: The Role of Responsibility Sharing

Lior Gazit (lgazit01@campus.haifa.ac.il)
University of Haifa

Ofer Arazy
University of Haifa

Uri Hertz
University of Haifa

Article

Keywords:

Posted Date: December 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2324206/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Algorithms are increasingly employed to provide accurate advice across domains, yet in many cases people tend to prefer human advisors, a phenomenon termed algorithm aversion. To date, studies have focused mainly on the effects of advisor's perceived competence, ability to give accurate advice, on people's willingness to accept advice from human and algorithmic advisors and to arbitrate between them. Building on studies showing differences in responsibility attribution between human and algorithmic advisors, we hypothesize that the ability to psychologically offload responsibility for the decision's potential consequences on the advisor is an important factor affecting advice takers' choice between human and algorithmic advisors. In an experiment in medical and financial domains (N = 806), participants were asked to rate advisors' perceived responsibility and choose between a human and algorithmic advisor. Our results show that human advisors were perceived as more responsible than algorithmic advisors and that the perception of the advisor's responsibility affected the advice takers' choice of advisor. Furthermore, we found that an experimental manipulation that impeded advice takers' ability to offload responsibility affected the extent to which human, but not algorithmic, advisors were perceived as responsible. Together, our findings highlight the role of responsibility sharing in shaping algorithm aversion.

Introduction

Humans often rely on advice from others to inform their decisions\(^1\)\(^,\)\(^2\). Advances in information technology have given rise to a new source of advice: algorithmic recommendations. Algorithmic advisors have become prevalent in everyday life\(^3\): Netflix uses data on viewers' consumption habits to recommend movies and TV; OkCupid provides recommendations for potential spouses; Wealthfront uses automation to manage customers' investment portfolios; Medal diagnoses and treats patients by relying on algorithms and medical calculators; and numerous other examples prevail. Recently, Amazon's chief scientist, Rohit Prasad, stated that SIRI, Amazon's AI assistant, "keep[s] growing from more of an assistant to an advisor"\(^4\). Notwithstanding these advancements, studies show that in many cases people are more likely to follow human advisors' advice than algorithmic advice\(^5\)\(^–\)\(^8\). Here, we conjecture that one reason for this discrepancy is that human advisors provide a benefit for advice takers that goes beyond the content of their advice, by offering the opportunity to share responsibility for the decision's potential outcome. In this study, we examine whether responsibility sharing is also applicable to algorithmic advisors, and investigate the extent to which this factor affects advice takers' choice of human/algorithmic advisors.

The most obvious and most studied factor affecting the choice of advisor, as well as compliance with advice, is the perceived competencies of the advisor. When choosing between human advisors, people tend to choose the advisor they perceive to be the most capable and therefore most likely to help them reach a good decision\(^9\)\(^–\)\(^11\). Accordingly, people tend to comply with algorithmic advice when the algorithm is perceived as proficient. Prior studies have shown that people tend to follow advisors they
perceive as trustworthy, where trustworthiness is a composite construct that includes three dimensions: competence, benevolence, and integrity\textsuperscript{12,13}. Therefore, the likelihood that advice is accurate is associated with advisors’ trustworthiness, and particularly their competence\textsuperscript{14}. For example, people tend to trust and rely more on algorithmic advisors when facing objective numerical tasks, such as estimating the weight of a person\textsuperscript{15}. People show less likelihood of following algorithmic advice after seeing the algorithm make a mistake, a phenomenon referred to as “algorithm aversion”, as even one mistake indicates that the algorithm is imperfect\textsuperscript{5}. These works indicate that the advisor’s competence and expertise contribute to advice takers’ perception of advice accuracy, affecting their choice of advisors and their willingness to embrace the advice.

