OPEN FORUM

The ABC of algorithmic aversion: not agent, but benefits and control determine the acceptance of automated decision-making

Gabi Schaap[1] · Tibor Bosse[1] · Paul Hendriks Vettehen[1]

Abstract

While algorithmic decision-making (ADM) is projected to increase exponentially in the coming decades, the academic debate on whether people are ready to accept, trust, and use ADM as opposed to human decision-making is ongoing. The current research aims at reconciling conflicting findings on 'algorithmic aversion' in the literature. It does so by investigating algorithmic aversion while controlling for two important characteristics that are often associated with ADM: increased benefits (monetary and accuracy) and decreased user control. Across three high-powered ($N_{\text{total}} = 1192$), preregistered 2 (agent: algorithm/human) $\times$ 2 (benefits: high/low) $\times$ 2 (control: user control/no control) between-subjects experiments, and two domains (finance and dating), the results were quite consistent: there is little evidence for a default aversion against algorithms and in favor of human decision makers. Instead, users accept or reject decisions and decisional agents based on their predicted benefits and the ability to exercise control over the decision.

Keywords Automated decision-making · Algorithmic aversion · Acceptance · User control · Technology benefits

Algorithmic, or automated, decision-making (ADM) is rapidly becoming ubiquitous in today's society. Aided by big data and advanced machine learning, algorithms are applied to do tasks that were previously exclusively a human prerogative (Jussupow et al. 2020). ADM ranges from professional to more private domains such as financial investment (cf. Lourenço et al. 2020), medical diagnoses (cf. Longoni et al. 2019), dating (cf. Tong et al. 2016), and personal fitness (cf. Busch et al. 2022).

When it comes to decision-making, in some instances algorithms are getting as good as or better than we are (e.g., Cheng et al. 2016; Grove et al. 2000; Kleinberg et al. 2018; O'Toole et al. 2007; Yeomans et al. 2019). However, algorithms have also been criticized for a number of downsides such as bias and fairness, shifting autonomy away from

humans, and other issues (for a brief overview cf. Rahwan et al. 2019). If projections are true that ADM use will increase exponentially in the coming decades and algorithms will continue to improve in many domains (Jussupow et al. 2020; Rahwan et al. 2019), the multi-faceted prospects of ADM raise an important question: How, or in what circumstances, do we want machines as opposed to humans to make consequential decisions in our personal lives?

The answer to this question from recent research is far from clear (Jussupow et al. 2020). Some evidence suggests that users are less inclined to accept, trust, or use algorithmic advice or decisions compared to human advisors. The phenomenon is quite prominent in medical contexts, where algorithms are consistently trusted and used less than human doctors (e.g., Bigman and Gray 2018; Longoni et al. 2019; Promberger and Baron 2006), but is also found in other contexts (for an overview see Jussupow et al. 2020). For instance, Önkal et al. (2009) find a main effect of agent in stock exchange scenarios, where people give greater weight and attention to advice on pricing from a human expert than from a statistical model. Dietvorst et al. (2015, 2018) find that when users see an algorithm make mistakes, they prefer the human forecaster, even when an algorithm is clearly the best overall forecasting option available. They have termed this phenomenon 'algorithmic aversion'.

✉ Gabi Schaap
  gabi.schaap@ru.nl

  Tibor Bosse
  Tibor.bosse@ru.nl

  Paul Hendriks Vettehen
  Paul.hendriksvettehen@ru.nl

1   Department of Communication and Media, Behavioural
    Science Institute, Radboud University, Thomas Van
    Aquinostraat 4, 6525 GD Nijmegen, The Netherlands

Although a majority of studies suggest that algorithmic aversion exists at least in specific contexts and situations (Jussupow et al. 2020), and according to Logg et al. (2019) some have taken this to mean that users have a generalized aversion against ADM, other studies question whether this aversion exists as a default attitude. Logg et al. (2019) for instance, find that 'algorithmic appreciation' is also possible, where people rely more on algorithmic advice than human advice (including their own judgment) in head-to-head comparisons, while Araujo et al. (2020) find that automated decisions are judged on par or better in terms of fairness than decisions taken by human experts. Similarly, one study found that people were willing to disclose more of their personal information and feelings to a virtual agent compared to a human agent (Lucas et al. 2014).

Moreover, on closer inspection, a number of studies do not necessarily suggest a generalized aversion, but instead a fairly equal appreciation of humans and algorithms, only finding a decreased use of algorithms when the algorithm makes mistakes or gives bad advice (cf. Castelo et al. 2019; Dietvorst et al. 2015; Prahl and Van Swol 2017), or when the human decision maker is suddenly replaced by an algorithm (Prahl and Van Swol 2021). Even then, algorithmic aversion is not characterized by outright rejection, but by a *relatively* stronger decrease of use, trust, or acceptance compared to the human agent. Finally, Himmelstein and Budescu (2022) find that people's ratings of the trustworthiness of human and algorithmic agents are not reflected in how persuasive the advice is. In sum, some authors suggest that people may have a default aversion against any and all automated decision makers—as opposed to humans—in any context. However, there is also evidence that users are often quite willing to rely on algorithmic decisions, or that any aversion is dependent on the circumstances. The current research attempts to clarify these contrary outcomes by addressing some of the issues that may have caused them. Below, we list several explanations for the mixed findings, and propose a way to disentangle them. We then focus on the question of whether acceptance of decisions is merely a function of whether the decision is made by a human or algorithmic decisional agent, or whether other factors play a role. To answer it, we will focus on the role of agent (human or algorithmic), the level of beneficial outcomes of the decision (high or low), and the degree of user control (high or low).

At least three factors may account for the diverging findings. A first factor relates to an imbalance in the way human vs. machine decision makers are presented in existing studies. Based on their extensive literature review, Jussupow et al. (2020, p. 11) conclude that the "current literature is inconclusive because researchers often involuntarily use different algorithm and human agent characteristics in their experimental investigations". So, if information is given about the attributes of the agent, it is often unequal between the agents. For instance, sometimes it is explicitly stated that a human agent has certain relevant expertise or experience, whereas the algorithm is portrayed more generally as a computer able to make calculations. Comparing just these two conditions makes it difficult to meaningfully attribute any observed aversion or appreciation among participants to a difference between the two agents. According to Prahl and Van Swol (2021), expertise has one of the strongest and most robust effects on trust in an advisor. Matching the two attributes 'expertise' and 'computational ability' in a factorial design would enable researchers to estimate the extent to which these attributes affect the appreciation or aversion toward the agents. Similarly, when Prahl and Van Swol (2017) offered the same performance information about the human and algorithm, they could not replicate the finding that there is a generalized aversion toward the algorithm.

A related issue in many experimental designs is that little or no information is provided to the participant about the algorithmic or human agent, such as how they work, their level of expertise, or performance levels (Jussupow et al. 2020; Logg et al. 2019). In such cases, no clues at all exist to investigate what caused any observed aversion or appreciation toward the algorithm or human agent. As a result, a host of explanations might account for any result (Jussupow et al. 2020). Research suggests that (information about) the quality of the performance is important. For instance, when participants are provided with information signifying that the algorithm outperforms the human, findings show aversion is reduced (Castelo et al. 2019), or turned into algorithmic appreciation (Bigman and Gray 2018).

