MUNI FSS

JOURNAL OF PSYCHOSOCIAL RESEARCH ON CYBERSPACE
CYBERPSYCHOLOGY

Schaap, G., Bosse, T., & Hendriks Vettehen, P. (2025). Beyond algorithm aversion: How users trade off decisional agency and optimal outcomes when choosing between algorithms and human decision-makers. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 19*(5), Article 8. https://doi.org/10.5817/CP2025-5-8

Beyond Algorithm Aversion: How Users Trade Off Decisional Agency and Optimal Outcomes When Choosing Between Algorithms and Human Decision-Makers

Gabi Schaap, Tibor Bosse, & Paul Hendriks Vettehen

Behavioural Science Institute, Department of Communication & Media, Radboud University, Nijmegen, Netherlands

Abstract

By definition, algorithmic decision-making (ADM) transfers human agency (i.e., decisional control) to the machine in exchange for optimal outcomes. Whereas algorithmic aversion literature suggests a default hesitance to use machine decisions, these conclusions have recently been disputed. This study examines the conditions under which individuals choose to delegate decisions to algorithms versus human agents, focusing on how trade-offs between agency and ADM benefits influence comparative evaluations. In three experiments (N_{total} = 841) in financial and romantic contexts, participants made a joint evaluation of algorithmic and human agents manipulated to offer varying levels of decisional agency and benefits. Contrary to prior research, we find limited evidence for algorithmic aversion as a default phenomenon. Instead, 1) both agency and benefits play critical roles in shaping ADM acceptance, and 2) individuals weigh these factors differently depending on the context. The findings highlight the complex context-dependent trade-offs between control and benefits shaping delegating preferences, and call for a more nuanced understanding beyond generalized algorithmic aversion.

Editorial Record

First submission received:
April 5, 2024

Revisions received:
April 2, 2025
August 14, 2025

Accepted for publication:
August 19, 2025

Editor in charge:
Alexander P. Schouten

Keywords: algorithmic aversion; technology acceptance; agency; delegation; trade-off; conjoint-choice experiment

Introduction

As AI technology becomes increasingly prominent in decision-making in our daily lives, the question becomes urgent if and under which conditions people will accept the decisions of evermore efficient and accurate intelligent machines. Is algorithmic decision-making (ADM, sometimes called automated decision-making) just a new tool in the human toolbox in this respect, evaluated by users in the same way as other technology? Or is it something fundamentally different? Research suggests that people are averse to using ADM agents, for instance if they are seen as imperfect or error-prone in making forecasts (Dietvorst et al., 2015, 2018; Prahl & Van Swol, 2017), because they are seen as 'mindless' when dealing with moral decisions (Bigman & Gray, 2018), or failing to account for human uniqueness in medical contexts (Longoni et al., 2019). Importantly, these studies suggest 'algorithmic

aversion' is present even when the machine offers great benefits in superior decisions and optimal outcomes, compared to human agents. The converging findings from these studies have been taken as proof of a default algorithmic aversion: A generalized rejection of algorithmic decision-making simply because it comes from a machine (see Burton et al., 2020; Jussupow et al., 2020 for overviews). This implies that indeed end-users see automated decision-making agents as unlike other technology, as adoption appears not to be determined mainly by its perceived benefits (cf. Ghazizadeh et al., 2012; Marangunić & Granić, 2015).

However, recently it has been suggested this conclusion may be premature, and a default aversion to algorithms may not exist (Logg & Schlund, 2024; Zehnle et al., 2025; Zhang & Gosline, 2023). Emergent recent research suggests that many times users are quite willing to rely on algorithmic advice in equal circumstances (Araujo, et al., 2020; Rabinovitch et al., 2024), for instance when they do not see the algorithm make mistakes (Logg et al., 2019), when the predicted outcomes are superior (Dietvorst & Bharti, 2020; Schaap et al., 2023), or when they have some measure of control over the algorithmic decision (Dietvorst et al., 2018; Schaap et al., 2023). Furthermore, while people may occasionally express having a preference for human advice, this does not predict actual usage of the advice (Himmelstein & Budescu, 2023). Researchers have pointed to a number of study design and measurement issues and overgeneralized interpretations of findings prohibiting a definitive conclusion (Pezzo & Beckstead, 2020a, b; Logg et al., 2019; Logg & Schlund, 2024; Schaap et al., 2023). One issue is that some studies provide no detailed (Pezzo & Beckstead, 2020b) or unequal information on the expertise, accuracy, or decision-making processes of the algorithmic agent, stacking the odds in favor of the human (Schaap et al., 2023). This makes meaningful attribution of any observed aversion or appreciation among participants to any objective difference between the two agents challenging (Jussupow et al., 2020).

In all, these divergent results raise the question exactly what processes may be at work to determine a choice between algorithmic and human agents. Are users 'automatically' averse to decisions merely because they come from machines? Or are perhaps other decision-making processes at play in determining preferences?

The current study aims to investigate in greater detail to what extent 'default' algorithm aversion or information about other prominent attributes of decisional agents play a role in people's preference for an algorithm or human decision maker. Many technology-related (e.g., efficiency, the potential for offloading), user-related (education, self-efficacy, familiarity), and (social) context-related factors (e.g., subjective norms) are likely important in predicting ADM use. Investigating them all falls outside the scope of the current research. But by definition, one crucial attribute separating ADM from prior technology is the transfer of decisional agency ('control' over decisions and their implementations) from human user to machine (Sundar, 2020). This attribute may simultaneously define what is so groundbreaking about this new technology and what occupies users in evaluating whether it is opportune to use it. Although retaining human agency has been flagged as an important concern by both scholars and the public (e.g., Araujo, 2018; Jia et al., 2012; Pew Research Center 2018, 2023; Sundar, 2020), its potential causal relation to actual algorithm use has never been investigated. We hypothesize the transfer of agency may be a vital factor in determining acceptance of this technology by users. Our goal is to investigate the relative importance of agency and relevant potential benefits of ADM in its acceptance. Moreover, we examine how this cost-benefit evaluation compares to a default aversion as a factor predicting ADM preference.

We do this in a series of conjoint-choice experiments in which users choose between using an algorithmic and human decisional agent, which are both described in terms of their associated user agency and outcome benefits. This paradigm allows us to ascertain whether and how people trade off agency and optimizing outcomes to choose a decisional agent. Understanding how information about these relevant attributes causes users to accept or reject ADM is important because it increases understanding of what we do and do not accept from machines. In the near future machines are poised to take over many decision-making tasks from humans in virtually all life domains, making decisions for us and about us. It is critical to discuss the intended and unintended consequences of this technology (Castelo & Lehmann, 2019), to decide the extent to which we want it to take over our decision-making faculties (Hannon et al., 2024). By investigating how individual users weigh their loss of agency against the potential advantages of ADM, our research provides relevant input for this debate.

Preference for Algorithmic Decisions as a Function of Cost-Benefit Trade-Off Between Decisional Agency and Benefits

Based on decision-making research in evolutionary biology, psychology, and affiliated disciplines such as economics, an explanation for how people choose a decisional agent assumes that people trade off the costs and

benefits of a choice according to the subjective value attributed to them (Berkman et al., 2017). As any decision must be made within the constraints imposed by the environment, cost-benefit evaluations are adaptive responses to that environment. In evolutionary biology, fitness benefits such as increased food intake must be weighed against fitness costs, such as predation risk by spending more time foraging (Stevens, 2008). This balancing act should result in an optimum where the benefits are deemed sufficient to offset the costs (Parker et al., 1990). In psychology, decisions are bounded by constraints on for instance capabilities, time, and effort. The requisite trade-off evaluation is expressed in for instance how people spend physical effort to attain some reward (Richter et al., 2016), or choose to eat food that is unhealthy but tasty (Stroebe et al., 2013). Neuroscience suggests people trade off cognitive effort and time investment costs versus rewards (Shenhav et al., 2017), or trade off spending computational resources versus accurate predictions (Friston, 2010). In economics, predicted utility—the result of cost-benefit evaluations—is central in financial decisions (Sanfey et al., 2006).