However, beyond seeking to maximize the accuracy of advice, advice takers are driven by other goals as well, namely their desire to offload the responsibility for the decision’s potential consequences by psychologically sharing it with the advisor\textsuperscript{11}. When faced with decisions that involve risk, people displayed sensitivity to anticipated regret, avoiding information on decisions with a potentially negative outcome\textsuperscript{16}, and changing their risk preferences when counterfactual information on the outcome of an unchosen option is available\textsuperscript{17}. Accordingly, when reaching decisions that involve risky outcomes, embracing advice allows the advice taker to share responsibility with the advisor for the consequences of errors, and therefore to alleviate some of the anticipated regret\textsuperscript{18}. For example, managers facing risky decisions, such as allocating funds for speculative investments, were shown to be more willing to use an advisor who can help them share some of the responsibility\textsuperscript{19}. In addition, when the cost of making a wrong decision increase, people are more likely to comply with advice. For example, when asked to choose between different ways of address a disease outbreak in cattle, people were more likely to follow advice when the disease was more severe\textsuperscript{2}. These works indicate that advice seekers may be guided by more than simply information capable of leading to a desirable outcome.

Responsibility sharing motivation is linked to the notion of responsibility attribution. According to attribution theory\textsuperscript{20}, following an action or decision outcome, people form a causal relationship between the action and the outcome and attribute responsibility to the agent making the recommendation. Studies have shown that once the decision outcome becomes known, people tend to seek justifications for bad decisions or congratulate themselves for good ones\textsuperscript{21,22}. When following an advisor’s advice, some of the responsibility can be attributed to the advisor, offloading responsibility from the decision maker. Several studies have investigated whether responsibility is equally attributed to human agents and algorithmic agents once the recommendation’s outcome becomes known, for example money loss after an investment recommendation or a deterioration in health after medical advice\textsuperscript{6,23}. Responsibility attribution was then compared between conditions in which the agent providing the information is human or an algorithm. A number of studies found that people tend to attribute less responsibility to an algorithmic advisor compared to a human advisor\textsuperscript{24–26} and feel less responsible after following human advice than when following an algorithm’s advice. However, other works found a more complex relationship regarding responsibility attribution in the case of algorithmic advisors, pointing to the
valence of the outcome and the nature of relationship between the advisor and the advice taker as key factors. It was further suggested that attribution may serve as a defense mechanism, where one attributes responsibility to a source that is very different from her/himself. This is particularly applicable to identity-threatening decisions, whose outcome may reflect badly on the decision maker. These findings indicate a difference in the way people attribute responsibility for an outcome when relying on the advice of algorithmic or human agents. It remains to be clarified whether such responsibility attribution considerations affect the choice of advisor even prior to reaching a decision and discovery of the outcome.

Here we propose that, similar to the findings concerning human advisors, responsibility sharing is an important factor affecting the decision of whether to prefer a human or algorithmic advisor, in addition to the effect of other well-studied factors of the advisor's perceived competence. Following the works on responsibility attribution to algorithmic advisors, we hypothesize that human advisors are perceived as more responsible than algorithmic advisors and therefore provide more opportunity for responsibility sharing. All else being equal, this difference in perceived responsibility is expected, therefore, to affect advice takers’ choice of advisor, making them more likely to choose a human advisor. We further conjecture that advice takers’ perception of a human advisor’s responsibility will be more sensitive to responsibility sharing manipulations (for example, signing a responsibility waver) when compared to the effect of similar manipulations on the perceived responsibility of algorithmic advisors.

To test these hypotheses, we designed an experimental task in which participants were required to make a decision and were asked to choose between two advisors – a human advisor and an algorithmic advisor. Prior to making the decision, participants rated the advisors in terms of their trustworthiness and responsibility. Our first set of hypotheses concerned the advisor's perceived responsibility. We hypothesized that human advisors would be perceived as more responsible (H1a). In order to test the effect of responsibility sharing, in one experimental condition advice takers were presented with a disclaimer stating that the advisor is not responsible for the decision's possible outcome. We hypothesized that the disclaimer manipulation would affect the advisors' perceived responsibility (H1b), and that the disclaimer's effect would be more pronounced for human advisors than for algorithmic advisors (H1c). Our second set of hypotheses concerned the association between the advisors’ perceived responsibility and the likelihood of their choice as a source of advice. We hypothesized that the more advice takers perceive advisors as responsible, the more they would be likely to opt for them (H2a). Furthermore, we conjectured that the experimental manipulation limiting responsibility sharing would reduce advice takers’ likelihood of choosing a human advisor (H2b).