A final issue is the domain of judgment (Logg et al. 2019). Of course, in different domains—for instance medical decisions, versus financial decisions, versus dating—different types of outcomes and consequences are relevant, and as a result, different expertise and qualities are expected of the decisional agent. Jussupow et al. (2020) suggest that there may be a fundamental difference between medical contexts and most other contexts, which may explain the consistent finding of aversion in medical settings. Across other domains varying effects are found (e.g., Araujo et al. 2020; Prahl and Van Swol 2021). For instance, Castelo et al. (2019) find that users more often click on online dating advertisements if they imply a human decision maker instead of an algorithmic agent, but do not find the same patterns with financial advertisements. They suggest that users are relatively averse to algorithms in domains that are perceived as being subjective (dating) as opposed to objective (finance). However, their findings also suggest that providing information about the effectiveness of algorithmic decision-making increases the willingness to use them in subjective domains as well.

To conclude, to assess whether a generalizable default algorithmic aversion exists, it is necessary for research designs to level the playing field between human and algorithmic agents.

This can be accomplished by making explicit the potentially relevant attributes of these agents in a particular domain and systematically varying them. In the current research, we apply this strategy. We report on three preregistered experiments ($N_{\text{total}} = 1192$) in which scenarios were presented of decisions made by algorithms across two different high-impact domains of everyday private life: finance (stock market investment) and romantic relationships (dating). In the experiments, we investigate which factors have the biggest impact on acceptance: the nature of the agent (human or algorithmic), the level of beneficial outcomes of the decision (high or low benefits), and the degree of user control (high or low).

1 Main effect of agent

By definition, using automated decision-making hinges on at least two aspects. First, it presupposes delegating one's decision-making authority (at least partially) to a decision-making agent (a machine in the case of automated decision-making). Second, one does this in order to achieve some desired outcome. Following Jussupow et al. (2020), in the current research we assess the existence of algorithmic aversion if we control for (1) the (greater or smaller) *benefits* (outcomes) of the decision and (2) the (higher or lower) degree of user *control* (agency) over the decision-making, while keeping information on agent attributes such as expertise and capabilities equal between the human and algorithm. Moreover, to evaluate the stability of findings, we replicate the study in two domains that are different but in which both benefits and control over the decision-making are expected to matter: financial investment and dating. In addition, we use decision paradigms related to decisions that affect the users themselves, and not abstract others. This three-factor design enables us to assess preferences for machine-made versus human decisions, controlled for the predicted benefits and the potential loss of control over the decision. In other words, it enables us to see whether participants have preferences for machine-made or human decisions if they set aside considerations regarding potential benefits or loss of control. As stated, extant research is inconsistent on the question of algorithmic aversion. Thus, our first research question is:

> RQ1: What (if any) are the main effects of agent (human vs machine) on the acceptance of the agent and its decision?

2 Benefits: optimizing outcomes

Whereas prior research has found an aversion or preference of decisions seemingly based on the human or machine nature of the decisional agent, the current study investigates whether the potential beneficial impact on optimizing outcomes may also play a role in the acceptance of decisions and decision makers. The potential benefits of ADM over human decision-making may lie in optimizing outcomes of decisions, such as offering higher chances of a high material reward, greater accuracy and efficiency (cf. Ghazizadeh, et al. 2012; Venkatesh 2000), or higher levels of quality and reliability of a decision (Lee and See 2004) compared to human decision makers. The question here is whether the potential of a decision maker to yield (more) optimal outcomes than another decision maker determines the acceptance of the decision-making (and decision maker)—irrespective of whether the decision maker is a human or a machine.

A growing body of evidence on theoretical models of technology acceptance, such as the technology acceptance model (TAM; cf. for reviews, cf. Marangunić and Granić, 2015; Yousafzai et al. 2007) and the unified theory of acceptance and use of technology (UTAUT; Venkatesh et al. 2003, 2012; for reviews see Tamilmani et al. 2021; Williams et al. 2015) in their various iterations, and models on the acceptance of automated decisions (Ghazizadeh et al. 2012) consistently points to a technology's perceived utility or performance expectancy at reaching goals as a central predictor of acceptance and use. According to a meta-analysis, perceived utility is the most consistent and strongest predictor of all concepts of usage, intention, and attitude in TAM research (Yousafzai et al. 2007). Although the concept of perceived utility is defined as the belief that a particular technology will enhance job performance (Dawes et al. 1989), it has strong links to outcome expectations (Venkatesh 2000). In addition, cross-sectional ADM research shows that, in terms of benefits, users have a greater preference for computers or humans that exhibit a higher success rate (Kramer et al. 2018). In sum, this suggests that ultimately, the acceptance of machine-made decisions over human decisions may depend on the perceived, expected or real benefits associated with the decision-making agent.

As said, experimental research directly addressing how such benefits affect the acceptance of (both automated and human) decision-making, although scarce, suggests that giving information about an agent's performance affects the acceptance and use of that agent. Several studies indicate that users will prefer the agent that is expected to yield the most optimal outcome (Bigman and Gray 2018; Castelo et al. 2019; Prahl and Van Swol 2017). In the current research, we present participants with a decision that is predicted to reap either high or low monetary benefits or higher or lower accuracy in date matching. If the above reasoning is correct, participants will have a greater acceptance of the agent that offers the greatest such benefits, regardless of whether that agent is human or machine. Therefore, based on the literature review above, we expect that, all else

being equal (amount of information given, decision domain, expertise etc.):

> H1: A high-benefit decision results in a greater acceptance of the decision and the decision-making agent (human or algorithm) than a low-benefit decision.

3 User control

As said, by definition, ADM means delegating at least part of the user's control to a machine (Sundar 2020). Much of the societal debate surrounding automated decision-making centers on the loss of decisional control (e.g., Harari 2016; Williams 2018; Zuboff 2019). Research suggests people value 'control' as one of the primary concerns regarding automated decision-making (Araujo et al. 2018; Stein et al. 2019; The Royal Society 2017). In psychology, one's ability to exert control over the environment and to produce desired results is seen as a basic human need and crucial for a person's functioning (cf. Bandura 1997; Deci and Ryan 2000; Haggard and Eitam 2015; Leotti et al. 2010; Rotter 1966) and is believed to be represented in the way our neural reward systems respond when control is given or withheld (Hassall et al. 2019; Leotti et al. 2015; Samejima 2005).

Later iterations of TAM and UTAUT recognize the importance of control in the acceptance and use of new technology (Marangunić and Granić 2015; Venkatesh 2000). In these models, control is often defined as both internal control—the knowledge, resources and opportunities required to perform specific behavior—and external control—conditions facilitating ease of use, such as availability of support staff (Venkatesh 2000). Although the models primarily define control as indirectly affecting acceptance, via perceived ease of use (cf. Venkatesh 2000), it is also noted that there may be a more direct relation between 'internal' control and behavioral intention or achievement (Marangunić and Granić 2015; Venkatesh 2000). In the context of technology acceptance in automation, Ghazizadeh et al. (2012, p. 40) conclude that "systems that heavily restrict operators' behavior or those that force behavioral changes are less likely to be accepted when compared with nonrestrictive, informative systems."

Although there is to date little research directly addressing how control affects the acceptance of automated and human decision-making alike, based on a comprehensive literature review, Jussupow et al. (2020) conclude that algorithmic aversion is in particular related to the agency of an algorithm. Several recent studies suggest that the ability to influence an automated decision may lead to a greater acceptance of, or preference for, an agent. Users are more likely to rate an advisory algorithm favorably compared to an algorithm that comes up with a decision on its own, without user influence (Palmeira and Spassova 2015), to adopt algorithmic predictions if they can tweak the predictions themselves (Dietvorst et al. 2018), can self-customize media content instead of having it tailored for them (Sundar and Marathe 2010), or experience more satisfaction with algorithmic recommendations on dating partners when they have more than one option to choose from (Tong et al. 2016). Likewise, users of self-driving cars are more accepting of collision warnings than automated control, even when automated control performs better (Inagaki et al. 2007; Navarro et al. 2008; El Jaafari et al. 2008). These findings suggest that differences in the appreciation of ADM may be influenced by the degree to which the user is able to exert control over the final outcome. Therefore, we predict that:

> H2: User control over a decision results in a higher acceptance of the decision and the decision-making agent (human or algorithm) than no user control.