In this research, we focus on the potential trade-off between two attributes that define ADM: degree of user agency (control) transferred from user to agent, and the degree of tangible benefits of the decisions made. To start with the latter, the potential tangible benefits of algorithmic decisions over human decisions are clear: they consist of every positive outcome from the decision resulting from its automated character, whether it be (higher) material rewards, efficiency, accuracy, or convenience. Cross-sectional research on algorithmic decision-making shows that users have a greater preference for computers or humans that exhibit a higher success rate (Kramer et al., 2018). Experimental studies indicate that users will prefer the agent that is expected to yield the most optimal outcome in terms of primary and secondary rewards, accuracy, or success rate (Bigman & Gray, 2018; Castelo et al., 2019; Pezzo & Beckstead, 2020a, b; Schaap et al., 2023).

The second attribute of ADM that is the focus of this research, is the degree of decisional control or *agency* – the capacity or perception of individuals to control actions to produce desired outcomes (Sundar, 2020) – it transfers from user to machine. Several principal psychological theories highlight the fundamental importance for humans to exert voluntary control over their environment for their physical and psychological functioning, in related concepts such as self-determination or autonomy (Deci & Ryan, 2000), locus of control (Rotter, 1966), and self-efficacy (Bandura, 2000): Humans need to have the sense that they are the agents of their own actions (Haggard & Eitam, 2015; Moore, 2016). Cross-sectional data suggest 'control' is one of the primary public concerns regarding algorithmic decision-making (Araujo et al., 2018; Castelo et al., 2019, Table 5; Pew Research Center, 2018; Stein et al., 2019; The Royal Society, 2017). Human-machine interaction research finds that people may be hesitant to cede too much decision-making control to machines (e.g., Chugunova & Sele, 2020; Coyle et al., 2012; Sundar & Marathe, 2010). Furthermore, various recent experimental studies provide first indications that a greater ability to influence an algorithmic decision may lead to a greater preference for an agent (Bigman & Gray, 2018; Dietvorst et al., 2018; Palmeira & Spassova, 2015; Schaap et al., 2023). This suggests that one of the defining characteristics of automated decision-making, the fact that it seizes agency from its human user, may be a crucial factor in its acceptance. In the current research, we operationalize agency as the 'amount of control over the decisions and their implementation' granted by the algorithmic of human agent. Unfortunately, there is a paucity of research explicitly manipulating user agency while keeping equal other human and algorithm attributes. Furthermore, it is unknown how the value attached to agency is weighed against the value of superior decisional outcomes.

In short, we argue that the preference for an agent, whether an intelligent machine or a human being, is primarily a function of evaluations in which the costs of relying on the agent are pitted against potential tangible benefits of the outcome of a decision. In our experiments, we ask participants to choose between an algorithm and a human agent to make decisions in a 'joint evaluation' paradigm. Here, users explicitly compare both agents on the basis of information presented about the relative costs (either having no control over a decision or a lower reward) and benefits (either having full control or greater rewards) associated with each agent. Offering information on the costs and benefits equally for both the human and algorithmic agent provides users an opportunity for fair comparison based on the agents' merits (Hsee et al., 1999), that has been sometimes lacking in ADM research (Jussupow et al., 2020; Logg et al., 2019; Schaap et al., 2023). In doing this, we are able to ascertain 1) whether a default algorithmic aversion exists, or 2) whether and how people employ a cost-benefit evaluation between agency and optimal outcomes. Although it would be fruitful for debate, design, and policy making to understand the trade-offs (Sundar, 2020), to date, no research exists investigating the influence of these explicit trade-off evaluations. Based on the above we expect that in a joint evaluation, users will have a relative preference for the agent (human or machine) that offers the greatest agency or the greatest benefits, regardless of whether that agent is human or machine. In addition, recent research suggests that people may have different preferences for human or algorithmic forecasting advice depending on the domain of the advice (Himmelstein & Budescu, 2023).

Therefore, the current research investigates the processes involved in choosing algorithmic vs. human decision making in two distinct domains: finance (stock market investment) and romance (dating).

Study 1: Stock Market

A first study investigated the preferences of users in the context of stock market investments, in which human and algorithmic decision-makers offered higher or lower return on investments and provided the user decisional control or not. Today, an estimated 60–75% of overall trading volume in major markets is done by algorithms (Groette, 2024). Stock investment scenarios have been successfully used in previous algorithmic acceptance studies (Castelo et al., 2019; Niszczota & Kaszás, 2020; Önkal et al., 2009).

Methods

Preregistration

This study was preregistered on the Open Science Framework; all hypotheses, measures and materials, as well as data and analysis protocols can be found at https://osf.io/49r3x/overview.

Design

We used stock investment scenarios in a 2 (Human control yes/no) X 2 (Human benefits high/low) X 2(Algorithm control: yes/no) X 2(Algorithm benefits high/low) between-subjects vignette experiment. Participants were randomly assigned to one of 16 decision scenarios in a paired conjoint paradigm. A paired conjoint design jointly presents two choices and their respective attributes, thus maximizing the amount of relevant information available to the user (Hainmueller et al., 2015).

Participants were asked to imagine that they were to invest part of their income, and that they could choose either an algorithm or a human expert to make the decision whether to invest their money. In a joint evaluation paradigm, the descriptions of both the algorithmic and human expert were presented simultaneously, and both the amount of control over the decision and the height of benefits resulting from the decision were manipulated. An overview of the full design can be found in the Appendix A.

At the end, participants were asked to choose between an algorithm and a human as decision-maker. Depending on the conditions to which they were assigned, the algorithmic and human decision maker differed or not regarding the level of benefits from the decision, and the level of control over the final decision they offered. The ethical committee of the authors' institution approved the project under identification code ECSW-2019-169.

Sample

We used the Prolific Academic survey platform to recruit participants (Peer et al., 2017). We included adult UK participants whose first language is English. Also, we included only Prolific users with an approval rate above 80%. Participants were paid £1,35 for max. 10 minutes of participation. We used G*Power software (Faul et al., 2007) to conduct a sample size analysis. Our goal was to obtain .95 power to detect a medium effect size of .25 at the standard .05 alpha error probability. The analyses included 16 treatment groups, with main effects and interactions. This showed that a sample of 210 would be sufficient to detect a medium effect size. After recruitment, exclusion of incomplete questionnaires and replacement of participants who failed attention checks (N=13), our final sample was 211 (Mean age = 40.72, SD = 13.36; 73% female; 15% at least Graduate degree; 8% had a background in Computer Science, AI, or related field; and 6.6% in finance, stock exchange or related field).

Procedure

Before proceeding to the study proper, eligible participants were first required to give their informed consent. Subsequently they were asked to read a scenario, and imagine the events "as if they were really happening to you". They then read the scenario, which was divided into several separate parts, each on its own screen, followed by a summary table. After reading the scenario, the participants answered questions regarding their choice of

agent. This was followed by two questions to check their attention to the scenario, and finally by demographics, and questions on education and employment in relevant fields such as computer science or programming. Finally, participants were thanked and redirected to the Prolific portal to collect their participation incentive.

Materials

In each condition participants read the same scenario in which a decision is made regarding their own financial investment in the stock market. Participants were instructed to "Imagine you have sufficient money to invest in the stock market, and that you are interested in investing 10% of your monthly income." Subsequently, they were asked to choose whether to use an algorithm or a human expert to make this decision. To facilitate the choice between these two options, the scenario first provided descriptions of both the algorithmic and the human expert option. Finally, we presented a table summarizing the relevant attributes of each option conjointly (see Figure 1). In the descriptions of both options, we manipulated the amount of user agency (control over the decision) and the height of benefits resulting from the decision, as follows:

Control (levels: yes/no). The 'Control' factor presented the scenario as either an automatic decision that will be executed by the (algorithmic or human) agent beyond the participant's control or a recommendation that awaits the participant's approval. No Control: "the decision is made for you. The choice which investments to make will be made automatically. You cannot influence the decision. The sum will automatically be transferred from your bank account." Control: "you have final say over the decision. You will be advised which investments to make. You may choose to follow the advice or not. You may or may not give permission to transfer the sum from your bank account. It is up to you."

Benefits (levels: high/low). This factor was manipulated by adding either a high or low return on investment rate to each scenario, and was based on the average return on investment in the global stock market being around 10% (Campbell & Safane, 2025): Low Benefits: your profit is estimated at 3% return on your investment. High Benefits: your profit is estimated at 21% return on your investment.

Figure 1. *Summary as presented to participants (Example: condition 4).*

	Who is assisting?	Type of assistance	What is your estimated profit?
Option A	EXPERT:	Advice: the final decision is up to you	3%
Option B	ALGORITHM:	Decision: Decision is made for you	21%

Measures

Dependent Variable. The primary dependent variable was 'agent choice', measured on a 7-point semantic differential scale: *reviewing the options above, which would you choose to assist you?*, with polar points of (1) *definitely the expert* through (7) *definitely the algorithm* (M = 3.75, SD = 2.34). The range of scores between the extremes is intended to capture the nuances that underly many discrete choices in human life, because choices between options are based on relative preferences for each option, which may vary.

Attention Check. To ensure that participants understood the options offered in the scenario, we asked two questions on the estimated profit offered by choosing the algorithm (98% correct answers), and the control offered by the human expert (95% correct).

Analysis

Participants who failed to correctly identify the options offered in the scenario (N = 13 of the initial N = 210) were excluded from the analyses, and new participants recruited. The analyses for the hypotheses test were rerun with the complete sample (N = 223), including the participants who failed the check (cf. Aronow et al., 2019; Montgomery et al., 2018). This produced essentially the same results.

We found a number of outliers with regard to completion time in the remaining sample with a standardized score over 3 (N = 5). Below, we report the analysis on the full final sample (after re-recruitment, N = 211), including the outliers. Analyses on the sample without the outliers produced the same results. We used four-way ANOVAs with Human Control, Human Benefits, Algorithm Control, and Algorithm Benefits as factors, and with Agent Choice as dependent variable.

Results

In answer to our research question, we find that a preference for the human agent, all else being equal, is not readily apparent. The mean scores of Agent Choice in a one-sample *t*-test, in which we set the test value at the midpoint of the scale (4), signify no preference for either agent. The test shows a mean score of 3.75 (SD = 2.34), which indicates a slight but statistically insignificant preference for the human agent (p = .127, Cohen's d = −.11).

We expected that users would prefer the decision agent that offered them the greatest benefits and control over the decision (Table 1: for reasons of brevity and clarity, we limit the information in the running text to Means, Standard Deviations, and Effect Sizes, and refer to the Tables for further information on the ANOVAs). Benefits have a large effect on Agent Choice, both when the benefits are associated with the algorithm (η^2 = .255) and with the human agent (η^2 = .201). Inspection of the means shows that these effects are in line with the hypothesis, with higher benefits leading to greater preference for that agent (for Algorithm: high Benefits M = 4.97, SD = 2.11; low Benefits M = 2.57, SD = 1.90, with a score closer to 7 meaning a stronger preference for the Algorithm; for Human: high Benefits M = 2.67, SD = 2.03; low Benefits M = 4.80, SD= 2.14, with a score closer to 1 designating a stronger preference for the Human). Control has a small effect on Agent Choice, both when the level of control is associated with the algorithm (η^2 = .014) and with the human agent (η^2 = .020). Once more the patterns for the means are in line with the hypotheses (For Algorithm: Control M = 3.99, SD = 2.37, No Control M = 3.49, SD = 2.28; for Human: Control M = 3.43, SD = 2.30, No Control M = 4.10, SD = 2.34). Thus, we conclude that our hypothesis is confirmed: Users have a greater preference for agents that offer them greater benefits and greater control.

Table 1. *The Effect of Control and Benefits on Agent Preference: Monetary Rewards (ANOVA).*

	Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
Corrected Model	598.935	15	39.929	14.202	<.001	.522
Benefits for Algorithm	289.500	1	289.500	102.968	<.001	.255
Control for Algorithm	15.367	1	15.367	5.466	.020	.014
Benefits for Human	227.741	1	227.741	81.002	<.001	.201
Control for Human	22.934	1	22.934	8.157	.005	.020
Benefits for Algorithm * Benefits for human	13.059	1	13.059	4.645	.032	.012
Residuals	548.250	195	2.812			

Note. Type III Sum of Squares. Only significant interactions are shown. Full Table in Appendix B.

Out of 11 potential interactions there is only one relatively weak interaction, between Benefits offered by the algorithm and those offered by the human expert (Table 1). Inspection of the means suggests that when either agent offers larger benefits than the other, users prefer that agent, in the same way for the algorithm and the human: The difference to the maximum choice score for both is 0.8 point on the 7-point scale (High Benefits Algorithm * Low Benefits Human M = 6.20, SD = 1.17; High Benefits Human * Low Benefits Algorithm M = 1.78, SD = 1.49). If however both human and algorithm offer equally small (M = 3.38, SD = 1.95) or equally great benefits (M = 3.64, SD = 2.10), users seem to slightly prefer the human expert (as 4 is the midpoint of the scale).

Conclusion Study 1

The first study yields little evidence of 'algorithmic aversion', nor of its opposite, algorithmic appreciation. Our analysis shows no significant difference in preference for any agent. The results suggest that indeed a cost-benefit trade-off exists. In line with our hypothesis, both monetary rewards and decisional control are relevant in determining the use of automated (and human) agents: Users opt for the agent that offers them the largest benefits and the greatest control, regardless of whether the agent is human or a machine. However, compared to benefits, control has only a modest impact on preference.

Study 2: Incentivized Interaction Experiment

The previous study used scenario-based conjoint choice paradigms to test the hypotheses. Research suggests that vignettes and conjoint choice designs are capable of mimicking choices made in real-world situations (Hainmueller et al., 2015). However, recent studies on ADM acceptance provide strong hints that stated preferences do not necessarily translate into actual behavior in using algorithmic agents, and that reactions to hypothetical scenarios differ from real judgments (Himmelstein & Budescu, 2023; Logg & Schlund, 2024). As these studies may offer clues as to why some prior research has found algorithmic aversion, and to further improve mimicking of real-world choice, in Study 2, we attempt to replicate the results of Study 1 by adapting the stock market scenario used in Study 1 into a paradigm with realistic interaction and in which decisions made by the participant had actual monetary consequences.

Methods

Preregistration

This study was preregistered on the Open Science Framework; all hypotheses, measures and materials, as well as data and analysis protocols can be found at https://osf.io/u43z9/overview.

Design and Sample

We used the same 2 (Human control yes/no) X 2 (Human profit high/low) X 2 (Algorithm control: yes/no) X 2 (Algorithm profit high/low) between-subjects design as in the prior studies. Here, participants were randomly assigned to one of 16 'stock market investment' conditions. They received £1 to invest in the stock market. In each condition they had either final control over the investment, or not; they also got either high or low predicted return of investment. The dependent variable was the same as in the prior studies. The sample size and its rationale were the same as previous studies (N = 210; Age M = 39,36, SD = 12.67; 50% female, 19% at least Graduate degree, 16.2% had a background in Computer Science, AI, or related field, and 11% in finance, stock exchange, or related field). The ethical committee of the authors' institution approved the project under identification code ECSW-2023-057R1.

Procedure

After giving informed consent, participants were told they would be given £1 to invest in the stock market. They read descriptions of the various human and algorithmic options available to make the investment, before picking one of the available options. Subsequently, they rated their preference for either agent, and answered questions on a number of control variables. At the end of the experiment they were told the outcome of the investment, which was added to their participation fee of £1.20 for approx. eight minutes.

Materials and Measure

Participants read five screens with short texts introducing them to the investment scheme. Based on offerings of real-world online investment companies, they were told that in addition to their participation fee, they would receive £1 to invest in short-term stock movements. Using a real-world investment technique called '1-minute scalping', they would know the result of any investment by the end of the experiment. It was made clear that any investment could result in a profit or loss. They were then presented with two options: Either an algorithm or a human expert would make the investment for them. Both options presented the same range in terms of agency and expected benefits as in Study 1. If participants picked an agent that offered them decisional control, they were subsequently presented with the opportunity to either close the sale (via a 'BUY!' button) or withdraw from the investment (through clicking an "ABORT!' button). If they chose the latter, they were told they would receive the £1 investment budget in addition to the default participation fee. In all other cases, all participants received £1.21 (the maximum 21% expected profit as presented to them), thus doubling their participation fee. In line with the

previous studies, preference was measured both by a continuous measure (*Reviewing the options above, please tell us how strongly you prefer one option over the other for this investment*) with a 7-point scale running from 1 (*I definitely prefer the expert*) to 7 (*I definitely prefer the algorithm*).

Results

Results from Study 2 show that in a situation with live interaction and real material consequences, the findings from scenario Study 1 become more pronounced (Table 2). There were no indications of a default preference for the human or algorithmic agent, in a one-sample *t*-test (M = 4.09, SD = 2.06, p = .525) t = .637 (df = 209).

Here, only benefits have a significant effect on agent preference, with higher return on investment leading to a greater preference for an agent compared to a lower expected profit. This is true for both the algorithm (high M = 5.11, SD = 1.84; low M = 3.07, SD = 1.74), and the human benefits (high M = 3.14, SD = 1.82; low M = 4.99, SD = 1.86), with again a higher mean pointing to a stronger algorithmic preference.

Table 2. *The Effect of Control and Benefits on Agent Preference: Incentivized Stock Market Investment (ANOVA).*

	Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
Corrected Model	437.867	15	29.191	12.657	<.001	.495
Benefits for Algorithm	220.731	1	220.731	95.709	<.001	.248
Control for Algorithm	4.461	1	4.461	1.934	.166	.005
Benefits for Human	190.055	1	190.055	82.408	< .001	.214
Control for Human	0.644	1	0.644	0.279	.598	.000
Residuals	447.414	194	2.306			

Note. Type III Sum of Squares. Only significant interactions are shown. Full Table in Appendix B.

Conclusion Study 2

In line with Study 1, the findings of the interactive and incentivized experiment show no evidence for aversion towards algorithmic decisions. Regarding the question whether a cost-benefit evaluation between optimal outcomes and control takes place, the experiment provided a slightly different picture than Study 1. We can now state with greater confidence that indeed, at least in a situation in which real-life primary benefits are at stake, users prefer the agent that most likely assures the greatest benefits. Moreover, the findings show that in this real stakes context, agency is not important when picking a decisional agent.

Study 3: Dating

Our first two experiments suggest that optimizing outcomes is important in choices regarding ADM. Control may be of some relevance, whereas the nature of the agent is irrelevant. However, both experiments tested our assumptions in a financial domain, where criteria of what constitutes an advantageous outcome are fairly clear: profit is good, loss is bad. To test the robustness of these findings, it is prudent to test them in a different one, as different domains may result in different evaluations of ADM (Araujo et al., 2020). Morewedge (2022) proposes that a preference for humans over algorithms may only occur where evaluative criteria are ambiguous, and in domains where our individual identity is threatened. Indeed, prior research found that the decision domain may be important in this regard. Himmelstein and Budescu (2023) found that judges express a preference for human forecasters in one domain, while preferring algorithmic advice in another. And Castelo et al. (2019) suggest that aversion is stronger in more 'subjective' domains (e.g., related to taste or emotions) than in 'objective' domains' such as stock market predictions. Finally, Longoni et al. (2019) find that a concern of AI potentially neglecting their individual uniqueness, leads to patients rejecting medical AI. The question is whether agency has the same weight in these evaluations when the evaluative criteria of what constitutes a beneficial outcome are less clearly defined, and where decisions may potentially threaten individual identity. Based on these notions, we conducted an experiment using a scenario from a domain where evaluation criteria are likely more ambiguous than in finance, and in which individual identity may be more directly threatened by adverse decisions: romantic relationships.

We use the same basic set-up used in Study 1, to present participants with a scenario, in which a human or algorithmic decision-maker picks a person for a romantic date in an online dating setting. In this study, agency is

operationalized in the same way as in Studies 1 and 2. Benefits are slightly different; instead of monetary rewards, the benefits now relate to a higher accuracy of the match made with a potential partner. Accuracy has been known to affect trust in (Martens et al., 2024) and acceptance of ADM (Burton et al., 2018; Jussupow et al., 2020), but research comparing its impact all else being equal, is scarce. Finally, one recent study demonstrates that when users must trade-off AI explainability and accuracy, they prioritize accuracy (Nussberger et al., 2022). Based on the above, we hypothesize that both agency and accuracy (as an alternative operationalization of benefits) play a role in accepting algorithmic decisions, and investigate how they are weighed against another in choosing a decision agent.

Methods

Design and procedures were the same as in Study 1, and preregistered at the same location. Following the same power analysis as in Study 1, a sample (N = 210) was recruited from Prolific.co, with one criterion added: only ages 18–50 were eligible, to ensure greater familiarity with dating services and dating websites (Mean Age 32.72, SD = 8.49; 70% female; 27% at least Graduate degree; 9,5% had a background in Computer Science, AI, or related field, and 59% had used dating websites before). All measures were the same as in Study 1. In Study 3, we used scenarios in which participants imagined using a dating service. Similar to Study 1, in the scenarios they were presented with two conjoint options, where a match with a potential romantic partner was made by either a human expert or an algorithm. Each option offered either decisional control or no control for the user, and a high (95%) or low (75%) accuracy in the match with their potential date (cf. Alexander et al., 2018; Bigman & Gray, 2018).

Results

In contrast to our previous studies, a one-sample t-test shows a slight and significant preference for the human expert (M = 3.5; SD = 2.21, p = .001, Cohen's d = −.23).

A four-way ANOVA with Human Control, Human Benefits, Algorithm Control, and Algorithm Benefits as the factors, and choice as the dependent variable, also produced somewhat different results compared to Study 1 and 2 (Table 3).

Table 3. *The Effect of Control and Benefits on Agent Preference: Accuracy (ANOVA).*

	Sum of Squares	*df*	Mean Square	*F*	*p*	η²
Corrected Model	548.098	15	36.540	15.070	<.001	.538
Benefits for Algorithm	101.325	1	101.325	41.788	<.001	.101
Control for Algorithm	149.105	1	149.105	61.494	<.001	.148
Benefits for Human	75.590	1	75.590	31.175	<.001	.075
Control for Human	134.579	1	134.579	55.503	<.001	.133
Control for Alg. * Control for human	30.925	1	30.925	12.754	<.001	.031
Benefits for Alg. * Benefits for human * Control for human	11.161	1	11.161	4.603	.033	.011
Control for Alg. * Benefits for human * Control for human	19.648	1	19.648	8.103	.005	.019
Residuals	470.397	194	2.425			

Note. Type III Sum of Squares. Only significant interactions are shown. Full Table in Appendix B.

Our hypothesis was confirmed: Users have a greater preference for agents that offer them greater accuracy and greater control. Higher accuracy leads to greater preference for that agent, with a score closer to 7 meaning a stronger preference for the Algorithm, and a score closer to 1 designating a stronger preference for the Human: for Algorithm: high Benefits M = 4.19, SD = 2.23; low Benefits M = 2.89, SD = 1.96; for Human: high Benefits M = 2.90, SD = 2.09; low Benefits M = 4.11, SD = 2.16, and the same pattern for control (For Algorithm: Control M = 4.36, SD = 2.16, No Control M = 2.66, SD = 1.92; for Human: Control M = 2.64, SD = 1.83, No Control M = 4.33, SD = 2.24). In deviation from Study 1 and 2, the effects for Control are now large, and greater than the medium effects for Benefits, though not as great as the effects of Benefits in Study 1 and 2 (cf. Table 1 and 2).

Three interactions were significant, but with small effect sizes. The largest of these is the two-way interaction Control for Algorithm * Control for Human: p <.001, η^2 = .031. The means show that if the algorithm and human agent offer the same amount of control (equally high or low), participants' preference score is somewhat below

the midpoint of the 7-point scale, indicating a slight preference for human agents (Algorithm No Control * Human No Control M = 3.11; Algorithm Control * Human Control M = 3.10). If the agents differ in terms of level of control, participants prefer the agent offering the highest control: Algorithm Control * Human No control M = 5.55 (preference for Algorithm); Algorithm No Control * Human Control M = 2.22 (preference for Human).

Conclusion Study 3

While Studies 1 and 2 used financial scenarios to test our hypotheses, Study 3 focused on the domain of romantic relationships. This yielded partly different results. First, agency had a stronger impact on agent preference in the dating domain as compared to the financial domain. Moreover, its impact in the dating domain was also greater relative to the benefits.

Study 3 provides some evidence for a relative preference for human agents. One explanation for this result might be that dating represents a domain that is perceived as having more subjective aspects, compared to stock market investments (Castelo et al., 2019; Morewedge, 2022), and thus is (relatively) better left to a human decision maker. Once again, we conclude that these results show that people evaluate an agent based on the amount of control they are offered as well as the benefits of a decision. Only if there is no clear advantage or disadvantage on these criteria, they have a slight preference for a human agent.

General Discussion

By definition, algorithmic decision-making diminishes human agency in exchange for optimal outcomes. Whereas algorithmic aversion literature argues that people have a default hesitance to delegate decisions to machine, an emergent volume of recent research shows this debate is by no means settled and that factors beyond the machine nature of ADM may be considerably more important. Therefore, the current research investigated the extent to which a user's agency (control over decisions and their implementation) is weighed against potential benefits of optimal outcomes of ADM in choosing between algorithmic and human decision-makers. Our aim was to investigate the respective roles in this process of user decision agency and the benefits of optimal outcomes and to simultaneously assess the role of algorithmic aversion.

In contrast to prior research, we offered elaborate and equal information on important characteristics of both human and algorithmic agents in a paired conjoint paradigm, thereby providing the opportunity for a fully informed and explicit comparison. Furthermore, to explore whether the processes involved varied across everyday decision domains, we conducted experiments in two separate domains: finance and dating. We also applied two types of benefits to match the respective domains: monetary rewards and decision accuracy. Finally, using both scenario-based and incentivized interactive experimental designs allowed for drawing of causal conclusions.

The research yielded two main findings. First, we find limited evidence for the existence of algorithmic aversion as a default, generalized phenomenon. Although in the dating domain, there is a slight relative preference for a human decisional agent, this human preference is not consistent across studies, and effect sizes are small, certainly when compared to the effects sizes associated with the agency and benefits factors. In line with recent research, we conclude that evidence is too weak to maintain that a default algorithm aversion exists (Zhang & Gosline, 2023).

Second, we show that rather than the machine or human nature, both agency and benefits play critical roles in shaping ADM use. Importantly, a trade-off evaluation is employed by which users weigh agency against outcome benefits to determine which agent to pick for making the decision. Importantly, the outcome of this trade-off, and whether agency or benefits are weighed the heaviest, depends on the decisional context.

The research contributes to the literature in several ways. First, we provide important input in the nascent debate on algorithmic aversion. Counter to arguments in prior aversion literature that users reject using decision-making algorithms – even if they have greater benefits – the current research finds no proof of a default aversive setting overruling all other considerations in decision-making. In part, new literature questioning the existence of aversion as a default state (cf. Himmelstein & Budescu, 2023; Logg & Schlund, 2024; Morewedge, 2022; Zehnle et al., 2025; Zhang & Gosline, 2023) has posited that previous aversion effects may stem from measurement issues and the absence in stimulus material of information, or uneven information between human and algorithm relevant to decision-making, and not necessarily from a general antipathy against algorithmic agents (Jussupow et al., 2020;

Pezzo & Beckstead 2020a, b; Logg et al., 2019; Logg & Schlund, 2024; Schaap et al., 2023). This may lead participants to not realizing the usefulness of algorithms in a given context, or resorting to falsely estimating that an algorithm is inferior to a human in the required task (Arkes, 2008). By keeping the odds even for human and algorithm (giving participants detailed information about the algorithm's expertise and performance), we show that if comparison is maximized by providing relevant information in a paired conjoint design, people seem to care less who or what the agent is, and more about the costs and benefits of a decision. In that sense, our findings do not necessarily contradict prominent studies often cited as evidence of algorithmic aversion (cf. Bigman & Gray, 2018; Dietvorst et al., 2015; Longoni et al., 2019), as these works typically demonstrate *reduced* algorithm preference under specific conditions—such as errors, abrupt agent changes, or unequal information—rather than default aversion, and frequently do not claim otherwise. Our findings and those of the recent research cited above, suggest it may be time to abandon the label 'algorithmic aversion' as a general phenomenon, as it does not adequately represent the complex processes at work in determining reliance on ADM.

Secondly, we contribute to scientific knowledge by showing for the first time the causal role of a fundamental attribute of ADM: agency (decisional control). Because it is such a central characteristic, many in society see this as one of the great concerns associated with decision-making technology (Araujo et al., 2018; Pew Research Center, 2018, 2023). This stems from the fact that having agency is a basic human need (cf. Deci & Ryan, 2000), and more specifically, that users are unwilling to fully cede (decision) authority to algorithmic or other actors (cf. Chugunova & Sele, 2020; Coyle et al., 2012). Our experimental research shows that agency is a crucial issue for users, not only in general attitudes as found in survey research, but also as a decisive psychological factor at the moment a decision must be made.

Third, we are the first to demonstrate the explicit trade-off between the fundamental attribute of agency and the tangible benefits offered by ADM. By using a conjoint choice design, we made this trade-off explicit for the participant: findings show that when dealing with algorithmic decisions, people assign varying values to the prospect of optimal outcomes from ADM versus the need to keep decisional agency. The context-dependent nature of the outcome of this trade-off evaluations, highlights the complex and nuanced nature of the psychology of algorithmic delegation.

Relatedly, our finding of a preference for a human agent alongside the cost-benefit factors in the domain of romantic relationships adds credence to novel ideas that, while perhaps not default, a preference for human agents may nonetheless be a significant factor in some domains, even when the user is fully informed about the pros and cons of using each agent. Recently, researchers have suggested that a preference for algorithms (algorithmic appreciation) may be related to domains that have clearly defined evaluative, or 'objective' criteria (such as in our case, stock market investment), whereas more 'subjective' domains in which criteria of success or failure are less evident may elicit a relative preference for human agents, such as selecting a romantic date (Castelo et al., 2019; Himmelstein & Budescu, 2023; Morewedge, 2022). Future research may be tasked to find out what other factors are included in user evaluations, for instance in different life domains, and how they are valued in relation to each other.

Combined, these findings provide a theoretical contribution, suggesting that acceptance of ADM technology depends both on value-based cost-benefit trade-off processes (Studies 1–3) and more automatic processes in which 'machine heuristics' (Sundar & Kim, 2019), such as algorithmic aversion may play a role (Study 3). This may imply a dual process model of algorithmic acceptance, comparable to for instance one such model proposed previously in research on management information systems acceptance (Bhattacherjee & Sanford, 2006). It would be important to find out whether such dual processing interpretations are applicable to acceptance of ADM technologies, to explore the factors predicting heuristic and systematic trade-off routes to acceptance or rejection, and to see whether aversion or preference entail a heuristic process or whether they emanate from more thoughtful processes as well.

In the current research, we posited that user agency and benefits such as monetary rewards and accuracy are crucial factors determining ADM acceptance. To be sure, as both the costs and benefits involved in these evaluative processes are likely multifaceted, a broader model of value-based evaluations of ADM would incorporate predicting factors beyond those two. Theoretical models on the acceptance of more traditional (i.e., non-intelligent) technology, such as the Technology Acceptance Model and its derivatives (Davis, 1989), and the Information Systems Success Model (DeLone & McLean, 1992) suggest similar cost-benefit processes, and the cross-sectional research associated with these models points to factors such as ease of use, and usefulness (such as efficiency, quality). However, as the above models predate them, they do not account for factors that may be

unique to the 'intelligent' technologies addressed in the current research, such as agency (De Freitas et al., 2023). We propose that in future models on the acceptance of or reliance on decision-making technology, agency may be worth incorporating. Further steps in this line of research should include disentangling in detail the multidimensionality of the costs and benefits of ADM acceptance. Value-based approaches from psychology, such as Berkman et al.'s (2017) value-based choice model may be of use here. The model distinguishes between three broad types of costs and benefits: tangible (e.g., primary rewards and costs and efficiency benefits: these are the benefits addressed in our paper), self-related (among which agency), and social (such as norm conformity and status). All of them are likely highly relevant in determining ADM acceptance and should be considered in future research.

In sum, this study contributes to the understanding the psychological process of how people accept or reject algorithmic decision-making by suggesting a new way of thinking about how people approach the questions posed by algorithms as decision makers. Rather than treating ADM technologies as uniquely different from other technologies, much of the general user processes involved in accepting them may not be that different from other technology, being largely based on cost-benefit evaluations. In a way, although conscious of their unique and possibly era-defining qualities, from a user perspective we may regard AI and algorithms as just another 'tool' in the shed (cf. Pew Research Center, 2020). Humans will meet the questions posed by these new technologies and associated challenges as they have been doing for centuries: by using value-based evaluations in which the costs and benefits of their actions are weighed against one another. If these subjective 'calculations' lead to a summation in favor or against using algorithmic decision-making, we will act accordingly. But looked at a more detailed level, the current research adds to prior research and theory involving cost-benefit processes in technology acceptance (cf. Marangunić & Granić, 2015) that it accounts for the novel dimension brought by ADM: the reshuffling of agency in the decision-making process.

The societal impact of this research may be significant. If our findings are to be believed, especially the fact that people frequently opt for the agent that allows them the *least* amount of agency in exchange for benefits, suggests this technology that is able to take over or diminish human agency will in many instances be quite easily adopted by the general public, although perhaps not as easily in every societal domain. In addition to being informative for designers, this is vital input for a public debate on this kind of technology, in which we must decide whether we want it adopted at large scale, precisely because of the agency issue. National and supranational governments are starting to develop and implement policies to stem the unbridled development of autonomous agents and related intelligent technologies, to guarantee safety, fundamental rights and human-centric AI (cf. Buther & Beridze, 2019; Galindo et al., 2021; Gstrein et al., 2024; Langman et al., 2021; Winfield, 2016; Yeung, 2020 for various overviews). But much of the acceptance of technology is determined in day-to-day interactions with users. Our findings illustrate that, alongside the potential advantages, we must carefully consider the costs to individual users, including, but not limited to loss of autonomy, before diving wholesale into a society dominated by decision-making digital agents (Castelo & Lehmann, 2019; Hannon et al., 2024).

Conflict of Interest

The authors have no conflicts of interest to declare.

Use of AI Services

The authors declare they have not used any AI services to generate or edit any part of the manuscript or data.

Authors' Contribution

Gabi Schaap: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, writing—original draft. **Paul Hendriks Vettehen:** conceptualization, methodology, writing—original draft. **Tibor Bosse:** conceptualization, methodology, writing—original draft.

References

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

Araujo, T., De Vreese, C., Helberger, N., Kruikemeier, S., Van Weert, J., Bol, N., Oberski, D., Pechenizkiy, M., Schaap, G., & Taylor, L. (2018). *Automated decision-making fairness in an AI-driven world: Public perceptions, hopes and concerns.* Digital Communication Methods Lab. https://dare.uva.nl/search?identifier=369fdda8-69f1-4e28-b2c7-ed4ff2f70cf6

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Arkes, H. R. (2008). Being an advocate for linear models of judgment is not an easy life. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (pp. 47–70). Psychology Press.

Aronow, P. M., Baron, J., & Pinson, L. (2019). A note on dropping experimental subjects who fail a manipulation check. *Political Analysis*, *27*(4), 572–589. https://doi.org/10.1017/pan.2019.5

Bandura, A. (2000). Self-efficacy, the foundation of agency. In W. J. Perrig & A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness* (pp. 16–30). Psychology Press.

Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-control as value-based choice. *Current Directions in Psychological Science*, *26*(5), 422–428. https://doi.org/10.1177/0963721417704394

Bhattacherjee, A., & Sanford, C. (2006). Influence processes for information technology acceptance: An elaboration likelihood model. *MIS quarterly*, *30*(4), 805–825. https://www.jstor.org/stable/25148755

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

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

Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally? *The RUSI Journal*, *164*(5–6), 88–96. https://doi.org/10.1080/03071847.2019.1694260

Castelo, N., & Lehmann, D. R. (2019). Be careful what you wish for: Unintended consequences of increasing reliance on technology. *Journal of Marketing Behavior*, *4*(1), 31–42. http://dx.doi.org/10.1561/107.00000059

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Chugunova, M., & Sele, D. (2020). We and it: An interdisciplinary review of the experimental evidence on human-machine interaction. *Center for Law & Economics Working Paper Series*, *99,* Article 101897. https://doi.org/10.1016/j.socec.2022.101897

Coyle, D., Moore, J., Kristensson, P. O., Fletcher, P., & Blackwell, A. (2012). I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2025–2034). https://doi.org/10.1145/2207676.2208350

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

De Freitas, J., Agarwal, S., Schmitt, B., & Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. *Nature Human Behaviour*, *7*(11), 1845–1854. https://doi.org/10.1038/s41562-023-01734-2

Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research 3*(1), 60–95. https://doi.org/10.1287/isre.3.1.60

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177%2F0956797620948841

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Galindo, L., Perset, K., & Sheeka, F. (2021). *An overview of national AI strategies and policies* (OECD Going Digital Toolkit Notes No. 14). OECD Publishing. https://doi.org/10.1787/c05140d9-en

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work, 14*(1), 39–49. https://doi.org/10.1007/s10111-011-0194-3

Groette, O. (2024, April 7). *What percentage of trading is algorithmic? (Algo Trading Market Statistics).* QuantifiedStrategies.com. https://www.quantifiedstrategies.com/what-percentage-of-trading-is-algorithmic/

Gstrein, O. J., Haleem, N., & Zwitter, A. (2024). General-purpose AI regulation and the European Union AI Act. *Internet Policy Review*, *13*(3), 1–26. https://doi.org/10.14763/2024.3.1790

Haggard, P., & Eitam, B. (2015). *The sense of agency*. Oxford University Press.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–2400. https://doi.org/10.1073/pnas.1416587112

Hannon, O., Ciriello, R., & Gal, U. (2024). Just because we can, doesn't mean we should: Algorithm aversion as a principled resistance. In *Proceedings of the 57th Hawaii International Conference on System Sciences* (pp. 6076–6085). https://hdl.handle.net/10125/107115

Himmelstein, M., & Budescu, D. V. (2023). Preference for human or algorithmic forecasting advice does not predict if and how it is used. *Journal of Behavioral Decision Making*, *36*(1), Article e2285. https://doi.org/10.1002/bdm.2285

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*(5), 576–590. https://psycnet.apa.org/doi/10.1037/0033-2909.125.5.576

Jia, H., Wu, M., Jung, E., Shapiro, A., & Sundar, S. S. (2012). Balancing human agency and object agency: An end-user interview study of the internet of things. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 1185–1188). https://doi.org/10.1145/2370216.2370470

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *Research Papers*. 168. https://aisel.aisnet.org/ecis2020_rp/168

Campbell, T., & Safane, J. (2025). *Average Stock Market Return: A Historical Perspective and Future Outlook* https://www.businessinsider.com/personal-finance/investing/average-stock-market-return

Kramer, M. F., Schaich Borg, J., Conitzer, V., & Sinnott-Armstrong, W. (2018). When do people want AI to make decisions? In *Proceedings of the 28th European Conference on Information Systems (ECIS)* (pp. 204–209). Association for Information Systems. https://doi.org/10.1145/3278721.3278752

Langman, S., Capicotto, N., Maddahi, Y., & Zareinia, K. (2021). Roboethics principles and policies in Europe and North America. *SN Applied Sciences*, *3*(12), Article 857. https://doi.org/10.1007/s42452-021-04853-5

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

Logg, J., & Schlund, R. (2024). *A simple explanation reconciles "algorithm aversion" and "algorithm appreciation": Hypotheticals vs. real judgments* (Working Paper No. 4687557). SSRN Electronic Journal. https://dx.doi.org/10.2139/ssrn.4687557

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

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society, 14*(1), 81–95. https://doi.org/10.1007/s10209-014-0348-1

Martens, M., De Wolf, R., & De Marez, L. (2024). Trust in algorithmic decision-making systems in health: A comparison between ADA health and IBM Watson. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *18*(1), Article 5. https://doi.org/10.5817/CP2024-1-5

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, *62*(3), 760–775. https://doi.org/10.1111/ajps.12357

Moore, J. W. (2016). What is the sense of agency and why does it matter? *Frontiers in Psychology*, *7*, Article 1272. https://doi.org/10.3389/fpsyg.2016.01272

Morewedge, C. K. (2022). Preference for human, not algorithm aversion. *Trends in Cognitive Sciences*, *26*(10), 824–826. https://doi.org/10.1016/j.tics.2022.07.007

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

Nussberger, A. M., Luo, L., Celis, L. E., & Crockett, M. J. (2022). Public attitudes value interpretability but prioritize accuracy in artificial intelligence. *Nature Communications*, *13*(1), Article 5821. https://doi.org/10.1038/s41467-022-33417-3

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

Palmeira, M., & Spassova, G. (2015). Consumer reactions to professionals who use decision aids. *European Journal of Marketing*, *49*(3/4), 302–326. https://doi.org/10.1108/EJM-07-2013-0390

Parker, G., Smith, A., & Maynard, J. (1990). Optimality theory in evolutionary biology. *Nature*, *348*(6296), 27–33. https://doi.org/10.1038/348027a0

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pew Research Center (2018, December, 10). *Artificial intelligence and the future of humans.* https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/

Pew Research Center (2020, June 30). *Experts predict more digital innovation by 2030 aimed at enhancing democracy*. https://www.pewresearch.org/internet/2020/06/30/experts-predict-more-digital-innovation-by-2030-aimed-at-enhancing-democracy/

Pew Research Center (2023, February 24). *The future of human agency*. https://www.pewresearch.org/internet/2023/02/24/the-future-of-human-agency/

Pezzo, M. V., & Beckstead, J. W. (2020a). Patients prefer artificial intelligence to a human provider, provided the AI is better than the human: A commentary on Longoni, Bonezzi and Morewedge (2019). *Judgment and Decision Making*, *15*(3), 443–445. https://doi.org/10.1017/S1930297500007221

Pezzo, M. V., & Beckstead, J. W. (2020b). Algorithm aversion is too often presented as though it were non-compensatory: A reply to Longoni et al. (2020). *Judgment and Decision Making*, *15*(3), 449–451. https://doi.org/10.1017/S1930297500007245

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702. https://doi.org/10.1002/for.2464

Rabinovitch, H., Budescu, D. V., & Meyer, Y. B. (2024). Algorithms in selection decisions: Effective, but unappreciated. *Journal of Behavioral Decision Making*, *37*(2), Article e2368. https://doi.org/10.1002/bdm.2368

Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know. *Advances in Motivation Science, 3*, 149–186. https://doi.org/10.1016/bs.adms.2016.02.001

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1–28. https://doi.org/10.1037/h0092976

Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen, J. D. (2006). Neuroeconomics: Cross-currents in research on decision-making. *Trends in Cognitive Sciences*, *10*(3), 108–116. https://doi.org/10.1016/j.tics.2006.01.009

Schaap, G., Bosse, T., & Hendriks Vettehen, P. (2023). The ABC of algorithmic aversion: Not agent, but benefits and control determine the acceptance of automated decision-making. *AI & Society*, *39*(4), 1947–1960. https://doi.org/10.1007/s00146-023-01649-6

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Stein, J. P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior*, *95*, 73–82. https://doi.org/10.1016/j.chb.2019.01.021.

Stevens, J. R. (2008). The evolutionary biology of decision making. In C. Engel & W. Singer (Eds.), *Better than conscious? Decision making, the human mind, and implications for institutions* (pp. 285–304). The MIT Press.

Stroebe, W., van Koningsbruggen, G. M., Papies, E. K., & Aarts, H. (2013). Why most dieters fail but some succeed: A goal conflict model of eating behavior. *Psychological Review, 120*(1), 110–138. https://doi.org/10.1037/a0030849

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of the 2019 CHI Conference on human factors in computing systems* (pp. 1–9). https://doi.org/10.1145/3290605.3300768

Sundar, S. S., & Marathe, S. S. (2010). Personalization vs. customization: The importance of agency, privacy and power usage. *Human Communication Research*, *36*(3), 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

The Royal Society. (2017). *Public views of machine learning: Findings from public research and engagement conducted on behalf of the Royal Society.* Ipsos MORI. https://royalsociety.org/-/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf

Winfield, A. F. (2016). *Written evidence submitted to the UK Parliamentary Select Committee on Science and Technology inquiry on robotics and artificial intelligence* [Evidence submitted to Parliament]. UK Parliament. https://committees.parliament.uk/work/4658/robotics-and-artificial-intelligence-inquiry/

Yeung, K. (2020). Recommendation of the council on artificial intelligence (OECD). *International Legal Materials*, *59*(1), 27–34. https://doi.org/10.1017/ilm.2020.5

Zehnle, M., Hildebrand, C., & Valenzuela, A. (2025). Not all AI is created equal: A meta-analysis revealing drivers of AI resistance across markets, methods, and time. *International Journal of Research in Marketing*. Advance online publication. https://doi.org/10.1016/j.ijresmar.2025.02.005

Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making*, *18*, Article e41. https://doi.org/10.1017/jdm.2023.37

Appendices

Appendix A

Figure A1. *Overview Experimental Design*.

		HUMAN Agent			
		Con− Ben-	**Con− Ben+**	**Con+ Ben−**	**Con+ Ben+**
AI Agent	**Con− Ben-**	1 HNL AINL	2 HNH AINL	3 HYL AINL	4 HYH AINL
	Con− Ben+	5 HNL AINH	6 HNH AINH	7 HYL AINH	8 HYH AINH
	Con+ Ben−	9 HNL AIYL	10 HNH AIYL	11 HYL AIYL	12 HYH AIYL
	Con+ Ben+	13 HNL AIYH	14 HNH AIYH	15 HYL AIYH	16 HYH AIYH

Note. Con=Control: − =no, + =yes. Ben=Benefits: − =low, + =high. H=Human Expert. AI=AI agent. Y=Control, N= No control. H=High benefits, L=Low benefits.

Appendix B: Full ANOVAs

Study 1

Table B1. *The Effect of Control and Benefits on Agent Preference: Stock Market (ANOVA).*

Cases	Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
benefits_AI	289.500	1	289.500	102.968	<.001	.255
Control_AI	15.367	1	15.367	5.466	.020	.014
benefits_human	227.741	1	227.741	81.002	<.001	.201
Control_human	22.934	1	22.934	8.157	.005	.020
benefits_AI * Control_AI	0.099	1	0.099	0.035	.851	<.001
benefits_AI * benefits_human	13.059	1	13.059	4.645	.032	.012
benefits_AI * Control_human	0.018	1	0.018	0.006	.936	<.001
Control_AI * benefits_human	0.331	1	0.331	0.118	.732	<.001
Control_AI * Control_human	<.001	1	<.001	<.001	.986	<.001
benefits_human * Control_human	0.092	1	0.092	0.033	.856	<.001
benefits_AI * Control_AI * benefits_human	3.591	1	3.591	1.277	.260	.003
benefits_AI * Control_AI * Control_human	7.815	1	7.815	2.780	.097	.007
benefits_AI * benefits_human * Control_human	3.338	1	3.338	1.187	.277	.003
Control_AI * benefits_human * Control_human	1.160	1	1.160	0.413	.521	.001
benefits_AI * Control_AI * benefits_human * Control_human	0.199	1	0.199	0.071	.791	<.001
Residuals	548.250	195	2.812			

Note. Type III Sum of Squares.

Study 2

Table B2. *The Effect of Control and Benefits on Agent Preference: Incentivized Stock Market Investment (ANOVA).*

Cases	Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
benefits_AI	220.731	1	220.731	95.709	<.001	.248
Control_AI	4.461	1	4.461	1.934	.166	.005
benefits_human	190.055	1	190.055	82.408	<.001	.214
Control_human	0.644	1	0.644	0.279	.598	<.001
benefits_AI * Control_AI	0.878	1	0.878	0.381	.538	<.001
benefits_AI * benefits_human	1.379	1	1.379	0.598	.440	.002
benefits_AI * Control_human	1.510	1	1.510	0.655	.419	.002
Control_AI * benefits_human	2.874	1	2.874	1.246	.266	.003
Control_AI * Control_human	0.982	1	0.982	0.426	.515	.001
benefits_human * Control_human	3.948	1	3.948	1.712	.192	.004
benefits_AI * Control_AI * benefits_human	<.001	1	<.001	<.001	.987	<.001
benefits_AI * Control_AI * Control_human	1.778	1	1.778	0.771	.381	.002
benefits_AI * benefits_human * Control_human	0.994	1	0.994	0.431	.512	.001
Control_AI * benefits_human * Control_human	4.917	1	4.917	2.132	.146	.006
benefits_AI * Control_AI * benefits_human * Control_human	6.481	1	6.481	2.810	.095	.007
Residuals	447.414	194	2.306			

Note. Type III Sum of Squares.

Study 3

Table B3. *The Effect of Control and Benefits on Agent Preference: Dating (ANOVA)*

Cases	Sum of Squares	*df*	Mean Square	*F*	*p*	η^2
benefits_AI	101.325	1	220.731	41.788	<.001	.101
Control_AI	149.105	1	4.461	61.494	<.001	.148
benefits_human	75.590	1	190.055	31.175	<.001	.075
Control_human	134.579	1	0.644	55.503	<.001	.133
benefits_AI * Control_AI	4.381	1	0.878	1.807	.180	.004
benefits_AI * benefits_human	0.643	1	1.379	0.265	.607	<.001
benefits_AI * Control_human	1.188	1	1.510	0.490	.485	.001
Control_AI * benefits_human	6.031	1	2.874	2.487	.116	.006
Control_AI * Control_human	30.925	1	0.982	12.754	<.001	.031
benefits_human * Control_human	0.249	1	3.948	0.103	.749	<.001
benefits_AI * Control_AI * benefits_human	1.954	1	<.001	0.806	.370	.002
benefits_AI * Control_AI * Control_human	0.725	1	1.778	0.299	.585	<.001
benefits_AI * benefits_human * Control_human	11.161	1	0.994	4.603	.033	.011
Control_AI * benefits_human * Control_human	19.648	1	4.917	8.103	.005	.019
benefits_AI * Control_AI * benefits_human * Control_human	0.188	1	6.481	0.078	.781	.001
Residuals	470.397	194	2.306			

Note. Type III Sum of Squares.

About Authors

Gabi Schaap (PhD, Radboud University), is Assistant Professor at the Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands. He has studied various topics related to the use and acceptance of technology, and journalism and news effects. He is particularly interested in if and when we accept and use Automated Decision Making by AI, and technologies such as second screens, and how they affect us psychologically. He has published on these and other topics in several top-ranked journals.

https://orcid.org/0000-0002-4661-701X

Paul Hendriks Vettehen is an assistant professor in communication science at Radboud University, Netherlands. His work aims at understanding the roles of technology, economics, and psychology in communication processes and their consequences. He has published articles on a variety of topics, including the role of competition in news production, the social character of media use, and the role of content, form, and technology in information processing. These articles have been published in journals like Communication Research, Poetics, Plos One, and Computers in Human Behavior.

https://orcid.org/0000-0001-9628-2476

Tibor Bosse is a Full Professor in the Communication and Media Group at Radboud University. He mainly conducts multi-disciplinary research in the intersection of Computer Science and the Social Sciences. He combines insights from these disciplines in order to design, implement and evaluate 'Social AI' systems, i.e. intelligent computer systems that have the ability to engage in natural social interactions with human beings. Such systems include, among others, social robots, virtual agents and chatbots. From a technical perspective, Bosse develops new algorithms to endow social AI systems with more human-like behavior (e.g., using techniques such as emotion recognition and natural language processing). From a social perspective, he studies the psychological effects of interacting with social AI systems on people's perceptions and behavior. Finally, from an applied perspective, he is interested in the use of social AI systems for various practical purposes, including social skills training and behavior change.

https://orcid.org/0000-0003-4233-0406

Correspondence to
Gabi Schaap, Behavioural Science Institute, Department of Communication & Media, Radboud University
Thomas van Aquinostraat 4, 6525 GD, Nijmegen, Netherlands, gabi.schaap@ru.nl