**Methods**

**Participants**

Participants were recruited on the Prolific online experiment platform. We recruited participants who were living in the UK and who had an approval rate of at least 91. Each participant received 1GBP for
participating in the experiment, and the experimental task required approximately seven minutes to complete. All participants provided informed consent and received monetary compensation.

We set the sample size at 400 per experiment (200 participants per condition), in order to detect an effect of 10% of our manipulation with power of 0.8\textsuperscript{31}. For the Medical domain experiment (see below), we collected data from 404 participants. The questionnaire contained several questions examining participants’ concentration (e.g., “please select the 'Neutral' option for this item”). One participant failed to provide the correct answers to these questions and was excluded from the dataset, leaving us with a sample of 403 participants (163 male, aged 39±14; 235 female, aged 39±13; 5 others, aged 35±15). Of these, 203 participants were assigned to the "Neutral disclaimer" condition and the remaining 200 participants were assigned to the "No Responsibility disclaimer" condition (participants were randomly assigned to the conditions; see details below). For the Financial domain experiment, we collected data from 402 participants and 4 participants were excluded, leaving a sample of 398 participants (172 male, aged 38±13; 222 female, aged 40 ± 14; 4 participants identified their gender as “other,” aged 29±5). Of these, 201 participants were assigned to the "Neutral disclaimer" condition and the remaining 197 participants were assigned to the "No Responsibility disclaimer" condition (participants were randomly assigned to the conditions). Summary statistics for the different demographic variables are provided in Supplementary Table S1.

Procedure

Our experiment investigated advice taking in two domains: Financial and Medical. The choice of these domains was based on evidence whereby people prefer human advisors over algorithmic advisors in both the Financial domain\textsuperscript{7,8} and the Medical domain\textsuperscript{6,8}. The experiments in both domains followed a similar design, which included: a description of the domain and decision problem, presentation of a disclaimer based on the experimental condition, rating of advisors, and choosing a preferred advisor (see Figure 1). In addition, we collected demographic information (age, education level, gender). Participants also completed a Ten Item Personality Inventory (TIPI) questionnaire to measure the Big-Five personality dimensions of Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience\textsuperscript{32}.

In the Medical domain experiment, each participant was asked to imagine that s/he is a member of a committee for clinical trials, and that s/he is required to make decisions regarding five different oncological patients who are candidates for participating in a new clinical trial. The clinical trial could either improve the patient’s medical condition or alternately cause a deterioration in the patient’s condition. For each patient, the only information that was presented was age, marital status, weight, and place of residence. In the Financial domain experiment, participants were asked to imagine that they are members of a Venture Capital’s Board of Directors and each was required to make an investment decision regarding five different startup companies. For each startup, the only information provided was the startup’s name, country, domain, and founding year.
Next, participants were asked to agree to a disclaimer. They were randomly assigned to two groups. Group 1 (“Neutral”) received the disclaimer: “I am fully aware of the fact that I can select any advisor that I desire,” whereas Group 2 (“No Responsibility”) received the disclaimer: “I am fully aware of the fact that the advisor I will receive has no responsibility for any consequence resulting from my decision.”

Then, participants were offered advice by one of two expert advisors: a human or an algorithm. The human advisor was described as an expert oncologist (Medical domain experiment) or an expert investment analyst (Financial domain experiment). The algorithm was described as an algorithmic advisor expertly trained on oncological data (Medical) or an algorithmic advisor expertly trained on investments data (Financial). Next, participants were asked to rate both advisors regarding the advisor’s perceived trustworthiness in terms of competence, benevolence, and integrity\(^\text{13}\), using a 1-5 ordinal scale. They were also asked to rate the advisor’s responsibility using a similar 1-5 scale. After rating both advisors, participants were asked to make their choice regarding the preferred advisor. Once a choice had been made, the chosen source of advice – human or algorithm – was named as the advisor and the disclaimer (see above) appeared at the bottom of the selection page.