4 Interaction between agent, benefits, and control

One might wonder whether the benefits or the cost of ceding control are felt differently for a human or algorithmic agent. For instance, it could be argued that the benefits are less important in accepting a decision made by an algorithm, because algorithms are supposed to be 'perfect' (Dietvorst et al. 2015). Likewise, it could be that ceding control is less problematic when dealing with a human decision agent, because users feel that a human can be reasoned with or influenced by the user (cf. Dietvorst et al. 2018). However, it is difficult to formulate a priori hypotheses about expected conditional effects because prior research does not provide clear indications of any pattern. Experiments have demonstrated significant interactions of agent and level of expertise (Bigman and Gray 2018, study 5; Madhavan and Wiegmann 2007), and providing performance data and the type of task (objective vs subjective; Castelo et al. 2019), but failed to find interactions in others (cf. Bigman and Gray 2018, study 9). We conclude that the possibility of conditional effects should not be excluded a priori, but also that, because of mixed findings, there is no unambiguous indication for their strength and direction. Thus, our second Research Question is:

> RQ2: How do benefits (high or low) and user control (high or low) interact with the nature of the agent (human or algorithm) to determine the acceptance of the agent and its decision?

5 Method

5.1 Design and preregistration

We conducted three preregistered vignette experiments. The design of all three experiments was the same: a 2 (agent: algorithm/human) × 2 (benefits: high/low) × 2 (control: user control/no control) between-subjects experiment. Participants were randomly assigned to one of eight 'decision scenarios'. In each condition they read the same scenario in which a decision is made regarding their personal life. Afterward they answered questions about their acceptance of the decision and the decisional agent. We used the Prolific Academic survey platform to recruit participants (www.prolific.co). The ethical committee of the authors' institution approved the project under identification code ECSW-2019-169.

All studies were preregistered. Study 1, on automated decision-making in finance, started out with different exploratory hypotheses and research questions than Study 2 and Study 3 (preregistration Study 1 in OSF see here: Study 1). Based on the post hoc analyses of that study included in this report, we developed hypotheses and research questions that were leading for this report. These were preregistered on OSF as Study 2 (a replication of the post hoc findings of Study 1: Study 2), and Study 3 (an application of the design of Study 1 and Study 2 to the domain of personal relationships: Study 3).

5.2 Samples

Our goal was to obtain 0.95 power to detect a medium effect size f of 0.25 at the standard 0.05 alpha error probability. Each experiment included eight treatment groups, with main effects and interactions. Analyses using G*Power (Faul et al. 2007) showed that a sample size of 400 was sufficient to detect this medium effect size.

In all studies, residents from the UK only were eligible for participation. In studies 1 and 2, we recruited adult (> 18 y.o.) participants. In Study 3, we only included 18–50 y.o. participants to minimize unfamiliarity with the phenomenon of dating sites. Table 1 shows that the study samples were generally comparable regarding age, gender, and educational distributions. Participants were paid €1,42 for max. 10 min of participation. After exclusions the final Ns of our studies were 399, 396, and 397 for Studies 1, 2, and 3 respectively, for a total of ($N = 1192$).

5.3 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 four separate parts, each on its own screen, followed by a short summary. After reading the scenario, the participants then answered questions regarding their acceptance of the decision, and of the agent making the decision. A number of exploratory attitude measures were followed by demographics items, and

Table 1 Sample characteristics and descriptives

	Study 1	Study 2	Study 3
Age M (SD)	37.6 (10.96)	32.87 (10.78)	32.97 (7.96)[a]
Gender (%F)	71	73	68
Education (%)			
No formal qualification	0.3	0.3	0.8
Secondary school/GCSE	12.8	9.3	8.1
College/A levels	34.5	33.3	25.2
Undergraduate degree (BA/BSc/other)	37.5	42.7	43.6
Graduate degree (MA/MSc/MPhil/other)	14.0	12.1	19.9
Doctorate degree (PhD/MD/other)	1.0	2.3	2.5
Background in computer science, information technology, AI, or data science (%Y)	11.3	13.1	10.6
Background in finance, such as stock exchange, banking, financial analytics (S1–2) (%Y)	7.0	9.6	–
Prior experience with dating services (S3) (%Y)	–	–	52.9
Acceptance decision M (SD)	4.99 (1.52)	4.68 (1.57)	4.39 (1.59)
Acceptance decision maker M (SD)	4.96 (1.53)	4.99 (1.39)	4.94 (1.40)
N	399	396	397

[a]NB Only ages 18–50 were eligible for participation in Study 3

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.

5.4 Materials

A pilot study ($N = 400$) was used to test measures and materials. Based on this pilot, we slightly adapted the design to ensure that participants spent sufficient time reading the scenario they were assigned to. Participants were asked to imagine they were in the situation described in the scenario. The scenario described how they would be subject to a decision that (a) was made by either an algorithm or human expert. Depending on the condition, they were further told that (b) the benefits for accepting the decision were either great or small, and that (c) they would have either final control or no control over the decision. In all other respects, the scenarios in each condition were the same. To promote processing of the three main features (a, b, and c) of the scenario, we added brief summaries of these features at the end of each scenario.

In Studies 1 and 2, participants read a scenario in which a decision is made regarding the investment of their personal money in the stock market. In Study 3, participants read a scenario in which a decision is made regarding selection of a potential dating partner on a dating website. These scenarios represent two high-impact domains of everyday private life: finance/stock market investment, and romantic relationships/dating.

Both domains share a number of other characteristics. First, both are currently already heavily dependent on automated decision-making and will be doing much more so in the near future. Furthermore, they are decision-making domains that are inherently uncertain, as many impactful real-life decisions are, which may be relevant in the acceptance or rejection of algorithmic decisions (Dietvorst and Bharti 2020). So people are not likely to accept decisions made by others (either by human experts or algorithms) unreflectingly. Third, they are personally relevant domains, with relevant everyday-life impact. In much prior research, algorithmic predictions or decisions have remained rather abstract in terms of their impact on the user, experimental scenarios having for instance a generic 'statistical forecasting model', and tasks involving prediction of mining revenues from planet Zorba (e.g., Dietvorst and Bharti 2020), or guessing the weight of a person on a photograph or the place of a song on the Billboard Hot 100 (Logg et al. 2019). While this research is certainly informative, our research focuses on decision scenarios with a real-life personal (financial or relational) impact on the subject of the decision. In the current research, we use less abstract decision domains, without

venturing into domains of life altering or indeed life-or-death decisions used in for instance medical contexts.

We chose two domains to check the stability of findings, and we selected these specific domains because research shows they represent two domains that likely differ in how they are perceived in terms of objectivity and subjectivity of the decisional task required, with finance likely seen as the more objective, or 'demonstrable' of the two (Castelo et al. 2019; Prahl and Van Swol 2021).

5.5 Manipulations

The three factors were defined as follows (Full texts of the scenarios are in the supplemental materials).