Analysis

To analyze the advisors’ perceived responsibility, we used mixed-effects linear regressions, using group level coefficients (fixed effects) to model population level effects and individual level coefficients (random effects) to capture average individual responses\(^\text{33}\). To analyze the choice of advisor, we used a logistic regression with group level coefficients only, as each participant made a single choice. We included the average trustworthiness ratings from the three scales as a control variable, as well as personality traits from the TIPI questionnaire and demographic variables, as reported in the Supplementary Materials. In the tables below, we report the standardized coefficients, \(t\) values, and \(p\) values for the relevant variables (full details are provided in the Supplementary Materials). Note that standardized coefficients represent the partial correlation between the dependent and independent variables and are therefore indicators of effect size\(^\text{34}\). We estimated the population level marginalized means for post-hoc evaluation of the effects and for post-hoc comparisons\(^\text{35}\). All analyses were conducted using R software (version 4.03; analysis packages are detailed in the Supplementary Materials).

Ethics

This study was approved by the University of Haifa's Department of Psychology Ethics Committee (No 058/20). All methods were performed in accordance with the relevant guidelines and regulations, and approved by the ethics committee.

Data Availability

All behavioral data and analysis scripts are available at: https://osf.io/ha4um/?view_only=54e9003e0690429f82f1c708545dc1c5
Results

Advisors’ Responsibility Ratings

To test hypotheses **H1A** (human advisors will be perceived as more responsible than algorithmic advisors) and **H1B** (impeding the ability to share responsibility with the advisor will reduce the advisor’s perceived responsibility), a linear mixed-effect regression was conducted, with the advisor’s perceived responsibility as the dependent variable and advisor identity (Human/Algorithm) and disclaimer (Neutral/No Responsibility) as fixed effects. The regression included individual level intercept random effects and control variables (age, gender, and TIPI). The regression analyses were conducted separately for the Medical domain and Financial domain experiments.

In the Medical domain experiment (see Figure 2), we observed a significant difference in advisors’ perceived responsibility when relating to an algorithm advisor (M=2.41, SD=1.03) or a human advisor (M=3.91, SD=0.96) (standard coefficient [95% CI]: 0.91 [0.8 1.02], t(402) = 16.53, p<0.001. See full details in Supplementary Materials, Table S4). A significant effect of the manipulation (i.e., disclaimer) on the advisor’s perceived responsibility was observed, with higher ratings in the Neutral disclaimer condition (M=3.16, SD=1.24) than in the No-Responsibility disclaimer condition (M=2.88, SD=1.23) (standard coefficient [95% CI]: -0.24 [-0.37 -0.1], t(391) = -3.41, p<0.001).

A similar pattern was observed in the Financial domain experiment. A significant difference was found in advisors’ perceived responsibility when relating to an algorithmic advisor (M=2.62, SD=1.15) or a human advisor (M=3.47, SD=1.09) (standard coefficient [95% CI]: 0.58 [0.47 0.68], t(397) = 10.81, p<0.001). Also, the manipulation had a significant effect, with higher perceived responsibility in the Neutral disclaimer condition (M=3.04, SD=1.20) than in the No-Responsibility disclaimer condition (M=2.5, SD=1.22), t(386) = -4.83, p<0.001. See details in Supplementary Materials, Table S5).

To test **H1C** - human advisors’ perceived responsibility is more sensitive to the manipulation that involved impeding the ability to offload responsibility than the algorithmic advisor’s perceived responsibility - we conducted a separate linear mixed-effects regression. This regression analysis was similar to the regressions described above but included an additional interaction effect between the source of advice (human/algorithm) and the manipulation (Neutral/No-Responsibility disclaimer). Again, analyses were performed separately for the Medical and Financial domain experiments.

In the Medical domain experiment (see Figure 2) a significant interaction effect was found (t(401) = -5.48, p<0.01; see full details in Supplementary Materials, Table S6). For human advisors, perceived responsibility was higher in the Neutral condition (M=3.91, SD=0.96) than in the No-Responsibility condition (M=3.27, SD=1.21). For algorithmic advisors, however, the perceived responsibility did not differ much between the Neutral condition (M=2.41, SD=1.03) and the No-Responsibility condition (M=2.50, SD=1.14). A similar pattern was observed in the Financial domain experiment where a significant interaction effect was found (t(396) = -2.12, p=0.03); see full details in Supplementary Material, Table S7). For human advisors, the perceived responsibility was higher in the Neutral condition (M=3.47,
SD=1.09) than in the No-Responsibility condition (M=0.57, SD=1.27), whereas for algorithmic advisors the perceived responsibility did not differ much between the Neutral (M=2.61, SD=1.15) and the No-Responsibility conditions (M=2.29, SD=1.14). See Figure 2 below.

As an additional test, we explored whether the manipulation (i.e., disclaimer) also affected the advisor’s perceived trustworthiness, performing similar analyses with trustworthiness as the dependent variable. We observed no significant effects of the manipulation on advisors’ trustworthiness scores. However, we noted significant differences in trustworthiness scores between human and algorithmic advisors, in both the Medical and the Financial domains. See details and results in the Supplementary Materials, Tables S8, S9, S10, S11, and Figure S11. We, therefore, included perceived trustworthiness as a control variable in our subsequent analyses regarding participants’ choice of advisors (details below).

**Choice of Advisors**

We hypothesized that the advisors’ perceived responsibility would affect the choice of advisor (H2A). To test this hypothesis, we conducted a multiple logistic regression, which included choice of advisor (human/algorithmic) as dependent variable and perceived responsibility of both algorithmic and human advisors as independent variables. As control variables, we included the advisors’ perceived trustworthiness (for both algorithmic and human advisors) and participants’ demographics. The analyses were performed separately for the Medical and Financial experiments. In the Medical domain, we found that most participants preferred a human advisor (82.37%; \( \chi^2 = 169.03, p<0.001 \)). The choice of advisor was significantly influenced by the human advisor’s perceived responsibility (standard coefficient [95% CI]: 0.38 [0.07 0.70], \( z(388)=2.4, p=0.02 \)) but not by the algorithmic advisor’s perceived responsibility (standard coefficient [95% CI]: -0.08 [-0.40 0.24], \( z(388)=-0.5, p=0.61 \); see details in Supplementary Materials, Table S8). Similarly, in the Financial domain most participants preferred a human advisor, but the effect was less pronounced (66.82%; \( \chi^2 = 45.12, p<0.001 \); see Figure 3B). Participants’ choice was significantly affected by the perceived responsibility of both the human advisor (standard coefficient [95% CI]: 0.49 [0.15 0.83], \( z(383)=2.83, p<0.01 \)) and the algorithmic advisor (standard coefficient [95% CI]: -0.40 [-0.74 -0.07], \( z(383)=-2.36, p=0.02 \); see details in Supplementary Materials, Table S9). In both experimental domains, the choice of advisor – human or algorithm – was significantly affected by the control variable of the advisor’s perceived trustworthiness, in line with prior studies in this field.

Finally, we examined the hypothesis that impeding advice takers’ ability to offload responsibility would shift their choice towards the algorithmic advisor (H2B). We conducted a multiple logistic regression, with choice of advisor (human/algorithim) as dependent variable and manipulation (Neutral/No-Responsibility disclaimer) as independent variable. We also included the control variables of perceived trustworthiness, perceived responsibility, and participants’ demographics. In both the Medical and Financial domain experiments, the effect of the manipulation was statistically insignificant (Medical: standard coefficient [95% CI]: 0.17 [-0.44 0.79], \( z(397)=.17, p=0.59 \); Financial: standard coefficient [95% CI]: <0.01 [-0.59 0.58], \( z(382)=-0.01, p=0.98 \); see details in Supplementary Materials, Tables S10 and S11 for the Medical and Financial domains, respectively). Interestingly, although the disclaimer manipulation significantly
affected the advisor’s perceived responsibility and perceived responsibility significantly affected the choice of advisor, the direct effect of disclaimer manipulation on the choice of advisor was very weak and statistically insignificant in both domains (in the Medical domain, 83.25% and 81.5% preferred a human advisor in the Neutral and No-Responsibility disclaimers, respectively; in the Financial domain, 68.16% and 65.48% preferred a human advisor in the Neutral and No-Responsibility disclaimers, respectively). See Figure 3 below.