5.5.1 Agent (levels: algorithm or human)

The 'Agent' factor presented the decision as made by either an algorithm or a human expert, both of whom base their decision on the same extensive data. The scenario stated whether the agent making the decision was an algorithm or an 'expert'.

5.5.2 Benefits (levels: high/low)

Benefits were manipulated by predicting either a high (20%) or low (2%) return on investment rate (study 1 and 2), or either a high (92%) or low (57%) match between the participant's and their date's personal profile (Study 3). The average return on investment in the global stock market has been around 10% for the past 10 years (Knueven 2021). Therefore, we used 2% as a low and 20% as a high return in this research. Many dating websites use mathematical algorithms to 'match' their users to a potential partner based on compatibility of characteristics. The assumption is that a highly compatible pair will have a greater chance of positive romantic outcomes (Finkel et al. 2012; Tong et al. 2016). No statistics are available for the average match on dating websites. eHarmony, a well-known US-based dating website, uses an algorithm that provides users with a compatibility score between 60 and 140, with higher scores signifying a greater compatibility with a potential partner, and any score above 100 regarded as high compatibility. Our score of 57% translates to 71 points on the eHarmony scale, and 92% to 115 points.

5.5.3 Control (levels: control/no control)

The 'Control' factor presented the scenario as either an automatic decision that will be executed without the participant's control or a decision that may ultimately be ignored by the user (i.e., an advice). 'No control' was made explicit by telling the participant: "The algorithm/expert will automatically

choose for you. You cannot influence the decision; the algorithm/expert takes it for you." Control was phrased as: "The algorithm/expert will advise which investments to make for you. However, you will have control over the decision; you can choose to follow the advice or not". Finally, participants were told that the decision would be put in motion automatically (no control), or that they "may or may not give permission. It is up to you."

5.6 Measures

5.6.1 Acceptance of decision

To measure acceptance of the decision, we asked: "We want to know how you feel about the decision [regarding the investment of your money in the stock market/to select a date for you]". Answers were measured on a seven-point Likert scale, which was previously used in work on the acceptance of policy decisions (Visschers and Siegrist 2012): "I accept this decision", and "I agree with the decision" (1 = not at all, 7 = completely). The Spearman–Brown coefficient (Eisinga et al. 2013) ranged from 0.924 to 0.933 across the studies.

5.6.2 Acceptance of decision maker

Acceptance of the decision-making agent was measured with two items, based on Technology Acceptance Model research (Venkatesh 2000): "The decision [to invest your money in the stock market/to select a date for you] was made by an [algorithm/expert]. We want to know how you feel about the [algorithm/expert]. Assuming that in real life you would be willing to [invest your money in the stock market/step into online dating], how would you feel about using the [algorithm/expert]?". Responses were measured on a two-item seven-point Likert scale: "Assuming I had access to the system/expert, I intend to use [it/him or her].", and "Given that I had access to the system/expert, I predict that I would use [it/him or her]", [1 = strongly disagree, 7 = strongly agree], Spearman–Brown coefficient = 0.955–0.963.

5.6.3 Attention check

To check whether participants paid attention to the questions, we included an instructed response test, to check for attention (cf. Meade and Craig 2012): "To check your attention, please respond with 'strongly agree' for this item".

5.6.4 Demographics

We asked for demographic characteristics (age, gender, education) at the end of the questionnaire. We also included two further items asking for (1) professional or educational background in computer science, information technology,

Artificial Intelligence (AI), or data science, or related disciplines, and (2) background in finance, such as stock exchange, banking, financial analytics, or related areas (Study 1 and 2), or prior experience with dating websites or applications (Study 3).

5.7 Statistical analysis

5.7.1 Data exclusion

As per the preregistration, participants failing the attention check were excluded ($n = 22$). Also, anyone who completed the questionnaire under 2 min was excluded ($n = 21$). Subsequently, we removed a suspect case with > 1 h of completion time, pointing to technical problems ($n = 1$). In all, we excluded 44 cases.

5.7.2 Outliers

Outliers in the remaining sample with a standardized score > 3 were detected (Study 1, $n = 7$; Study 2, $n = 9$; Study 3, $n = 7$). We ran all hypotheses tests with and without the outlying cases. This produced no differences in the outcomes. Below, analyses with the full samples (including 23 outliers) are reported. The dataset without the outlying cases can be found on osf.

5.7.3 Analysis strategy

As per the preregistration, we used two-way ANOVAs with agent, benefits and control as factors, and with acceptance of the decision and acceptance of the decision maker as dependent variables in two separate models. Significant interactions were explored using comparisons within the moderator variable.

5.8 Data availability

Preregistration of all studies, and data and analyses syntax, and all study materials are available through the following links: Study 1: osf link for study 1), Study 2: osf link for study 2), and Study 3: osf link for study 3).

6 Results

6.1 Randomization check

For each study, we checked randomization by running Chi^2 analyses for gender, educational level, having a professional background in computer science or related disciplines, having a professional or educational background in finance or related sectors (Study 1 and 2), and having used

dating websites or applications (Study 3), and one-way ANOVAs for age with condition as factor. We found no differences between the conditions on any of these variables. We therefore concluded that randomization was successful. Table 1 provides an overview of the descriptive statistics of the main variables.

6.2 Main analyses

To test our hypotheses, we employed two-way ANOVAs. Descriptives of the two dependent variables are in Table 2. For easy comparison, we present a summary of the main effects relating to the hypotheses testing in Table 3. Full ANOVAs are in the supplementary material.

Table 2 Means (standard deviations) per condition per study

	Study 1	Study 2	Study 3
Acceptance of decision			
Human			
Low benefits			
No control	4.24 (1.63)	4.0 (1.76)	3.72 (1.63)
Control	5.04 (1.43)	4.76 (1.54)	4.44 (1.51)
High benefits			
No control	4.73 (1.63)	4.57 (1.60)	4.44 (1.70)
Control	5.56 (1.15)	5.14 (1.46)	5.70 (0.93)
Algorithm			
Low benefits			
No control	4.40 (1.64)	4.13 (1.63)	3.11 (1.34)
Control	5.05 (1.51)	4.86 (1.47)	4.01 (1.40)
High benefits			
No control	4.98 (1.40)	4.59 (1.54)	4.43 (1.40)
Control	5.96 (0.95)	5.43 (0.98)	5.30 (1.13)
Acceptance of decisional agent			
Human			
Low benefits			
No control	4.66 (1.51)	4.92 (1.43)	4.47 (1.62)
Control	4.96 (1.65)	5.23 (1.41)	5.13 (1.36)
High benefits			
No control	4.86 (1.49)	5.22 (1.20)	4.74 (1.57)
Control	5.60 (1.11)	5.49 (1.22)	5.69 (0.76)
Algorithm			
Low benefits			
No control	4.07 (1.68)	4.12 (1.61)	3.23 (1.37)
Control	4.96 (1.49)	5.23 (1.10)	4.97 (1.37)
High benefits			
No control	4.70 (1.57)	4.44 (1.60)	4.70 (1.33)
Control	5.82 (0.96)	5.23 (0.92)	5.56 (1.0)

6.2.1 RQ1: effect of agent

We first assess whether acceptance of a decision and the decisional agent is affected by whether the agent is an algorithm or a human being. In each study, we first checked whether any main effects could be observed. For the three studies, this amounted to 2 (outcomes per study) × 3 (studies) = 6 possible main effects. There does not seem to be a strong case for the idea that users are meaningfully affected by whether the decision is made by a human or algorithm (see Fig. 1). Of the six possible main effects of agent type on the acceptance of the decision or the decision maker, only two turned out to be significant (see Table 3). In Study 3, a small main effect of agent on acceptance of the decision was found: $F(1,396) = 6.78$, $p = 0.010$, $\eta^2 = 0.013$. Participants were slightly more inclined to accept a decision when it was made by a human (M = 4.58, SD = 1.63), compared to a decision by an algorithm (M = 4.20, SD = 1.53). Similarly, in Study 2, participants more readily accepted the decision maker if the decision maker was a human (M = 5.22, SD = 1.32) versus an algorithm (M = 4.75, SD = 1.41), although again, this main effect of agent was weak (Cohen 1988): $F(1,395) = 11.61$, $p = 0.001$, $\eta^2 = 0.027$.