Discussion

We set out to examine the factors affecting advice takers’ choice between human and algorithmic advisors. Whereas most previous studies focused on factors related to the advisor’s competencies (e.g., trustworthiness) and advice content (e.g., accuracy), here we focused on an alternate factor that may affect advice takers’ choice: responsibility sharing, i.e., offloading to the advisor some of the responsibility associated with the decision’s potentially negative consequences. Participants were required to make a decision in the Medical and Financial domains, and were provided with a choice between a human and an algorithmic advisor. In addition, some participants received a No-Responsibility disclaimer, impeding their ability to share responsibility with their chosen advisor. We found evidence that responsibility sharing is applicable to human advisors more than to algorithmic advisors, and that human advisors’ perceived responsibility plays an important role in the choice between human and algorithmic advisors. Specifically, our findings show that human advisors’ perceived responsibility was higher than that of algorithmic advisors. Moreover, the perceived responsibility of human advisors was more strongly influenced by the No-Responsibility disclaimer. Finally, the choice of advisors was affected by human advisors’ perceived responsibility. Algorithmic advisors’ perceived responsibility too affected the choice of advisor, but this effect was relatively small. Generally speaking, the effects mentioned above were consistent for both the Medical and Financial domains, indicating the domain-general nature of advice takers’ inclination to offload responsibility by turning to a human advisor.

Theoretical Contribution

The primary theoretical contribution of this study is in demonstrating that responsibility attribution considerations affect advice takers’ choice of human/algorithmic advisors even prior to reaching a decision and before the outcome is known. Previous studies suggest that decision makers attribute responsibility once an outcome is revealed and the cause of the outcome is sought\[^{20-22}\]. As such, responsibility attribution seeks to form a direct causal link between the outcome and its source, and the literature distinguishes between attribution to the action and attribution to the agent\[^{36}\]. Decision makers’ responsibility attribution is sensitive to factors such as controllability, stability, and autonomy\[^{37}\]. These factors are linked to a broader responsibility perception, which is not directly dependent on a past outcome. Hence, when an advice taker selects a source of advice, broader perceptions regarding the advice giver’s responsibility may affect the former’s considerations regarding the possibility of attributing responsibility to the advice giver (in the case of an undesirable outcome). When comparing advice takers’
willingness to accept advice from either a human or an algorithmic advisor, previous studies have shown that responsibility attribution is more applicable to human advice givers once the recommendation's outcome becomes known\(^6,23,25,26\). Our study reveals that beyond the attribution associated with the action and its outcome, broader responsibility considerations associated with the advice giver, namely the possibility of offloading responsibility to the advisor in the case of a negative outcome, may affect an advice taker's choice between a human and an algorithmic advisor. Next, we discuss our study's specific findings and how they inform extant knowledge in the field.

Our finding that less responsibility is attributed to algorithmic advisors than to human advisors is in line with previous studies. For example, studies have shown that algorithms are perceived as less autonomous, having less control over their actions, and therefore there is a smaller likelihood that responsibility for wrongdoing will be attributed to them\(^ {26,38} \). Our finding further extends these results by showing that algorithmic advisors' perceived responsibility is less sensitive to explicit manipulations involving the future possibility of attributing responsibility to the advisor (i.e., the No-Responsibility disclaimer) than the effect on human advisors' perceived responsibility.

Interestingly, previous studies provide evidence for cases in which responsibility for an outcome was attributed to algorithmic agents, sometimes even more than the responsibility attributed to human agents. For example, one study found that responsibility for accidents was equally attributed to human and algorithmic drivers, and that algorithmic drivers were praised even more than humans for rescuing their passenger\(^ {36} \). Another study found that advice takers feel less regret when making bad decisions regarding stock investments when following the advice of robo-advisors\(^ {23} \). In different contexts, participants were found to blame algorithm mediators and advisors for their own immoral behavior\(^ {39–41} \). Such contradictory findings suggest that in some cases the need to attribute responsibility to a source that is very different from the human advice taker (i.e., an algorithmic advisor), a process known as defensive attribution\(^ {28,29} \), is so strong that it overshadows the effects of the stability, autonomy, and controllability considerations discussed above, leading people to blame algorithmic advisors\(^ {28,29} \). Whereas responsibility attribution that is linked to a decision's outcome is susceptible to defensive attribution, broader agent-based perceived responsibility may be less sensitive to such defensive considerations, as it is independent of a specific outcome. An interesting direction for future studies regarding the choice of advisors is to explore the extent to which the likelihood of a negative outcome shapes the salience of responsibility sharing considerations. We conjecture that when negative outcomes are more likely, responsibility sharing will be more pronounced, whereas positive outcomes may reduce the need for responsibility sharing.