To gauge the strength of the evidence for the null hypotheses, in addition to the preregistered analyses, we conducted post hoc Bayesian ANOVAs (prior r scale = 0.5) for main effects showing a null effect of agent on acceptance. In Study 1, this yielded a Bayes factor of $BF_{01} = 3.655$ for acceptance of the decision, in Study 2 $BF_{01} = 7.070$. For acceptance of the decision maker, the scores were $BF_{01} = 6.492$ for Study 1, and $BF_{01} = 5.318$ for Study 3. Thus, all scores are within the three to ten range, telling us that it is three to seven times more likely for these data to occur under the null model. This can be interpreted as moderate evidence for the null hypothesis (Lee and Wagenmakers 2014).

### 6.2.2 Hypothesis 1: effect of benefits

Our first hypothesis was that a decision with a great benefit results in a greater acceptance of the decision and the decision maker than one with a small benefit. In all three studies, benefits had a significant small to moderate/strong effect on the acceptance of the decision (see Table 3). The means were in line with our expectation that greater expected benefits of the decision lead to a greater willingness to accept the decision (see Fig. 1). The effects were as follows. Study 1: $F(1,398) = 19.03$, $p < 0.001$, $\eta^2 = 0.043$; low benefits M = 4.68 (SD = 1.59), high benefits M = 5.32 (SD = 1.38); Study 2: $F(1,395) = 10.60$, $p = 0.001$, $\eta^2 = 0.025$; low benefits M = 4.44 (SD = 1.64), high benefits M = 4.93 (SD = 1.45); Study 3: $F(1,396) = 65.44$, $p < 0.001$, $\eta^2 = 0.129$; low benefits M = 3.82 (SD = 1.54), high benefits M = 4.96 (SD = 1.42).

Table 3 Summary of main effects per study

Effects		Study 1 ($N=399$)	Study 2 ($N=396$)	Study 3 ($N=397$)
Agent → acceptance decision	F	2.05	0.76	6.78
	p	0.153	0.383	**0.01**
	η^2	0.005	0.002	0.013
Agent → acceptance decision maker	F	0.79	11.61	1.11
	p	0.375	**0.001**	0.292
	η^2	0.002	0.027	0.003
Benefits → acceptance decision	F	19.03	10.60	65.44
	p	**0.000**	**0.001**	**0.000**
	η^2	0.043	0.025	0.129
Benefits → acceptance decision maker	F	16.14	0.2.63	12.59
	p	**0.000**	0.106	**0.000**
	η^2	0.037	0.006	0.029
Control → acceptance decision	F	32.23	22.48	43.51
	p	**0.000**	**0.000**	**0.000**
	η^2	0.072	0.053	0.090
Control → acceptance decision maker	F	27.63	21.28	36.44
	p	**0.000**	**0.000**	**0.000**
	η^2	0.063	0.050	0.083
Df		7391	7388	7389

Bold font indicates significant effects ($p < 0.05$ two sided)
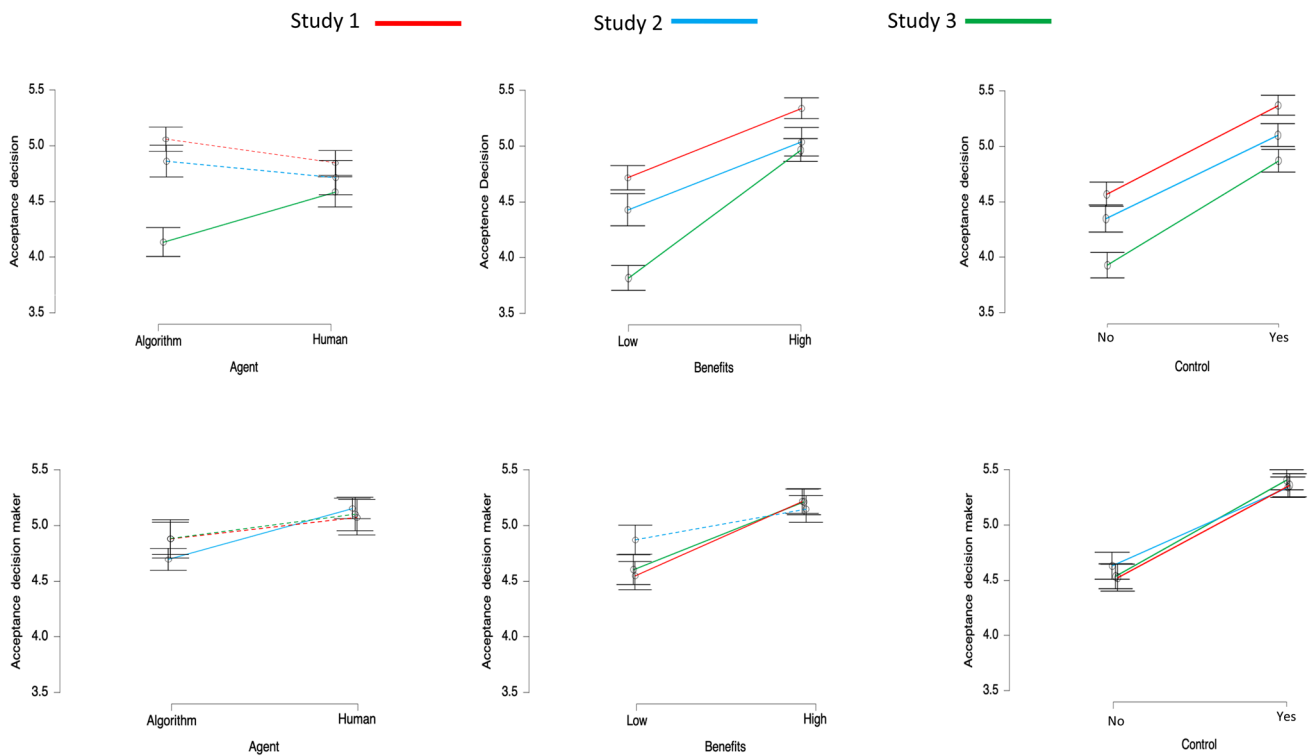
Fig. 1 Main effects of agent, benefits, and control on acceptance of the decision and the decision maker, per study. Error bars show SEs. Dotted lines show non-significant effects

The same pattern was found for the acceptance of the decision-making agent in two out of three studies, with effect sizes indicating small effects: Study 1: $F(1,398) = 16.14$, $p < 0.001$, $\eta^2 = 0.037$; low benefits M = 4.66 (SD = 1.61), high benefits M = 5.26 (SD = 1.37); Study 3: $F(1,396) = 12.59$, $p < 0.001$, $\eta^2 = 0.029$; low benefits M = 4.70 (SD = 1.47), high benefits M = 5.18 (SD = 1.28). However, the effect was not significant in Study 2: $F(1,395) = 2.63$, $p = 0.106$, $\eta^2 = 0.006$; low benefits M = 4.88 (SD = 1.45), high benefits M = 5.10 (SD = 1.32). There were no conditional effects of Benefits in any of the above instances. Overall, we conclude that hypothesis 1 can be accepted.