Our results showed that advice takers' perceived responsibility regarding a human advisor affect the likelihood of preferring a human advisor over an algorithmic one. This finding highlights the importance of factors that are not directly linked to promoting the decision's outcome in shaping advisor preferences. Namely, responsibility will not change the specific outcome of the decision (e.g., monetary loss or gain, health deterioration or improvement), yet it affects the choice of advisor. One line of explanation is that
responsibility sharing can alleviate psychological consequences, such as regret or reduced self-evaluation\textsuperscript{11,28,42}. This finding informs the discussion regarding the broader contextual conditions in which algorithmic advice is provided and the varying psychological and physical needs of potential advice takers in these different contexts\textsuperscript{30}. Morewedge (2022) argues that, similar to the effect of defensive attribution, in decisions that are identity-threatening (i.e. the outcome may reflect badly on one's character), advice takers are less likely to prefer algorithmic advice. The level of identity threat may be affected by the decision domain, for example decisions related to employee hiring or to character judgements seem to reflect highly on the decision maker's abilities and therefore lead to a preference for a human advisor\textsuperscript{43}. In contrast, when decisions concern the estimation of an objective quantity or a demanding calculation, where making a mistake poses less of a threat to the decision maker's identity, people tend to prefer algorithmic advisors\textsuperscript{15,44,45}. Identity threat may also vary by the individual's traits and perceptions. For example, people's sense of power with regard to the algorithmic advisor, which modulates their sense of threat, shapes their need to share responsibility\textsuperscript{27}. Hence, individuals with an elevated sense of self-responsibility are more likely to prefer algorithmic advice\textsuperscript{46}. Our findings provide support for the role of perceived responsibility in shaping the direct choice between human and algorithmic advisors. Future works may further examine the extent to which reducing the threat to the decision maker's identity affects the choice between algorithmic and human advisors.

**Applied Contribution**

In addition to its theoretical contribution, our research has practical applications as well. As algorithmic advisors become increasingly common, a growing body of evidence is showing that they can outperform human advisors in various domains. For example, algorithmic advisors were found to provide more accurate estimations regarding which employee will succeed in his job\textsuperscript{47}, which academic candidate will succeed in her studies\textsuperscript{48,49}, or which stock will perform better\textsuperscript{50}. Nonetheless, in many situations people still tend to prefer the advice of a human advisor, which means that in many cases people receive less accurate advice. Better understanding of the reasons for preferring human over algorithmic advisors can help find ways of promoting algorithmic advisors for tasks where algorithms are likely to outperform humans, thus contributing to general well-being.

Previous research offered insights regarding people's perceptions of algorithmic advisors' abilities, as well as regarding the conditions in which people tend to trust algorithmic advisors to provide them with beneficial advice\textsuperscript{5,15,51,52}. Our results indicate that other parameters, namely perceived responsibility, should also be considered. For example, in situations where there is an interest in promoting the adoption of algorithmic advice, foregrounding the name of the algorithm's developer (an individual or a company) may prove an effective tool for applying responsibility sharing to the algorithmic advisor, thus increasing the willingness to embrace the algorithm's advice.

Notwithstanding the earlier recommendations, there may be other situations where algorithms do not have a clear advantage over human advisors, where promoting human advisors may have societal
benefits. Designers of algorithmic and human recommendation systems should take responsibility sharing perceptions into account, and perhaps include a human-in-the-loop as a means of mitigating algorithm aversion. Studies predict that about 47% of jobs in the United States are at high risk of becoming automated\textsuperscript{53}, while others predict that about 14% of jobs in OECD countries participating in the PIAAC have the same risk\textsuperscript{54}. Either way, work environments are gradually becoming computerized, and more and more professions are at risk of being replaced by artificial intelligence. As people seek ways of remaining relevant in the workforce, the notion of responsibility sharing may become powerful for convincing the public to rely on human advice. More broadly, highlighting factors that satisfy psychological needs, such as responsibility sharing, reciprocity, and long-term relationships, may prove effective in persuading people to prefer interactions with humans.