6.2.3 Hypothesis 2: the effect of control

We hypothesized that when users have final control over a decision, this would result in a higher acceptance of the decision and the decision maker compared to a situation where users have no final say. The analyses confirm that control is an important factor impacting acceptance. In all three studies, control had a moderate and significant effect on both the acceptance of the decision made by the algorithm to invest in stocks or select a date and on the acceptance of the decision-making agent (whether an algorithm or a human) (see Table 3). In each case, the mean values show that participants were more willing to accept the decision or the agent if they had the final say over the decision (see Fig. 1). With regard to acceptance of the decisions the effects were as follows. Study 1: $F(1,398) = 32.23$, $p < 0.001$, $\eta^2 = 0.072$; no control: M = 4.58 (SD = 1.59), control: M = 5.41 (SD = 1.32); Study 2: $F(1,395) = 22.48$, $p < 0.001$, $\eta^2 = 0.053$; no control M = 4.32 (SD = 1.64), control M = 5.04 (SD = 1.40); Study 3: $F(1,396) = 43.51$, $p < 0.001$, $\eta^2 = 0.083$; no control M = 3.93 (SD = 1.61), control M = 4.87 (SD = 1.42). We found no interactions.

With regard to acceptance of the decision-making agent, the effects were again moderate: Study 1: $F(1,398) = 27.63$, $p < 0.001$, $\eta^2 = 0.063$; no control M = 4.57 (SD = 1.58), control M = 5.34 (SD = 1.37); Study 2: $F(1,395) = 21.28$, $p < 0.001$, $\eta^2 = 0.050$; no control M = 4.67 (SD = 1.51), control = 5.30 (SD = 1.17); Study 3: $F(1,396) = 36.44$, $p < 0.001$, $\eta^2 = 0.083$; no control M = 4.54 (SD = 1.48), control M = 5.34 (SD = 1.17). There were no significant conditional effects of Control beyond the weak agent × control effect on acceptance of the decision maker in Study 2 described above. Therefore, we conclude that hypothesis 2 is accepted.

6.2.4 RQ2: interactions

In each study, we also checked whether any conditional effects could be observed in which agent was involved. For the three studies, this amounted to 3 (agent × benefits; agent × control; and agent × benefits × control) possible

conditional effects per outcome × 2 outcomes per study × 3 studies = 18 possible conditional effects. However, we only observed one significant but weak conditional effect, were agent interacted significantly with control to affect acceptance of the decision maker, in Study 2: $F(1,388) = 5.91$, $p = 0.016$, $\eta^2 = 014$. Participants showed a greater acceptance of the human decision maker (M = 5.07, SD = 1.33) compared to the algorithm (M = 4.29, SD = 1.60), $p < 0.001$, in the no control condition, whereas there was no such difference in the control condition (human M = 5.37, SD = 1.31; algorithm M = 5.23, SD = 0.99), $p = 0.489$. No other interaction effects were found (see Supplementary Material). In all, we must conclude that the nature of the agent does not seem an important factor in the acceptance of algorithms and their decisions, neither independently nor conditionally.

7 Conclusion and discussion

While AI-assisted algorithmic decision-making is projected to increase exponentially in the coming decades, the academic debate on whether people are ready to accept, trust, and use ADM as opposed to human decision-making is ongoing. As pointed out by Jussupow et al. (2020), research has produced inconsistent findings on the existence of algorithmic aversion because previous research designs may have prevented to indisputably asses the existence of a default aversion. The current research set out to investigate whether a default aversion to algorithms and their decisions exists, when controlling for two important characteristics that participants otherwise might associate with humans or algorithms, and that thus might explain aversion or appreciation. The characteristics that were controlled for were (monetary and accuracy) benefits and user control, whereas other attributes were kept constant.

Across three high-powered, preregistered studies, including a replication study, two acceptance measures, two domains (finance and romantic relationships), and six possible main effects per experimental factor, the results were quite consistent: there is little evidence for a default aversion against algorithms and in favor of human decision makers. Users seem fairly unconcerned when decision-making is done by machines instead of humans. Instead, they accept or reject decisions and decisional agents based on their predicted monetary and accuracy benefits and the level of control they can exercise over the decision. All effects of control were significant, and generally moderate in size. They all lead to the same conclusion: that acceptance of decisions and the agent making the decisions is lower when people do not have decisional control. Furthermore, although generally the effects are somewhat weaker, five out of six effects for benefits were significant, with high benefits leading to a greater acceptance of both the decision and the agent making

the decisions. In contrast, only two effects relating to the agent making the decision (human or algorithm) were significant, and the effect sizes were small. Finally, these effects are generally unconditional.

In sum, when accepting automated decisions on everyday private life and their agents, people weigh the extent to which they preserve control over a decision and the predicted benefits of the decision. If these two criteria are held equal, along with a number of other agent attributes, it does not matter much to users whether a decision is made by a machine or a human being.

From prior research it was undetermined whether or not people are averse to machine-made decisions compared to human decisions (Jussupow et al. 2020). This study contributes to this line of research by employing a three-factorial design that for the first time allowed controlling for the impact of two important factors that are often associated with human vs machine decisions: benefits of the decision and ultimate control over the decision. Our findings are contrary to previous work that found that users have a preference for algorithms over human agents (e.g., Logg, et al. 2019). But although we do find some instances where a slight preference for human decisions was detected, pointing perhaps to algorithmic aversion, the evidence for algorithmic aversion is weak. Moreover, the Bayesian analysis revealed evidence for the null hypothesis that there were no differences in preference for human and algorithmic agents. Our findings show that when controlled for benefits and control, it does not seem to be the nature of the decision maker per se that is most important, but rather the expected benefits of and degree of user control over the decisions, regardless of the life domain in which they operate. This concurs with research showing that users appreciate algorithmic decisions when they are provided with information signifying that the algorithm outperforms the human agent (Bigman and Gray 2018).

These results provide a new perspective on the inconsistent findings regarding algorithm aversion or preference in earlier research (Araujo et al. 2020; Castelo et al. 2019; Prahl and Van Swol 2021). Jussupow et al. (2020) argue that results from prior research are inconsistent because it has sometimes manipulated different algorithm and human agent characteristics in their experimental investigations, or has omitted relevant information about the agents. The current study sought to alleviate these issues. By keeping multiple factors constant or statistically controlled for, while giving relevant and equal information about the expertise of the agents (cf. Prahl and Van Swol 2021), our study offers an opening for a reconciliation of the aversion vs appreciation views of ADM, by suggesting there may not be a 'generalized' attitude in either direction, and that aversion or appreciation simply may depend on the pros and cons of a decision.

Relatedly, it is interesting to see that in some prior research the deck is possibly stacked against the machine. Oftentimes, if information is given at all about the algorithm (beyond: 'it is a computer/software that is very good at calculating') it seems to emphasize what the algorithm is *not* (e.g., a person, or doctor), or is *in*capable of doing (e.g., empathizing), or at the least not giving fair account that what is often needed for the tasks involved is precisely being very good at calculating. For instance, in medical decisions, it may not be obvious to the average study participant whether the decisions involved (and used as cases in the experiments to test aversion) are at least partially based on being able to do 'objective' analyses or calculations (e.g., Bigman and Gray 2018; Longoni et al. 2019; Promberger and Baron 2006). This does not immediately imply that when users are informed of this, they might change their aversion to medical algorithms, much less that they should, given the current issues with algorithmic medical decision-making (e.g., Grote and Keeling 2022). But it does perhaps partially explain extant findings on aversion to (medical) algorithms. What is more, our findings concur with a number of previous studies showing that when users are informed about the superiority of algorithmic decisions, they prefer them to humans (Bigman and Gray 2018; Castelo et al. 2019; Prahl and Van Swol 2017). This suggests that in order to get a more final answer on the existence of aversion in any domain or context, future research should ensure that enough relevant information is provided on the functioning of the decision maker be they human or machine.

The acceptance of decision-making technology as a function of the concrete benefits it has to offer (compared to humans) has not been extensively studied. However, there is ample evidence from psychological and animal behavior research that concrete benefits affect choice behavior (Karsh and Eitam 2015; Samejima 2005). Research in the technology acceptance model tradition (Venkatesh 2000 Yousafzai et al. 2007), and models on the acceptance of automated decisions (Ghazizadeh et al. 2012), pointing to the fact that a greater expected utility or performance leads to easier acceptance of a technology and may be seen in a similar vein. Our research is in line with these prior notions, but adds to them the direct test of the effect of concrete benefits on acceptance of an automated decision and its maker that have not been previously studied. Furthermore, our research suggests, although ADM technology is new and different, in this respect it may be 'just another technology'. People will use it if the benefits outweigh the costs.

Our findings regarding the role of control correspond to a small number of studies which indirectly implicate control as an important factor. These studies found that people are more appreciative of algorithms when they have a more advisory role or offer choice (Palmeira and Spassova 2015; Tong et al. 2016), or when their decisions or attributes can

be modified (Dietvorst et al. 2018; Sundar and Marathe 2010). With our explicit test of the role of control, these earlier findings could be reinterpreted as having in common 'control' as the crucial factor.

Although at first glance, the effects of benefits and control may seem obvious to some, they are not necessarily so. First, as indicated above, much research suggests that users are averse to automated decisions, even if these produce better results, whereas our research—in line with other research—leads to other conclusions. Second, although it seems logical to prefer decisions and decisional agents that bring the greatest benefits and control, it is less clear beforehand how they would interact to affect acceptance, and which of the two would be more important. Moreover, it is surprising that we did not find interactions with the agent factor, as it would not be unreasonable to expect that for instance keeping control is more important when dealing with a machine than it is with a human being.

We must be cautious generalizing our findings. First, we employed a vignette paradigm, which tries to induce in participants the sense that they are in a real situation. However, the situations, and consequently their corresponding impact and behavior are in fact hypothetical. A meta-analysis of TAM research did find a strong correlation between behavioral intentions an actual technology usage, which may ease some of these reservations; however, the authors note that only a small number of TAM studies have included actual behavior as independent variable, so caution remains advised (Yousafzai et al. 2007). Second, our findings were consistent across two different everyday personal life domains. While it may not seem unwarranted to extrapolate the findings to other domains, such as financial loans, leisure, commodities consumption, and psychological and physical well-being, caution is required. Especially in domains involving overt moral judgments and life and death decisions such as in medicine, different mechanisms may be at work (Bigman and Gray 2018; Hidalgo et al. 2021; Longoni et al 2019). Future research may want to explore the role of various domains and the related advantages and disadvantages of using ADM in acceptance. Each domain may come with its own primary costs and benefits. For instance, in the medical domain users may judge the capability to show human-like qualities such as empathy highly important, whereas in a financial domain objectivity is more relevant. Likewise, the costs of (non-)acceptance of a decision (-maker) may differ across domains.

Furthermore, our experiments used 'faceless' algorithms (and humans) as decisional agents. It is uncertain whether our claims that the human–machine dimension is of little consequence in acceptance are equally true for decision-making technology where the agent has a more tangible, or even human(oid) appearance, as is the case when people deal with virtual agents or (chat)bots. In such cases issues such as similarity (e.g., in behavior) may play a role (e.g., Bernier and Scassellati 2010; You and Robert 2018). However, currently most everyday encounters with AI and algorithms probably are of the faceless, more 'abstract' kind. Finally, we have operationalized the concrete benefits in terms of revenue and accuracy. Future research may look into other types of concrete benefits associated with machine decisions, such as greater (monetary, effort, or time) efficiency, convenience, or fun, to see if they produce similar results.

Overall, our findings teach us two important things about how people now and in the future might accept or reject decisions by intelligent machines. First, that there does not seem to be a default aversion against, or appreciation for automated decision-making. Second, and equally important, users incorporate the benefits *and* the costs (control loss) when evaluating ADM technology. We conclude that when people are well-informed about the costs and benefits of algorithmic decision-making, neither aversion nor appreciation are likely. As algorithms will continue to increase their decision-making advantage over humans (possibly with the inclusion of fields usually not associated with them, such as emotions, empathy, and ethics), it would be fair to account for this in future research.

Declarations

References

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. Research Report

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

Bandura A (1997) Self-efficacy: the exercise of control. Freeman, New York

Bernier EP, Scassellati B (2010) The similarity-attraction effect in human-robot interaction. In: 2010 IEEE 9th international conference on development and learning. IEEE, pp 286–290

Bigman YE, 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

Busch L, Utesch T, Strauss B (2022) Normalised step targets in fitness apps affect users' autonomy need satisfaction, motivation and physical activity–a six-week RCT. Int J Sport Exerc Psychol 20(1):223–244

Castelo N, Bos MW, Lehmann DR (2019) Task-dependent algorithm aversion. J Mark Res 56(5):809–825. https://doi.org/10.1177/0022243719851788

Cheng JZ, Ni D, Chou YH et al (2016) Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Rep 6(1):1-13. https://doi.org/10.1038/srep24454

Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Erlbaum, Hillsdale

Dawes RM, Faust D, Meehl PE (1989) Clinical versus actuarial judgment. Science 243(4899):1668–1674. https://doi.org/10.1126/science.2648573

Deci EL, Ryan RM (2000) The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. Psychol Inq 11(4):227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dietvorst BJ, Bharti S (2020) People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. Psychol Sci 31(10):1302–1314. https://doi.org/10.1177/0956797620948841

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 144(1):114. https://doi.org/10.1037/xge0000033

Dietvorst BJ, Simmons JP, Massey C (2018) Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. Manage Sci 64(3):1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Eisinga R, Te Grotenhuis M, Pelzer B (2013) The reliability of a two-item scale: Pearson, Cronbach, or Spearman–Brown? Int J Public Health 58(4):637–642. https://doi.org/10.1007/s00038-012-0416-3

El Jaafari M, Forzy JF, Navarro J, Mars F, Hoc JM (2008) User acceptance and effectiveness of warning and motor priming assistance devices in car driving. In: Proceedings of European conference on human centred design for intelligent transport systems, p 311

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. Behav Res Methods 39:175–191. https://doi.org/10.3758/BF03193146

Finkel EJ, Eastwick PW, Karney BR, Reis HT, Sprecher S (2012) Online dating: a critical analysis from the perspective of psychological science. Psychol Sci Public Interest 13(1):3–66

Ghazizadeh M, Lee JD, Boyle LN (2012) Extending the Technology Acceptance Model to assess automation. Cogn Technol Work 14(1):39–49. https://doi.org/10.1007/s10111-011-0194-3

Grote T, Keeling G (2022) On algorithmic fairness in medical practice. Camb Q Healthc Ethics 31(1):83–94

Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C (2000) Clinical versus mechanical prediction: a meta-analysis. Psychol Assess 12:19–30. https://doi.org/10.1037/1040-3590.12.1.19

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

Harari YN (2016) Homo deus: a brief history of tomorrow. Harvill Secker, London

Hassall CD, Hajcak G, Krigolson OE (2019) The importance of agency in human reward processing. Cogn Affect Behav Neurosci 19(6):1458–1466. https://doi.org/10.3758/s13415-019-00730-2

Hidalgo CA, Orghiain D, Canals JA, De Almeida F, Martín N (2021) How humans judge machines. MIT Press, Cambridge

Himmelstein M, Budescu DV (2022) Preference for human or algorithmic forecasting advice does not predict if and how it is used. J Behav Decis Mak 36(1):e2285. https://doi.org/10.1002/bdm.2285

Inagaki T, Itoh M, Nagai Y (2007) Support by warning or by action: which is appropriate under mismatches between driver intent and traffic conditions? IEICE Trans Fundam Electron Commun Comput Sci 90(11):2540–2545

Jussupow E, Benbasat I, Heinzl A (2020) Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In: Proceedings of the 28th European conference on information systems (ECIS), an online AIS conference, June 15–17, 2020. https://aisel.aisnet.org/ecis2020_rp/168

Karsh N, Eitam B (2015) I control therefore I do: Judgments of agency influence action selection. Cognition 138:122–131. https://doi.org/10.1016/j.cognition.2015.02.002

Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2018) Human decisions and machine predictions. Q J Econ 133(1):237–293

Knueven L (2021) The average stock market return over the past 10 years. https://www.businessinsider.com/personal-finance/average-stock-market-return?international=true&r=US&IR=T. Accessed 9 July 2021

Kramer MF, Schaich Borg J, Conitzer V, Sinnott-Armstrong W (2018) When do people want AI to make decisions? In: Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society, pp 204–209

Lee JD, See KA (2004) Trust in automation: Designing for appropriate reliance. J Hum Factors Ergon Soc 46(1):50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee MD, Wagenmakers EJ (2014) Bayesian cognitive modeling: A practical course. Cambridge University Press, Cambridge

Leotti LA, Iyengar SS, Ochsner KN (2010) Born to choose: the origins and value of the need for control. Trends Cogn Sci 14(10):457–463

Leotti LA, Cho C, Delgado MR (2015) The neural basis underlying the experience of control in the human brain. In: Haggard P, Eitam B (eds) The sense of agency. Oxford University Press, Oxford, pp 145–169

Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process 151:90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to medical artificial intelligence. J Consum Res 46(4):629–650. https://doi.org/10.1093/jcr/ucz013

Lourenço CJ, Dellaert BG, Donkers B (2020) Whose algorithm says so: the relationships between type of firm, perceptions of trust and expertise, and the acceptance of financial Robo-advice. J Interact Mark 49:107–124

Lucas GM, Gratch J, King AA, Morency LP (2014) It's only a computer: the impact of human-agent interaction in clinical interviews. In: Proceedings of the 2014 international conference on autonomous agents and multi-agent systems, pp 85–92

Marangunić N, Granić A (2015) Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 14(1):81–95. https://doi.org/10.1007/s10209-014-0348-1

Meade AW, Craig SB (2012) Identifying careless responses in survey data. Psychol Methods 17(3):437–455. https://doi.org/10.1037/a0028085

Navarro J, Mars F, Forzy JF, El-Jaafari M, Hoc JM, Renault G (2008) Objective and subjective assessment of warning and motor priming assistance devices in car driving. In: de Waard D (eds) Human Factors for assistance and automation. Shaker, pp 273–283

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. J Behav Decis Mak 22(4):390–409. https://doi.org/10.1002/bdm.637

O'Toole AJ, Phillips PJ, Jiang F, Ayyad J, Penard N, Abdi H (2007) Face recognition algorithms surpass humans matching faces over changes in illumination. IEEE Trans Pattern Anal Mach Intell 29(9):1642–1646

Palmeira M, Spassova G (2015) Consumer reactions to professionals who use decision aids. Eur J Mark 49(3–4):302–326. https://doi.org/10.1108/EJM-07-2013-0390

Prahl A, Van Swol L (2017) Understanding algorithm aversion: When is advice from automation discounted? J Forecast 36(6):691–702

Prahl A, Van Swol LM (2021) Out with the humans, in with the machines? Investigating the behavioral and psychological effects of replacing human advisors with a machine. Hum Mach Commun 2:209–234

Promberger M, Baron J (2006) Do patients trust computers? J Behav Decis Mak 19(5):455–468. https://doi.org/10.1002/bdm.542

Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon JF, Breazeal C et al (2019) Machine behaviour. Nature 568(7753):477–486. https://doi.org/10.1038/s41586-019-1138-y

Rotter JB (1966) Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr Gen Appl 80(1):1–28. https://doi.org/10.1037/h0092976

Samejima K (2005) Representation of action-specific benefit values in the striatum. Science 310:1337–1340. https://doi.org/10.1126/science.1115270

Stein JP, Liebold B, Ohler P (2019) Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. Comput Hum Behav 95:73–82. https://doi.org/10.1016/j.chb.2019.01.021

Sundar SS (2020) Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). J Comput-Mediat Commun 25(1):74–88. https://doi.org/10.1093/jcmc/zmz026

Sundar SS, Marathe SS (2010) Personalization vs. customization: the importance of agency, privacy and power usage. Hum Commun Res 36:298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

Tamilmani K, Rana NP, Wamba SF, Dwivedi R (2021) The extended unified theory of acceptance and use of technology (UTAUT2): a systematic literature review and theory evaluation. Int J Inf Manag 57:102269. https://doi.org/10.1016/j.ijinfomgt.2020.102269

The Royal Society & Ipsos MORI (2017) Machine learning: what do the public think? https://royalsociety.org/-/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf. Accessed 23 Mar 2021

Tong ST, Hancock JT, Slatcher RB (2016) Online dating system design and relational decision-making: choice, algorithms, and control. Pers Relat 23(4):645–662. https://doi.org/10.1111/pere.12158

Venkatesh V (2000) Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf Syst Res 11(4):342–365. https://doi.org/10.1287/isre.11.4.342.11872

Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27(3):425–478

Venkatesh V, Thong JY, Xu X (2012) Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 36(1):157–178

Visschers VH, Siegrist M (2012) Fair play in energy policy decisions: procedural fairness, outcome fairness and acceptance of the decision to rebuild nuclear power plants. Energy Policy 46:292–300. https://doi.org/10.1016/j.enpol.2012.03.062

Williams J (2018) Stand out of our light: Freedom and resistance in the attention economy. Cambridge University Press, Cambridge

Williams MD, Rana NP, Dwivedi YK (2015) The unified theory of acceptance and use of technology (UTAUT): a literature review. J Enterp Inf Manag 28(3):443–488. https://doi.org/10.1108/JEIM-09-2014-0088

Yeomans M, Shah A, Mullainathan S, Kleinberg J (2019) Making sense of recommendations. J Behav Decis Mak 32(4):403–414. https://doi.org/10.1002/bdm.2118

You S, Robert LP (2018) Human–robot similarity and willingness to work with a robotic co-worker. In: 2018 13th ACM/ieee international conference on human-robot interaction (HRI). IEEE, pp 251–260

Yousafzai SY, Foxall GR, Pallister JG (2007) Technology acceptance: a meta-analysis of the TAM: part 2. J Model Manag 2(3):281–304. https://doi.org/10.1108/17465660710834462

Zuboff S (2019) The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile Books, London