**Limitations**

Our study has several limitations that should be noted. First, participants were presented with a hypothetical scenario (being part of a Board of Directors facing an investment decision or part of a committee for clinical trials) that was not directly linked to their everyday life. Whereas this practice is quite common in studies on advice taking and has yielded consistent findings\textsuperscript{2,19,24}, it has limited ecological validity and may not fully capture relevant cognitive processes (e.g., participants did not pay a price for the consequences of their decision). In all likelihood, if advice takers needed to face similar situations in real life, the potential risks associated with a decision would have increased the need to offload responsibility, and thus the results of this study represent the lower bound of the true effect of responsibility sharing.

It is also possible that our experimental design indirectly affected participants’ choices. Participants were first asked to provide their ratings of the (algorithmic and human) advisors’ trustworthiness and responsibility, and only then did they select their advisor of choice. It is possible that the rating questions prompted participants to consider responsibility, and thus affected the decision regarding the preferred advisor. It should be noted this experimental design allowed us to include in our statistical analysis the constructs of the advisor's perceived trustworthiness and responsibility as antecedents of the preferred advisor. Given that our results regarding perceived responsibility were consistent across domains, since we found that only human advisors’ responsibility ratings were affected by the disclaimer manipulation, and given that our findings are largely in line with the results of previous studies on responsibility attribution, we believe that the effect of our experimental design (i.e., asking participants to first rate the advisors and only then make the choice) was negligible. We propose that future studies explore the effects of responsibility sharing on the choice of advisors through alternative experimental designs.

**Conclusions**

To conclude, our study examined whether algorithm aversion is affected by properties that are not directly linked to a possible improvement in the decision outcome. Conceptual works in this field have called for
empirical studies that would further increase the understanding regarding the psychological processes that underlie decision makers’ algorithm aversion. Our study addresses these calls and our results, showing that advisors’ perceived responsibility is an important factor affecting the decision between human and algorithmic advisors, point to an important direction in understanding the sources of algorithm aversion. Namely, our work provides evidence for advice takers’ consideration regarding the need to offload responsibility for a potentially negative outcome. We, therefore, maintain that making algorithms more capable and accurate is not sufficient in order to alleviate algorithm aversion. Finally, we suggest that designers of recommendation systems who seek to promote the use of algorithmic advisors should take into account advice takers’ psychological goals or needs, beyond the goal of receiving accurate advice. We hope that this initial exploration of advice takers’ desire to offload responsibility will encourage future research into the factors underlying the preference for human/algorithmic advisors.

References

25. Awad, E. et al. Drivers are blamed more than their automated cars when both make mistakes. Nat Hum Behav 4, 134–143 (2020).


42. el Zein, M. & Bahrani, B. Joining a group diverts regret and responsibility away from the individual. Proceedings of the Royal Society B: Biological Sciences 287, (2020).


Figures

1. Background Story & Decision Problem
2. Signing a Disclaimer: Neutral \ No responsibility
3. Rating The Advisors
4. Advisor Selection: Human vs. Algorithm

Figure 1

Experimental design. The experiment consisted of four stages: (1) presentation of the background story (medical or financial) and the required decision: select either a human advisor or an algorithmic advisor; (2) signing a disclaimer; (3) ratings of both advisors; and (4) making the decision: human or algorithmic
Figure 2

Responsibility ratings of advisors in the two domains: Medical and Finance, and in the two conditions tested: Neutral disclaimer and No-Responsibility disclaimer. The bars indicate the mean value, and error bars represent 95% confidence interval.
Figure 3

The percentage of participants who prefer a human/algorithm advisor in the two domains - Medical and Finance - and for the two disclaimer conditions (Neutral and No Responsibility disclaimers)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx