

Gill, Andrej; Gillenkirch, Robert M.; Ortner, Julia; Velthuis, Louis

**Article — Published Version**

## Dynamics of Reliance on Algorithmic Advice

Journal of Behavioral Decision Making

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Gill, Andrej; Gillenkirch, Robert M.; Ortner, Julia; Velthuis, Louis (2024) : Dynamics of Reliance on Algorithmic Advice, Journal of Behavioral Decision Making, ISSN 1099-0771, Wiley, Hoboken, NJ, Vol. 37, Iss. 4, <https://doi.org/10.1002/bdm.2414>

This Version is available at:

<https://hdl.handle.net/10419/305825>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*


*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

## RESEARCH ARTICLE OPEN ACCESS

# Dynamics of Reliance on Algorithmic Advice

Andrej Gill<sup>1</sup> | Robert M. Gillenkirch<sup>2</sup> | Julia Ortner<sup>1</sup>  | Louis Velthuis<sup>1</sup><sup>1</sup>Faculty of Law, Management and Economics, University of Mainz, Mainz, Germany | <sup>2</sup>Faculty of Business and Economics, University of Osnabrueck, Osnabrück, Germany**Correspondence:** Julia Ortner ([julia.ortner@uni-mainz.de](mailto:julia.ortner@uni-mainz.de))**Received:** 20 September 2023 | **Revised:** 6 September 2024 | **Accepted:** 9 September 2024**Keywords:** algorithm aversion | algorithmic advice | decision support system | strategic interaction | technology acceptance | trust

## ABSTRACT

This study examines the dynamics of human reliance on algorithmic advice in a situation with strategic interaction. Participants played the strategic game of Rock–Paper–Scissors (RPS) under various conditions, receiving algorithmic decision support while facing human or algorithmic opponents. Results indicate that participants often underutilize algorithmic recommendations, particularly after early errors, but increasingly rely on the algorithm following successful early predictions. This behavior demonstrates a sensitivity to decision outcomes, with asymmetry: rejecting advice consistently reinforces rejecting advice again while accepting advice leads to varied reactions based on outcomes. We also investigate how personal characteristics, such as algorithm familiarity and domain experience, influence reliance on algorithmic advice. Both factors positively correlate with increased reliance, and algorithm familiarity significantly moderates the relationship between outcome feedback and reliance. Facing an algorithmic opponent increases advice rejection frequencies, and the determinants of trust and interaction dynamics differ from those with human opponents. Our findings enhance the understanding of algorithm aversion and reliance on AI, suggesting that increasing familiarity with algorithms can improve their integration into decision-making processes.

## 1 | Introduction

The field of machine learning and its rapid advancements in artificial intelligence (AI) have garnered significant attention in recent years (Littman et al. 2021). Machine learning focuses on generating knowledge from experience through the use of learning algorithms to develop complex models. Learning algorithms are poised to increasingly produce decision-relevant results. Consequently, organizations have increasingly employed algorithms over the past years to create business value. Despite this, decisions in many domains are unlikely to be fully delegated to algorithms. Instead, human decision-makers will interact with these algorithms, retaining the final decision rights. In these scenarios of human–machine decision-making, it is crucial to understand the conditions under which human decision-makers adhere to algorithmic advice (Burton, Stein, and Jensen 2020; Glikson and Woolley 2020). The decision to accept or reject algorithmic advice depends on factors inherent to both the individual and

the algorithm, as well as the dynamics of human–machine interaction. Dietvorst, Simmons, and Massey (2015), in their seminal study on algorithm aversion, highlight a dynamic phenomenon where decision-makers initially rely on an algorithm but lose trust in it after witnessing errors. Our exploratory study investigates the dynamics of human–algorithm interaction in that it analyzes determinants of reliance on algorithmic advice in a situation where individuals repeatedly interact with an algorithm and receive immediate outcome feedback after each decision to accept versus reject the algorithm's advice. Although most previous studies have focused on situations where the algorithm's task is to estimate an unknown fact or forecast an uncertain event, this study models a scenario of strategic interaction where decision-makers try to make accurate predictions of their strategic counterparts' behaviors in order to gain an advantage. Investigating such a scenario is crucial to deepen our understanding of human–algorithm interaction for several reasons. As the range of tasks AI can perform rapidly expands, it is essential to broaden research by exploring

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Journal of Behavioral Decision Making* published by John Wiley & Sons Ltd.

new decision domains (Glikson and Woolley 2020). Strategic interaction is one such domain. It is prevalent within markets and organizations, characterized by its repetitive and dynamic nature. Moreover, strategic interaction necessitates algorithms that extend beyond statistical models, which have been the primary focus of early research but are no longer representative of AI applications in many decision domains. In strategic interaction, a successful algorithm must predict the choices of its strategic counterpart, requiring it to be a sufficiently sophisticated learning algorithm that identifies patterns in the counterpart's behavior and continuously updates its predictions during the interaction.

Our analysis is guided by a framework that builds upon theoretical research on trust in automation and AI (Glikson and Woolley 2020; Hoff and Bashir 2015; Marsh and Dibben 2003) and focuses on three research questions we address. First, we aim to understand how humans dynamically interact with an algorithm in a strategic setting, specifically examining how an individual's decision to follow the algorithm, combined with the feedback the individual gets by observing the outcome of that decision, influences subsequent behavior. Second, we investigate whether prior evidence regarding the determinants of individuals' basic willingness to trust algorithms extends to the decision domain we explore. Additionally, we examine whether factors such as participants' familiarity with algorithms, experience with the decision domain and task, and other personal characteristics moderate the dynamics of human–algorithm interaction. Third, we assess whether the findings related to the first two research questions are unique to scenarios where the individual receiving decision support faces a human counterpart or if they also apply to situations where the counterpart is an algorithm.

To investigate these research questions, we conducted a controlled, incentivized experiment in which participants engaged in the strategic game of Rock–Paper–Scissors (RPS) over an extensive number of rounds. RPS is characterized by its symmetric structure and straightforward rules: Rock defeats scissors but is defeated by paper; scissors defeat paper but are defeated by rock; and paper defeats rock but is defeated by scissors. During the experiment, participants played RPS against either a human or an algorithm. They were provided with decision support from a heuristic reinforcement learning algorithm to guide their choices throughout the game. All three research questions refer to the determinants of human decision-makers' reliance on algorithmic advice (Burton, Stein, and Jensen 2020; Jussupow, Benbasat, and Heinzl 2020). Theoretical frameworks (Glikson and Woolley 2020; Hoff and Bashir 2015) categorize these determinants into factors that influence an individual's basic willingness to rely on an algorithm in a given situation and factors that determine the dynamics of human–algorithm interaction. To position this study, we review the research streams addressing both sets of determinants, with a particular focus on dynamic human–algorithm interaction.

## 1.1 | Dynamic Human-Algorithm Interaction

Building on early evidence of algorithm aversion (Dzindolet et al. 2003; Önköl et al. 2009), recent research has contributed to building theory about drivers of human trust in algorithms. In an influential study, Dietvorst, Simmons, and Massey (2015)

identified algorithm aversion as a dynamic phenomenon, showing that decision-makers lose trust in an algorithm after experiencing it making erroneous forecasts. Participants of their experiments are averse to algorithms in that they prefer to not delegate a forecasting task to an algorithm, and prefer to delegate the task to another human rather than to an algorithm, even after experiencing the algorithm to be more accurate than both their own and the other human's forecasts. Despite the influence of Dietvorst, Simmons, and Massey (2015), few subsequent experimental studies have replicated its dynamic setting. Prah and Van Swol (2017) manipulate the source of advice (human expert versus algorithm) and observe that after a round where advice turns out to be poor, the weight decision-makers give to algorithmic advice declines considerably more sharply than the weight they give to human advice. Similarly, Chacon, Kausel, and Reyes (2022) find that reliance on algorithmic advice in financial forecasting decreases more rapidly than reliance on human advice after witnessing algorithmic errors. Conversely, other studies (Bastani, Bastani, and Sinchaisri 2021; Berger et al. 2021; Filiz et al. 2021; Snyder, Keppler, and Leider 2022; Yin, Wortman Vaughan, and Wallach 2019; Zhang 2023) demonstrate an increase in reliance on algorithms when participants experience accurate algorithmic predictions. Although these studies focus on humans' evaluations of the algorithms they receive advice from, Chong et al. (2022) emphasize the role of self-confidence, showing that not only a decision-maker's confidence in the algorithm but also in herself decreases after inaccurate advice from the algorithm.

Field studies have further explored human–algorithm interaction in practical settings such as retail product planning and pricing (Caro and de Tejada Cuenca 2023; Kawaguchi 2021) and taxi positioning (Liu et al. 2023). However, few provide direct evidence of reactions to algorithmic recommendations. Liu et al. (2022) find that store managers maintain their initial attitudes toward an algorithm even after positive outcomes. Similarly, Lin, Kim, and Tong (2022) observe that diabetes patients' reliance on algorithmic advice increases following their own mistakes but not after the algorithm's errors.

Our study extends this research into a new decision domain—strategic interaction. We investigate whether human decision-makers show negative reactions to algorithm errors but also whether participants increase their reliance on the algorithm when they (repeatedly) receive helpful advice. Both observations—losing trust after seeing an algorithm err (Dietvorst, Simmons, and Massey 2015; Merritt et al. 2015; Prah and Van Swol 2017) and learning to rely on an algorithm (Bastani, Bastani, and Sinchaisri 2021; Berger et al. 2021; Filiz et al. 2021; Yin, Wortman Vaughan, and Wallach 2019) have been made in prior studies. In contrast to prior studies, though, our design allows us to capture reactions to correct and incorrect predictions made by both the algorithm and the human decision-maker herself.

## 1.2 | Reliance on Algorithms in Forecasting and in Strategic Interaction

This study also relates to research on determinants of an individual's basic willingness to rely on algorithms. Theoretical

frameworks (e.g., Hoff and Bashir 2015; Mahmud et al. 2022) have categorized determinants of trust in algorithms into sets referring to the characteristics of the person, the algorithm (e.g., its opaqueness and explainability; Gunning et al. 2019), the decision domain (e.g., business vs. medical decisions, objective vs. subjective judgments; Castelo, Bos, and Lehmann 2019; Lee 2018), and the organizational and social environment (e.g., how algorithm use affects perception by others or replaces human roles; Bigman and Gray 2018; Prahll and Van Swol 2021; Venkatesh and Davis 2000). We collected data on participants' personal characteristics, including familiarity with algorithms and domain and task experience, and analyzed their impact on reliance on a machine learning algorithm in strategic interaction. These determinants belong to the within the decision-maker category; studies investigating determinants from the other categories are reviewed, for example, in Cabiddu et al. (2022), De Freitas et al. (2023), or Mahmud et al. (2022). Research indicates that familiarity with algorithms positively affects reliance on algorithmic advice (Castelo, Bos, and Lehmann 2019; Logg, Minson, and Moore 2019; Whitecotton 1996), whereas domain experience has a negative (Kim et al. 2022; Logg, Minson, and Moore 2019) or inversely U-shaped effect (Allen and Choudhury 2022; Luo et al. 2021). Typically, studies investigating these determinants of reliance on algorithmic advice model tasks that involve predicting uncertain events or making judgments about unknown facts, and they show that reliance on algorithms varies by decision domain (Castelo, Bos, and Lehmann 2019; Himmelstein and Budescu 2023) and is lower for more subjective tasks (Commerford et al. 2022; Logg, Minson, and Moore 2019; Yeomans et al. 2019).

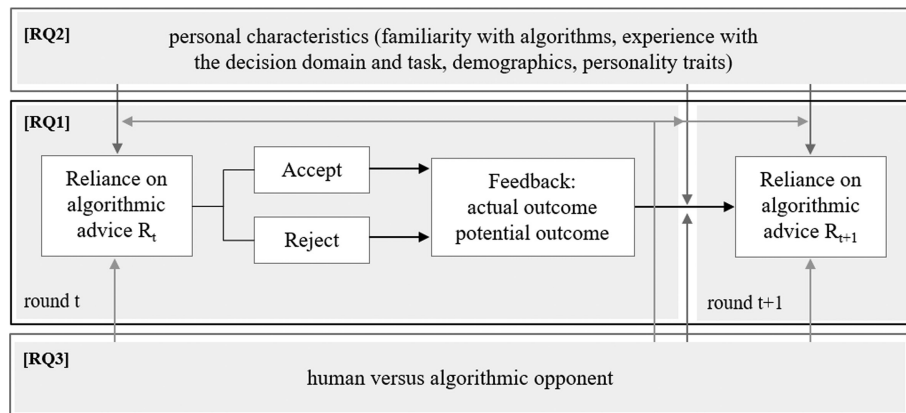
In this study, participants are tasked with predicting human behavior in real-time, a complex and subjective endeavor requiring the identification of behavioral patterns and consideration of strategic interdependencies. Although previous studies have introduced machine learning in strategic play settings (Crandall et al. 2018; Normann and Sternberg 2023), these typically involve human players facing algorithms without receiving decision support. An exception is Schauer and Schnurr (2023), who find that algorithmic decision support influences pricing decisions in a duopoly game when facing an algorithmic competitor but not a human one. Erlei et al. (2020) observe that algorithmic decision support benefits proposers in an ultimatum

game by promoting fairer offers, yet proposers underweight the algorithm's recommendations. Our study uniquely provides evidence on how the personal characteristics of human users affect reliance on an algorithm designed to assist in strategic encounters. Our analyses relate to prior findings about the influence of familiarity with algorithms (e.g., Araujo et al. 2020; Castelo, Bos, and Lehmann 2019; Logg, Minson, and Moore 2019; Önköl et al., 2009; Whitecotton 1996) and of domain experience (e.g., Logg, Minson, and Moore 2019; Luo et al. 2021; Kawaguchi 2021; Kim et al. 2022). Our analyses go beyond prior research in that we check whether such personal characteristics are moderators of dynamic human–algorithm interaction, that is, whether they affect the direction and/or strengths of individuals' reactions to outcomes following acceptance versus rejection of algorithmic advice.

### 1.3 | The Present Study

We designed the strategic interaction setting for this study around the repeated play of the RPS game over a substantial number of rounds. RPS is particularly compelling because the complexity arises not from the game itself, as seen in traditional strategy games like Go or chess, but from the behaviors exhibited by humans playing the game. Increasing the chances of winning a round necessitates anticipating the opponent's choice. Therefore, our experimental task requires an effective algorithm to predict patterns in human behavior and give recommendations based on these predictions. Although it is impossible to gain an advantage over an opponent who fully randomizes their play, research shows that human choices in RPS often follow non-random patterns (Dyson et al. 2016; Wang, Xu, and Zhou 2014). For instance, players tend to choose rock as their opening move (Dyson et al. 2016; Xu, Zhou, and Wang 2013) or employ a win–stay–lose–shift strategy (Wang, Xu, and Zhou 2014), where they repeat a winning choice but switch to a different option after losing. An effective algorithm exploits these patterns, and for this experiment, an algorithm was specifically programmed and trained to do so successfully.

Figure 1 illustrates our analysis framework which builds upon theoretical research on trust in algorithms (Glikson and Woolley 2020; Hoff and Bashir 2015; Marsh and Dibben 2003) and illustrates the three research questions we address. The



**FIGURE 1** | Analysis framework. The figure shows the framework and the three research questions (RQ1, RQ2, and RQ3) that guide the data analyses.

middle level of the figure depicts the dynamic interaction between a human decision-maker and an algorithm in our setting. In each round, the decision-maker makes an initial (provisional) choice, receives the algorithm's advice, and then decides either to follow the advice (switch to the recommended option) or reject it and stick with the initial choice. The round's result provides feedback about both the actual outcome of the choice made (accept vs. reject advice) and the potential outcome had the opposite choice been made. All outcomes (win, draw, and lose) are unambiguous, as are the evaluations of the algorithm's and the decision-maker's predictions. We first assume that the strategic opponent is a human without decision support.

Our first research question (RQ1) examines the impact of this feedback on the decision-maker's subsequent choice between accepting and rejecting advice in the following round. Specifically, we explore how the decision to accept or reject advice, combined with the feedback from the outcome of that decision, influences subsequent behavior. Although complementing prior studies (Dietvorst, Simmons, and Massey 2015; Prahla and Van Swol 2017) by contrasting early-round experiences with later-round reliance on advice, our focus is on the dynamics of the decision to follow the algorithm round-by-round. Previous research suggests that decision-makers may lose trust in an algorithm after observing errors (Chacon, Kausel, and Reyes 2022; Dietvorst, Simmons, and Massey 2015; Dzindolet et al. 2003; Muir and Moray 1996; Prahla and Van Swol 2017) but may also learn to trust the algorithm over time (Bastani, Bastani, and Sinchaisri 2021; Berger et al. 2021; Filiz et al. 2021; Snyder, Keppler, and Leider 2022; Yin, Wortman Vaughan, and Wallach 2019; Zhang 2023). In our setting, decision-makers observe both the actual and potential outcomes in each round, allowing us to draw conclusions about their perspectives and evaluations of the algorithm's advice versus their own choices.

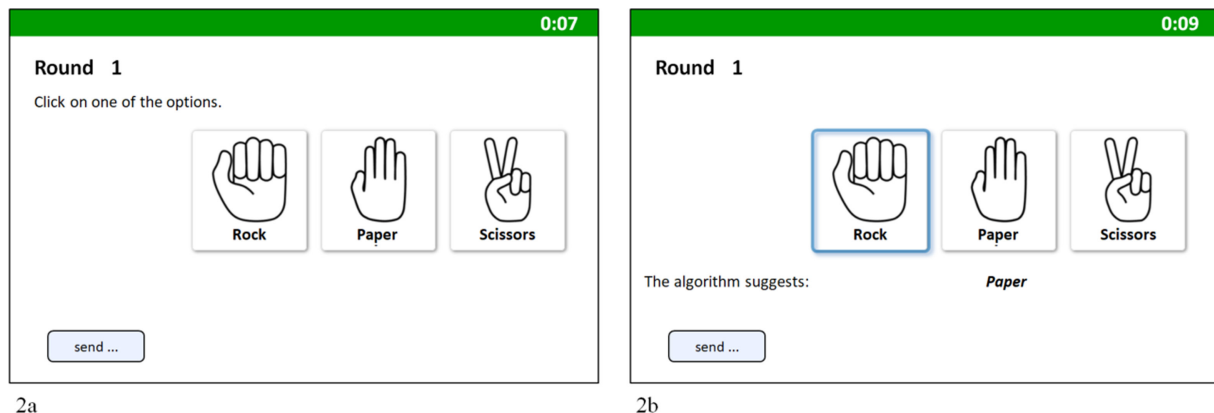
The top level of Figure 1 addresses our second research question (RQ2), showing selected determinants of an individual's basic willingness to trust algorithms derived from previous studies in other decision domains. These include demographic variables (e.g., gender and age), familiarity with algorithms (experience with algorithms in various decision domains and

understanding of their capabilities), experience with the decision domain and the specific task, and personality traits, including the Big Five (Barnett et al. 2015; Rammstedt and John 2007) and competitiveness. We investigate whether prior evidence about these determinants of trust in algorithms applies to the strategic interaction domain we study. More specifically, we analyze whether personal characteristics explain general reliance on algorithms in strategic interaction and whether these characteristics influence the dynamics of human–algorithm interaction. For instance, individuals familiar with algorithms may evaluate round outcomes differently than those with low familiarity; individuals confident in their task skills may react more strongly to algorithm errors. The bottom level of Figure 1 illustrates our third research question (RQ3), which is related to prior research on the impact of the decision context on trust in algorithms (Chiou and Lee 2023; Mahmud et al. 2022, with further references). In the experiment, participants play against both human and algorithmic opponents. Our third research question explores whether the results for the first two questions are unique to facing a human opponent or also apply when the opponent is an algorithm. This inquiry is crucial as the use of algorithms becomes more widespread, increasing the likelihood of facing an algorithmic opponent in strategic encounters.

## 2 | Experimental Design

### 2.1 | Treatments and Procedures

The experiment employs a mixed between-within-subject design to manipulate the conditions under which participants play RPS while receiving decision support from an algorithm (All participants provided informed consent upon registration for the experiment. The requirement for ethics approval was waived by the ethics committee based on its ethics approval principles). Participants engage in RPS over a computer network, completing four sets of 50 rounds each. In the first stage of the experiment, all participants play against each other (human vs. human, HvH). In the second stage, all participants play against the RPS algorithm (human vs. algorithm, HvA). Subsequently, two treatments follow in which participants receive algorithmic decision support: In the human



**FIGURE 2** | Computer interface of the RPS game. The figure shows the computer interface of the Rock–Paper–Scissors game. Players having no algorithm support (Figure 2a) choose an option, send it, and then see the outcome of the round. Players with algorithm support make an initial choice (Figure 2a), send it, then receive advice (Figure 2b), and make the decision to stay with their initial option or switch.



with algorithm versus human (HAvH) treatment, half of the participants are randomly selected to receive advice from the algorithm while playing against the other half, that is, against an unsupported human opponent. In the human with algorithm against algorithm (HAvA) treatment, all participants receive support from the algorithm while playing against the machine, essentially a clone of the algorithm. In each round of play with algorithmic decision support, the respective player makes an initial choice, receives advice from the algorithm, and then decides either to stay with the initial choice or revise it based on the recommendation. Participants select options by clicking on one of three pictograms. When they play with the algorithm, its recommendation is displayed after the initial choice has been made (see Figure 2). The first set of 50 rounds (HvH) begins after five training rounds against a human opponent. Participants receive a show-up fee of 8 EUR, plus a payoff of 2 EUR for a win, 1 EUR for a draw, and 0 EUR for a loss over 12 randomly selected rounds, three from each of the four stages.

Sessions were conducted at the university computer laboratory (The participant pool is managed using ORSEE; Greiner 2015). A total of 176 subjects participated, with 62% being business, economics, or information systems students, 50.6% male, and an average age of 23 years. All participants received algorithmic advice in the HAvA treatment, whereas only half of them ( $N=88$ ) had algorithm support in the HAvH treatment. We thus have a total of 4400 observations from participants interacting with the algorithm in the HAvH treatment ( $88 \text{ participants} \times 50 \text{ rounds}$ ) and 8800 observations for the HAvA treatment. Participants had 12s to make a choice. If they did not manage to do this, they would lose the respective round. Overall, there were 110 such

incidents, amounting to 0.3% of the data. HAvH was tested before HAvA in half of the sessions, and the sequence was reversed in the other half.

## 2.2 | The Algorithm

The algorithm was sourced from a website hosting a programming contest ([www.rpscontest.com](http://www.rpscontest.com)). We selected a top performer, “RPS\_FSm” by “TeleZ” (the source code is publicly available at <http://www.rpscontest.com/entry/5731716400414720>) and trained it over approximately 10,000 rounds of play against test subjects. It is a heuristic reinforcement learning algorithm, meaning that it does not explicitly model a Markov decision process with a reward function but relies on heuristics to determine the best choices for a round. Irrespective of the score, the algorithm continuously strives to win each round, which aligns with the incentive structure of the experiment. To derive the best choice, the algorithm analyzes the history of play in each round, comparing it with the cumulative experience from prior training rounds. It examines sequences in the opponent’s play history, starting with up to 25 previous rounds, and uses the frequencies of these sequences to make decisions. The algorithm learns to identify and exploit patterns in the opponent’s behavior, such as a tendency to use a win–stay–lose–shift strategy. In pretests, the algorithm aggressively exploited such naive strategies. In the experiment, the algorithm performed effectively: In the HvA stage, it won against 59.7% of participants, lost to 33.5%, and tied with 6.8%. In the HAvH treatment, participants who consistently accepted the algorithm’s advice increased their chances of winning from 30.8% to 35.2% and reduced their

**TABLE 1** | Descriptive data on algorithm utilization in the HAvH treatment.

Panel A: Game outcome and advice rejection frequencies					
Average <i>score</i> of participants having algorithmic decision support ( <i>N</i> = 88)				50.13	
Participants that win tie lose the 50-round game:				47.7% 2.3% 50.0%	
Frequencies of advice rejection	All rounds	Rounds 1–25		Rounds 26–50	
all advice	35.2%	33.8%		36.6%	
contradicting advice	50.2%	49.0%		51.3%	
supporting advice	6.1%	6.1%		6.3%	
Panel B: Early experience and subsequent advice rejections: Correlations					
Number of rounds from 1 to <i>t</i> , in which ...	<i>t</i> = 5	<i>t</i> = 10	<i>t</i> = 15	<i>t</i> = 20	<i>t</i> = 25
... the algorithm's advice would lose	0.14	0.17	0.22**	0.31***	0.30***
... the algorithm's advice would win	−0.29***	−0.26**	−0.22**	−0.26**	−0.26**
... the participant's initial choice would lose	−0.03	0.06	0.01	−0.02	0.02
... the participant's initial choice would win	0.14	0.12	0.10	0.08	0.01

*Note:* Panel A of the table shows, for the  $N=88$  participants receiving algorithmic decision support in the HAvH treatment, the average score and frequencies with which they win, tie, or lose the 50 rounds game, and the frequencies with which they reject advice over all 50 rounds, over the first 25 rounds, and over the last 25 rounds. Panel B of the table shows Spearman rank correlations between four proxies of the game’s history up to round  $t$  and the frequencies with which participants reject advice after round  $t$ . The histories span the first 5, 10, 15, 20, and 25 rounds, respectively, and the corresponding rejection frequencies are calculated over rounds 6–50, 11–50, 16–50, 21–50, and 26–50, respectively. Panel B includes only cases of contradicting advice.

– score is the sum of all 50 rounds’ outcomes, with a win (draw, loss) giving 2 (1, 0) points.

– supporting (contradicting) advice is a recommendation given by the algorithm that is equal (unequal) to the participant’s initial choice.

– advice rejection is an indicator variable equal to 1 if the participant’s final choice is not the recommended choice.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .

chances of losing from 35.3% to 30.6%, which translates into an overall chance of winning the 50-round game that roughly equal to the actual 59.7% win rate observed in the HvA stage.

## 2.3 | Measures

Advice is categorized as either supporting (the algorithm's recommendation matches the participant's initial choice) or contradicting. Contradicting advice allows a clear distinction between advice acceptance and rejection, whereas with supporting advice only rejection is clearly defined. We thus measure reliance on algorithmic advice with the indicator variable *reject*, which is equal to 1 when the final choice differs from the algorithm's suggestion. Each round's outcome is coded with scores of 2, 1, and 0 points for a win, draw, and loss, respectively; a total score of 50 ties a 50-round game. A postexperimental questionnaire collected information on participants' familiarity with algorithms, prior experience with the experimental task (RPS), experience in the decision domain (other strategic interaction tasks), demographics (e.g., age, gender, and work experience), and personality traits. Regarding algorithm familiarity, participants reported their experiences with online recommendation systems (e.g., online shops or streaming platforms), their evaluations of strategy game engines (e.g., chess, Go, and poker), and their programming knowledge. Since there is no public knowledge about the sophistication of RPS algorithms, participants likely had no preconceived notions about the algorithm's strength. Their performances in the HvA stage serve as a proxy for their experience with the algorithm. For domain experience, participants self-assessed their skills in strategy games and reported their experience with real-time online strategy games (e.g., Dota, League of Legends, and StarCraft), as these involve algorithms, interaction with other players, and quick decision-making. Participants also self-assessed their skills in playing RPS, and their scores in the initial HvH stage serve as a proxy for their actual RPS skills. For personality traits, we used a short version of the Big Five inventory (Rammstedt and John 2007) and included items on competitiveness. The postexperimental questionnaire also included three questions of the cognitive reflection test (Frederick 2005) and an incentivized dictator game decision that we use as a proxy for distributive fairness preferences. As none of the collected measures explains behavior in the experiment, we do not further comment on them.

## 3 | Results

### 3.1 | Dynamics of Algorithm Utilization in Playing RPS

We start the presentation of results addressing the first research question and show how participants interact with the algorithm in playing RPS. Our analyses focus on the HAvH treatment, where participants receive algorithmic decision support while facing a human opponent without such support. Table 1 presents descriptive statistics on scores and frequencies of rejecting algorithmic advice in the HAvH treatment. Given the algorithm's strong performance in the HvA stage, where it outperforms approximately 60% of participants, having algorithm support while facing an unsupported human opponent should theoretically

confer an advantage. However, this advantage is contingent on participants' willingness to accept the algorithm's advice. Panel A of Table 1 shows that participants do not fully leverage this advantage, achieving an average score of 50.13, only marginally higher than the neutral score of 50. Participants lose 50% (win 47.7%, tie 2.3%) of the 50-round games. They reject contradicting advice 50.2% of the time—approximately every second round—but rarely reject supporting advice (6.1%). Overall, the average overall rejection frequency is 35.2%. The table also presents rejection frequencies for early (rounds 1–25) and late (rounds 26–50) stages of the game, showing a slight increase in rejections in the latter half. There is considerable variation among participants: A median split of overall rejection rates reveals that participants in the below-median group have a rejection rate of 19.8%, outperforming their human opponents, whereas those in the above-median group have a rejection rate of 50.9% and underperform. Rejection rates between these groups do not converge but diverge over the course of the game.

Dietvorst, Simmons, and Massey (2015) coined the term algorithm aversion to describe the phenomenon where decision-makers lose faith in an algorithm after observing its errors. We test for such behavior by examining whether experience with the algorithm's correct or incorrect advice in early rounds influences participants' decisions to accept or reject advice in later rounds. Panel B of Table 1 presents Spearman rank correlations between proxies for early experiences and the frequency of rejecting advice in later rounds. Two proxies for early experiences with the algorithm are derived from counting the number of rounds from round 1 to round  $t$  ( $t = 5, 10, 15, 20$ , and  $25$ ) in which the algorithm has (i) erred (i.e., recommended a losing choice: “algorithm's advice would lose”) or (ii) has provided correct advice (i.e., recommended a winning choice: “algorithm's advice would win”). We calculate the same proxies for the participants' own judgments, that is, we count the number of rounds from 1 to  $t$  in which the participant's initial choice has been (iii) incorrect (“participant's initial choice would lose”) or (iv) correct (“participant's initial choice would win”). The analysis focuses on cases of contradicting advice, where acceptance and rejection are clearly defined. The data reveal both algorithm aversion and appreciation: Early algorithm errors correlate positively with later advice rejections, with the strength and significance of this correlation increasing over time. Conversely, early correct predictions correlate strongly with reliance on the algorithm, particularly in the earliest rounds (1–5 and 1–10). Unlike previous studies (Dietvorst, Simmons, and Massey 2015), we do not observe an asymmetry that exclusively establishes algorithm aversion. Moreover, participants' long-term reactions to their own initial choice accuracy are not significant, suggesting that reliance on advice is more influenced by the quality of the algorithm's advice than by their own choices.

Table 1 provides insights into whether a participant's reliance on advice is correlated with early algorithmic recommendations turning out right or wrong. We further investigate the dynamics of human–algorithm interaction on a round-by-round basis to address our first research question, as illustrated in Figure 1. We examine how a participant's decision to accept or reject advice in round  $t$  and the feedback stemming from observing the outcome of that round affect their behavior in the following round. We hypothesize that participants sensitive to the success of their

**TABLE 2** | Decision to reject advice in round  $t+1$  as a response to round  $t$ 's outcome.

<b>Panel A: Descriptive statistics: Frequencies of rejecting advice in round <math>t+1</math></b>					
<b>Reject advice in round <math>t+1</math>:</b>		<b>All advice</b>		<b>Contradicting advice only</b>	
<b>Frequency [observations groups]</b>					
<b>decision in round <math>t</math>:</b>		<b>Not reject</b>	<b>Reject</b>	<b>Not reject</b>	<b>Reject</b>
outcome of round $t$ :	win	21.1% [979 88]	55.2% [464 82]	15.5% [322 70]	79.0% [291 78]
	draw	28.1% [897 88]	56.3% [512 81]	27.5% [298 78]	79.4% [325 74]
	loss	33.4% [887 88]	38.9% [529 85]	37.1% [267 80]	58.0% [324 80]
	total	27.3% [2763 88]	49.8% [1505 86]	26.0% [887 85]	71.9% [940 85]
<b>Panel B: Logistic regressions</b>					
<b>Dependent variable: reject advice in round <math>t+1</math></b>		<b>All advice</b>	<b>Contradicting advice only</b>		
<b>Independent variables:</b>		<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	
prior round $t$ :	<i>not reject</i> $\cap$ <i>win</i>	−0.365***	−0.705***	−0.636***	
	<i>not reject</i> $\cap$ <i>win</i> $\times$ <i>potential loss</i>			−0.144	
	<i>not reject</i> $\cap$ <i>loss</i>	0.224**	0.451***	0.238	
	<i>not reject</i> $\cap$ <i>loss</i> $\times$ <i>potential win</i>			0.483**	
	<i>reject</i> $\cap$ <i>draw</i>	0.553***	0.954***	0.797***	
	<i>reject</i> $\cap$ <i>draw</i> $\times$ <i>potential loss</i>			0.315	
	<i>reject</i> $\cap$ <i>win</i>	0.569***	0.942***	0.905***	
	<i>reject</i> $\cap$ <i>win</i> $\times$ <i>potential loss</i>			0.088	
	<i>reject</i> $\cap$ <i>loss</i>	−0.122	0.276*	0.490***	
	<i>reject</i> $\cap$ <i>loss</i> $\times$ <i>potential win</i>			−0.434**	
<i>N</i> (groups)		4170 (86)	2719 (86)	2719 (86)	

Note: Panel A shows the frequencies with which participants having algorithm decision support in the HAvH treatment rejected advice in round  $t+1$ , classified by the six potential events of round  $t$  ( $\{\text{reject/not reject}\} \times \{\text{win/draw/lose}\}$ ). Panel B presents fixed effects logistic regressions estimating the decision to reject advice in round  $t+1$  based on the same round  $t$  events, with  $\text{not reject} \times \text{draw}$  being the baseline in each respective regression. As two participants never rejected the advice, the group size is 86 in all regressions. All regressions include round as control variable.

Wald tests of regression coefficients:

model (1),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{win})$ ,  $\chi^2 = 0.01$ ,  $p = .904$ ;

model (1),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{loss})$ ,  $\chi^2 = 25.57$ ,  $p < .001$ ;

model (2),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{win})$ ,  $\chi^2 = 0.01$ ,  $p = .931$ ;

model (2),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{loss})$ ,  $\chi^2 = 24.30$ ,  $p < .001$ ;

model (3),  $H_0$ :  $b(\text{reject} \cap \text{loss}) + b(\text{reject} \cap \text{loss} \cap \text{potential win}) = 0$ ,  $\chi^2 = 0.09$ ,  $p = .77$ .

– *reject* (*not reject*) is an indicator variable equal to 1 if the participant's final choice is not (equal to) the recommended choice.

– *win*, *draw*, and *loss* are indicator variables for the round's outcome.

– *not reject*  $\cap$  *win* etc. are indicator variables for the respective events, that is, a win after not rejecting advice, etc.

– *potential win* and *potential loss* are indicator variables for hypothetical outcomes of a round: In case the participant rejected the advice, *potential loss* (*potential win*) is equal to one if the actual outcome of the round was no loss (no win) but would have been a loss (a win) if the participant had not rejected the advice. In case the participant did not reject advice, *potential loss* (*potential win*) is equal to one if the actual outcome of the round was no loss (no win) but would have been a loss (a win) if the participant had rejected advice.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .

strategy in round  $t$  will be less likely to reject advice in round  $t+1$  after losing with rejected advice in  $t$  or winning with accepted advice in  $t$ , and more likely to reject advice after winning with rejected advice or losing with accepted advice. Panel A of Table 2 presents participants' advice rejection frequencies in round  $t+1$  based on the six possible events in round  $t$ :  $\{\text{reject/not reject advice}\} \times \{\text{win/draw/loss}\}$ . The descriptive data indicate distinct effects of round outcomes on subsequent decisions. After not rejecting advice, the rejection frequency in  $t+1$  is 21.1% following

a win, increases to 28.1% following a draw, and rises to 33.4% following a loss. Conversely, after rejecting advice, participants exhibit an asymmetry: They continue rejecting advice with similar frequencies after a win (55.2%) and a draw (56.3%), but the frequency drops to 39.9% after a loss. This pattern is consistent when considering only contradicting advice. The asymmetry in behavior suggests that participants' reactions are influenced more by their success in predicting their opponents' behavior than by their assessments of the algorithm's advice.



To compare participants' reactions after accepting versus rejecting advice, we conduct fixed effects logistic regressions estimating the likelihood of rejecting advice in round  $t + 1$ . Panel B of Table 2 presents three models: Model (1) includes all observations and estimates the likelihood of rejecting advice in round  $t + 1$  based on indicator variables reflecting the actual events of round  $t$ . Models (2) and (3) consider only contradicting advice in round  $t$ . Model (2) replicates Model (1) for robustness, whereas Model (3) includes additional measures to test the impact of potential outcomes on decision-making. All models control for the round number.

The baseline for all regression models is not reject  $\cap$  draw, that is, the case where the participant accepts advice and the outcome is a draw. Model (1) results conform to the descriptive data from Panel A. They show that relative to the baseline (not reject  $\cap$  draw), a different outcome after accepting advice significantly affects the likelihood of rejecting advice in the next round, with a win decreasing and a loss increasing it. Instead, for outcomes after rejecting advice, a win and a draw have nearly identical coefficients (0.569 and 0.553), that is, both outcomes equally increase the likelihood of rejecting advice in the next round, whereas a loss reduces the propensity to repeat rejecting advice relative to a draw as its

coefficient is significantly lower (coefficients: reject  $\cap$  loss vs. reject  $\cap$  draw, Wald test,  $\chi^2 = 25.57$ ,  $p < .001$ ). The coefficient for reject  $\cap$  loss is insignificant, indicating that losing after rejecting advice does not significantly alter the propensity to reject advice again relative to accepting advice and experiencing a draw. This asymmetry supports the notion of self-serving evaluations, as participants are stricter towards the algorithm than towards themselves. Model (2) corroborates these findings, with larger coefficients and a positive, though marginally significant, coefficient for experiencing a loss after rejecting advice ( $p = .089$ ). Model (3) examines the effect of potential outcomes. It reveals that participants ignore potential outcomes after wins, whether they accepted (coefficient:  $-0.144$ ,  $p = .630$ ) or rejected advice (coefficient:  $0.088$ ,  $p = .667$ ). Realizing that the potential outcome would have been a win after an actual loss affects behavior, and the effect is again asymmetric: After not rejecting contradicting advice, seeing that the decision resulted in a loss because the algorithm erred does not significantly impact behavior per se (not reject  $\cap$  loss:  $0.238$ ,  $p = .244$ ), but realizing that the own initial choice would have won—rather than only drawn—significantly increases the rejection likelihood in  $t + 1$  (not reject  $\cap$  loss  $\times$  potential win:  $0.483$ ,  $p = .043$ ). Conversely, after rejecting contradicting advice, an actual loss even increases the

**TABLE 3** | Random effects logistic regressions.

<b>Dependent variable:</b>							
<b>reject advice in round <math>t + 1</math></b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>	<b>(5)</b>	<b>(6)</b>	<b>(7)</b>
<i>Independent variables:</i>							
round $t$ : not reject $\cap$ win	−0.373***	−0.373***	−0.374***	−0.372***	−0.373***	−0.374***	−0.376***
round $t$ : not reject $\cap$ loss	0.230**	0.230**	0.229**	0.229**	0.230**	0.230**	0.226**
round $t$ : reject $\cap$ draw	0.671***	0.669***	0.673***	0.671***	0.671***	0.670***	0.670***
round $t$ : reject $\cap$ win	0.678***	0.678***	0.676***	0.681***	0.678***	0.677***	0.676***
round $t$ : reject $\cap$ loss	−0.025	−0.024	−0.025	−0.026	−0.025	−0.025	−0.026
familiarity with algorithms		−0.350***					−0.277***
experience with online strategy games			−0.280***				−0.192**
self-assessment of strategy game skills				−0.216**			−0.160*
score HvH					−0.003		0.087
score HvA						0.100	0.106
N (groups)	4170 (86)	4170 (86)	4170 (86)	4170 (86)	4170 (86)	4170 (86)	4170 (86)

*Note:* The table shows results from random effects logistic regressions estimating the likelihood of rejecting advice in round  $t + 1$  based on the events in round  $t$  and proxies for the experience made prior to or within the experiment. For all measures representing events, *not reject  $\times$  draw* is the baseline in each respective regression. As two participants never rejected the advice, the group size is 86 in all regressions. All regressions include round as control variable. *– reject (not reject)* is an indicator variable equal to 1 if the participant's final choice is not (equal to) the recommended choice.

*– win, draw, and loss* are indicator variables for the round's outcome.

*– not reject  $\cap$  win* etc. are indicator variables for the respective events, that is, a win after not rejecting advice, etc.

*– familiarity with algorithms* is the average of four items, scaled to lie in (0, 5). Items ask for assessments of (i) the helpfulness of recommendations in online shopping and (ii) on streaming platforms, of (iii) the strengths of strategy game (chess, Go, and poker) engines, and for a self-assessment of programming know-how.

*– experience with online strategy games* is a two-item average. Items ask for experience with both playing online strategy games and watching others playing such games on streaming platforms (like Twitch)

*– self-assessment of strategy game skills* is a two-item average, scaled to lie in (0, 5). Items ask for self-assessments of participants' skills in playing strategy games (card games and board gamers) in general as well as Rock–Paper–Scissors.

*– score HvH* is the standardized score in the HvH stage, and *score HvA* is the standardized score in the HvA stage. Standardization is based on the mean and standard deviation of scores from the  $N = 88$  participants having algorithm support.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .

likelihood of rejecting advice in the next round (reject  $\cap$  loss: 0.490,  $p = .009$ ); only if the decision-maker sees that following the algorithm would have given a win (reject  $\cap$  loss  $\times$  potential win:  $-0.434$ ,  $p = .026$ ) neutralizes this effect of the actual loss (Wald test,  $\chi^2 = 0.09$ ,  $p = .77$ ).

### 3.2 | Personal Characteristics as Determinants and Moderators of Algorithm Utilization

We now address our second research question, which has two parts: Do the determinants of individuals' basic willingness to trust an algorithm explain the heterogeneity in their behaviors? Do these determinants moderate the relationship between dynamic experience and reliance on advice identified in the main analyses? Our focus remains on the HAvH treatment. Descriptive data on the postexperimental questionnaire items are presented in Table A1 in Appendix A.

#### 3.2.1 | Main Effects of Personal Characteristics on Algorithm Utilization

We first examine whether participants' demographics, personality traits, familiarity with algorithms, and experience with the decision domain and the specific experimental task have a main effect on participants' decisions to reject versus accept algorithmic advice by incorporating each measure into the logistic regression of Panel B, Table 2, Model (1). Table 3 presents selected results. To include these measures in the regression, we estimate random effects models instead of fixed effects models. Model (1) compares random and fixed effects estimations, showing similar results. Models (2) to (6) include one measure each, whereas Model (7) includes all measures. The results indicate that decisions to accept or reject advice are significantly related to prior experience measures but not to the experience made in the first two stages of the experiment (scores in HvH and HvA): Participants who have outperformed their human counterparts in HvH do not show higher rejection frequencies, nor do those who have performed better against the algorithm in HvA use its advice significantly less. Measures of familiarity with algorithms, experience with online strategy games, and self-assessed strategy game skills all negatively correlate with rejecting advice. This finding is expected for algorithm familiarity and online strategy game experience, but the negative correlation with self-assessed strategy game skills is surprising, implying that participants who perceive themselves as stronger players rely more on the algorithm.

Table 3 does not present results for demographic characteristics or personality traits. Regarding demographics, we find a significant gender effect: Male participants are less likely to reject advice than female participants, but this effect disappears once experience measures are controlled for. No systematic relationships are found between personality traits (including competitiveness) and reliance on algorithmic advice. Model (7), which includes all variables, shows stable effects identified in Models (2) to (6), although the significance of experience with online strategy games and self-assessed strategy game skills diminishes. Robustness checks (untabulated) for contradicting advice only yield similar results.

#### 3.2.2 | Personal Characteristics as Moderators

The second part of the research question examines whether determinants of heterogeneity in algorithmic advice reliance moderate the relationship between dynamic experience and advice reliance found in the first research question. Specifically, we explore whether participants with varying familiarity with algorithms, domain and task experience, or personality traits respond differently to dynamic experiences with the algorithm. We repeat the regression analyses from Table 2, interacting the identified drivers of algorithm reliance with measures reflecting events of round  $t$ . Table 4 presents these results, starting with the reference model (Model [1] from Table 2). It then shows results for familiarity with algorithms, experience with online strategy games, self-assessment of strategy game skills, and scores in the HvH and HvA stages.

The results demonstrate that neither experience with online strategy games nor self-assessment of strategy game skills significantly moderates the effects of dynamic experience on behavior. Strategy game skills interactions are not significant, and only the experience with online strategy games interaction with losing after not rejecting advice shows an effect. Scores in the HvH and HvA stages also have minimal and selective effects on dynamic behavior. In stark contrast, familiarity with algorithms strongly moderates the effects of round outcomes after rejecting advice: Interactions with reject  $\cap$  draw and reject  $\cap$  win are highly significant. These effects only occur after rejecting advice; interactions after not rejecting advice are insignificant. This indicates that participants familiar with algorithms are significantly more sensitive to the consequences of rejecting, but not following, algorithmic advice. We also estimate regressions for demographic characteristics and personality traits (results untabulated), but we find no systematic moderation regarding any of these variables.

### 3.3 | Human Opponent Versus Algorithmic Opponent

In the settings investigated in the literature, decision support from algorithms typically refers to judgment and decision-making situations where the human decision-maker needs to forecast a future event or estimate an unknown fact. Our setting differs significantly as it involves the human decision-maker facing an opponent in a strategic interaction. To address our third research question, which examines whether our findings about the dynamics of human–algorithm interaction and overall trust in the algorithm are unique to the HAvH treatment, we implemented an additional treatment: HAvA. In this treatment, all human participants play RPS with the support of the algorithm, but their opponent is not another human; it is an algorithm. Specifically, for each participant, the computer system runs two independent clones of the identically trained algorithm: one serving as an advisor to the human participant and the other as the opponent. Appendix B (Tables B1, B2, B3, and B4) provides detailed results of repeating all analyses from Sections 3.1 and 3.2 for the HAvA treatment. These results indicate that facing an algorithmic opponent, as opposed to a human opponent, significantly alters participant behavior.

TABLE 4 | Personal characteristics as moderators of human–algorithm interaction.

Panel B: Logistic regressions: Decision to reject advice in round $t+1$					
Dependent variable: reject advice in round $t+1$	Referencemodel	Familiarity with algorithms	Experience with online strategy games	Strategy game skills	ScoreHvH      ScoreHvA
<i>Indep. variables (round t):</i>					
<i>not reject</i> $\cap$ win	−0.365***	−0.360***	−0.365***	−0.367***	−0.373***      −0.367***
<i>not reject</i> $\cap$ win interacted		0.148	−0.153	−0.063	−0.035      0.142
<i>not reject</i> $\cap$ loss	0.224**	0.229**	0.226**	0.224**	0.219**      0.225**
<i>not reject</i> $\cap$ loss interacted		0.002	−0.204*	−0.020	−0.084      0.214*
<i>reject</i> $\cap$ draw	0.553***	0.615***	0.583***	0.563***	0.546***      0.574***
<i>reject</i> $\cap$ draw interacted		0.431***	0.153	0.075	−0.087      −0.120
<i>reject</i> $\cap$ win	0.569***	0.607***	0.558***	0.568***	0.572***      0.580***
<i>reject</i> $\cap$ win interacted		0.357***	−0.116	−0.090	−0.218*      −0.008
<i>reject</i> $\cap$ loss	−0.122	−0.104	−0.085	−0.120	−0.130      −0.116
<i>reject</i> $\cap$ loss interacted		0.248*	0.126	0.005	−0.102      −0.034
<i>N (groups)</i>	4170 (86)	4170 (86)	4170 (86)	4170 (86)	4170 (86)      4170 (86)

*Note:* The table shows results from fixed effects logistic regressions estimating the likelihood of rejecting advice in round  $t+1$  based on the events in round  $t$ . In each model (except for the reference model), the indicator variable for the respective event is included both in isolation and interacted with the measure that gives the model its name. All measures that are interacted with the indicator variables are z-standardized. The event *not reject*  $\cap$  draw is the baseline in each respective regression. As two participants never rejected the advice, the group size is 86 in all regressions. All regressions include round as control variable.

- *reject (not reject)* is an indicator variable equal to 1 if the participant's final choice is not (equal to) the recommended choice.
- *win, draw, and loss* are indicator variables for the round's outcome.
- *not reject*  $\cap$  win etc. are indicator variables for the respective events, that is, a win after not rejecting advice, etc.
- *familiarity with algorithms* is the average of four items; *experience with online strategy games* is a two-item average; *self-assessment of strategy game skills* is a two-item average; see Table 3 for details.
- *score HvH* is the standardized score in the HvH stage, and *score HvA* is the standardized score in the HvA stage.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .

First, advice rejections become substantially more frequent in the HAvA treatment (43.9%) compared to the HAvH treatment (35.2%). Second, in the HAvA treatment, only the experience of the algorithm giving winning advice (not advice that loses the round) is significantly correlated with rejection frequencies in later rounds. Third, the dynamic behavior over the game rounds changes: After accepting advice, a loss no longer predicts the subsequent decision to reject advice, and after rejecting advice, a draw is no longer equivalent to a win in predicting behavior in the following round. Fourth, none of the proxies for experience with algorithms or the decision domain significantly relate to the decision to accept or reject advice. The only variable showing an effect is performance against the algorithm in the HvA stage, which negatively affects reliance on the algorithm's advice. Lastly, none of the determinants investigated moderates the effect of a round's outcome on the decision to follow the advice in the subsequent round. These findings suggest that the dynamics of human–algorithm interaction and trust in the algorithm observed in the HAvH treatment do not fully translate to the HAvA treatment, highlighting the impact of the type of opponent on behavior and decision-making processes.

## 4 | Discussion

This exploratory study investigates the dynamics of human–algorithm interaction in a setting of strategic interaction. We organize our discussion of its results and their implications along with the three research questions guiding the analyses. Then, we address some of the study's limitations.

### 4.1 | Dynamics of Human-Algorithm Interaction

Our observation that individuals make suboptimal use of algorithmic recommendations aligns with prior studies, but our results provide a more nuanced understanding of human–algorithm interaction dynamics and uniquely capture reactions to correct and incorrect predictions by the algorithm and participants' own predictions. The regression results in Table 2, Panel B, indicate that participants are highly sensitive to the outcome of their decision both to accept and to reject advice, but with asymmetric reactions: Participants who follow the algorithm's advice increase reliance after a successful prediction, and decrease reliance after a failed prediction, particularly when their initial choice was correct, whereas they treat a draw as neutral. Conversely, after rejecting advice, both a win and a draw reinforce the tendency to continue rejecting advice. Hence, participants who reject advice tend to continue doing so, showing greater tolerance for their own errors than for the algorithm's errors, consistent with findings from Dietvorst, Simmons, and Massey (2015). Even losing a round does not break their tendency to reject advice unless they see that following the algorithm would have led to a win (Table 2, Panel B, Model [3]). This is consistent with participants who are reluctant to follow the algorithm having a high threshold for changing their behavior, even though it is clear that in our setting the algorithm cannot guarantee a win. Such a threshold would also explain why participants' overall experience with the algorithm (HvA stage score) does

not significantly affect their reliance on it. Our findings on the dynamics of algorithmic advice reliance have no direct parallels in the literature, but related studies offer insights. Dietvorst and Bharti (2020) suggest that decision-makers are particularly averse to algorithms and less sensitive to prediction errors when facing (high) outcome uncertainty. They prefer their own judgment, despite its lower accuracy. Similarly, Burton, Stein, and Jensen (2023) argue that human–machine interaction dynamics change under uncertainty, where future events' probabilities cannot be reliably calculated. Our finding that participants persist in rejecting advice aligns with these arguments, as our setting involves a decision-maker who cannot expect the algorithm to eliminate uncertainty but only to offer a strategic edge. Participants evaluate algorithmic advice based on its perceived utility in predicting their opponent's behavior and their own predictive ability, indicating that they do not expect their opponents to randomize their play effectively.

A prevalent argument in discussions of algorithm aversion is that human behavior hinders the adoption of superior forecasting and decision-making methods (e.g., Dietvorst, Simmons, and Massey 2018). Although our study reveals under-reliance on algorithmic advice and persistence among those who reject it, we are careful with drawing the same conclusion without reservations. This is because this study's task involves predicting human behavior in strategic interactions, where algorithmic decision support can influence the very behavior being predicted. In the RPS game, no algorithm can aid a player if the opponent successfully randomizes their moves. Awareness of algorithmic support may prompt human players to attempt randomization. Broadly, this suggests that in strategic interaction settings, algorithm aversion is driven by multiple factors. These include the belief that personal intuition about human behavior surpasses the algorithm and the notion that winning is feasible only between human players.

### 4.2 | Personal Characteristics as Determinants of Reliance on Algorithms

Our second research question asks whether the personal characteristics of decision-makers receiving algorithmic advice explain their basic willingness to rely on the algorithm in our setting of strategic interaction and whether these characteristics influence the dynamics of human–algorithm interaction. Our findings on participants' willingness to rely on an algorithm align with the growing literature on trust in algorithms. Regarding familiarity with algorithms, our results confirm previous research that suggests “algorithm literacy” is necessary for appreciating algorithmic recommendations. Trust in algorithms is influenced by multiple factors, including the algorithm's opacity—how transparent and explainable its processes are (Burrell 2016). Research shows that trust increases when the “black box” is made transparent (Mahmud et al. 2022). Although users are generally more hesitant to trust opaque algorithms, familiarity can mitigate this reluctance. In our study, participants were only informed that the algorithm was professionally programmed for RPS, and their interactions with it during the HvA stage provided practical exposure. Our findings indicate that familiarity with algorithms



moderates dynamic behavior: Participants familiar with algorithms are more likely to follow its advice but are also more critical, as they tend to reject advice more frequently after a successful personal play. This suggests that familiarity with algorithms enhances both trust and critical evaluation, highlighting the need for future research to consider the dynamics of human–algorithm interaction.

We also discover that domain experience positively influences reliance on the algorithm. Participants with higher self-assessed strategy game skills and more experience with online strategy games were more likely to follow the algorithm. This contrasts with previous findings of a negative (Arkes, Dawes, and Christensen 1986; Kim et al. 2022; Logg, Minson, and Moore 2019; Sieck and Arkes 2005) or inversely U-shaped relationship (Allen and Choudhury 2022; Luo et al. 2021). Instead, we find a monotonic positive relationship. Allen and Choudhury (2022) argue that two opposing effects drive the inverse U-shaped relationship: More experienced individuals are more resistant to algorithms, but also have superior assessments, increasing reliance on algorithms when they prove helpful. In our setting, the latter effect appears dominant. Additionally, our finding that experience with online strategy games enhances reliance on advice aligns with Araujo et al. (2020), who found that online self-efficacy positively relates to algorithm reliance, unlike general self-efficacy.

Lastly, we examined whether personality traits affect reliance on the algorithm's recommendations. Previous studies have shown that personality traits can influence evaluations of algorithmic judgments, particularly dutifulness (a facet of conscientiousness), neuroticism, and extraversion (Barnett et al. 2015; Lee 2018; Mahmud et al. 2022; Neumann, Niessen, and Meijer 2023). We found no significant correlation between personality traits and participants' willingness to follow the algorithm's advice; finding no such effects might be due to a lack of reliability in our Big Five measures, though, which consisted of two items only.

### 4.3 | Facing a Human Versus Algorithmic Opponent in Strategic Interaction

Our final research question asks whether our results are unique to the HAvH treatment where decision-makers face a human opponent or carry over to the HAvA treatment where the opponent is an algorithm. The overall picture is that the results are unique to facing a human opponent: When decision-makers face the algorithm, overall rejection frequencies increase, dynamic behaviors change, and the determinants of individuals' basic willingness to follow the algorithm no longer contribute to explaining behavior. A potential explanation for these results can be derived from an argument in Burton, Stein, and Jensen (2020, 2023), who discuss incentive structures as potential determinants of trust in algorithms: If the incentive of a human decision-maker is to make the “best” judgment or decision, the decision-maker may refrain from using an algorithm if the situation is such that even the best algorithms cannot provide perfect predictions. In our study, when facing the algorithm as their opponent, participants may

have rejected algorithmic advice more frequently for a similar reason: Receiving algorithmic decision support while facing an algorithm of comparable strength does not give decision-makers a competitive advantage (but only levels the playing field). Overall, our finding that the dynamics of human–algorithm interaction crucially depend on whether the opponent is a human or an algorithm highlights the importance of the exact context in which algorithmic advice is used.

### 4.4 | Limitations

This study is subject to several limitations. First, we collected data on personal characteristics in a postexperimental questionnaire, that is, after participants played all four stages of the experiment. The self-assessments regarding familiarity with algorithms and especially domain experience we elicit may thus be influenced by the experience made within the experiment. However, placing the questionnaire before the experiment may have similar carryover effects (see Asay et al. 2022; Pirlott and MacKinnon 2016; and Spencer, Zanna, and Fong 2005, for methodological discussions of carryover and reverse causality effects in experiments). Second, when receiving algorithmic advice, participants have only the option to either keep their initial choice or switch to a different choice, that is, they cannot compromise between their own initial prediction and that of the algorithm. We do not see this as a weakness of the design, though, as many real-world applications are such that advice comes in the form of a recommendation that can only be fully accepted or fully rejected. Also, we control for individuals' general reluctance to follow advice when we estimate fixed-effects regressions. Still, we do not provide evidence on the role of individuals' fundamental tendency to stick with their initial judgments even though switching is the better strategy. Providing such direct evidence would have required contrasting the actual with a different manipulation of how the decision-maker interacts with the algorithm. A related point is that the experimental design does not allow participants to opt out of receiving recommendations, which may affect how advice is utilized in real-world settings. Furthermore, we do not contrast reliance on algorithmic advice with human advice directly, which could provide additional insights into algorithm aversion. Finally, the complexity of understanding the algorithm's functioning without explicit information may have influenced participants' behavior, suggesting the need for further research on the effects of algorithm transparency.

### 5 | Conclusion

This study provides crucial insights into human decision-makers' reliance on algorithmic advice, particularly within the context of strategic interactions exemplified by the RPS game. By exploring the dynamics of human–algorithm interaction, we address three key research questions concerning the utilization of algorithmic advice, the drivers of reliance on algorithms, and the influence of the type of opponent on decision-making behavior. Our findings reveal that participants generally underutilize advice from a machine learning algorithm, especially after experiencing algorithmic



errors early in the game. Conversely, those who witness the algorithm making successful predictions early on are more likely to rely on its advice in subsequent rounds. This dynamic behavior underscores the sensitivity of participants to the outcomes of their decisions, whether they accept or reject advice. Notably, participants show asymmetric reactions to these outcomes: Rejecting advice tends to reinforce this behavior regardless of the outcome, whereas accepting advice leads to varied responses based on the result of the round. We also examine whether individuals' basic willingness to trust algorithms is influenced by personal characteristics such as familiarity with algorithms, experience with the decision domain, and other traits. Our results indicate that familiarity with algorithms and domain experience positively correlate with reliance on algorithmic advice. Moreover, familiarity with algorithms significantly moderates dynamic behavior, with more familiar individuals exhibiting stronger reactions to outcomes after rejecting advice but not after accepting it. Our study further differentiates between the effects observed when participants face a human opponent versus an algorithmic one. When playing against an algorithmic opponent, participants' behavior notably changes: They reject advice more frequently, and their reliance on the algorithm's advice is primarily influenced by the algorithm's performance in early rounds. Additionally, the determinants of trust in the algorithm and the dynamics of interaction observed with a human opponent do not fully translate to interactions with an algorithmic opponent. This highlights the significant impact of the opponent's nature on decision-making processes and trust in algorithmic advice.

Our findings contribute to the expanding body of research on algorithm aversion and reliance, confirming and extending previous insights into how individuals interact with and trust algorithms. The evidence suggests that enhancing familiarity with algorithms can improve their integration into decision-making processes, potentially mitigating underutilization and fostering more effective human–algorithm collaboration. As algorithms have entered various areas of human life and are used not only for an increasing number of private but also business decisions, it is vital to deepen and broaden our understanding of the determinants of successful human–algorithm interaction. Future research should explore additional contexts and task domains to help build a broader body of knowledge about the factors driving the dynamics of human–algorithm interaction. Additionally, examining how different types of feedback and varying levels of transparency in algorithmic decision support influence trust and reliance could provide deeper insights into optimizing human–algorithm interaction. Overall, our study underscores the nuanced and dynamic nature of human–algorithm interaction, emphasizing the importance of experience and context in shaping trust and reliance on algorithmic decision support systems.

## Acknowledgements

The authors thank George Wright (editor), three anonymous reviewers and participants at the EAA annual congress, the ACMAR, and the GOR Workshop for helpful comments and suggestions. Open Access funding enabled and organized by Projekt DEAL.

## Data Availability Statement

The data that support the findings of this study are openly available in OSF at [https://osf.io/zr82p/?view\\_only=a257fb2bc693412881a19282e5cd4d4a](https://osf.io/zr82p/?view_only=a257fb2bc693412881a19282e5cd4d4a).

## References

- Allen, R., and P. Choudhury. 2022. "Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion." *Organization Science* 33, no. 1: 149–169. <https://doi.org/10.1287/orsc.2021.1554>.
- Araujo, T., N. Helberger, S. Kruikemeier, and C. H. De Vreese. 2020. "In AI We Trust? Perceptions About Automated Decision-Making by Artificial Intelligence." *Ai & Society* 35: 611–623. <https://doi.org/10.1007/s00146-019-00931-w>.
- Arkes, H. R., R. M. Dawes, and C. Christensen. 1986. "Factors Influencing the use of a Decision Rule in a Probabilistic Task." *Organizational Behavior and Human Decision Processes* 37, no. 1: 93–110. [https://doi.org/10.1016/0749-5978\(86\)90046-4](https://doi.org/10.1016/0749-5978(86)90046-4).
- Asay, H. S., R. D. Guggenmos, K. Kadous, L. Koonce, and R. Libby. 2022. "Theory Testing and Process Evidence in Accounting Experiments." *Accounting Review* 97, no. 6: 23–43. <https://doi.org/10.2308/TAR-2019-1001>.
- Barnett, T., A. W. Person, R. Pearson, and F. W. Kellermanns. 2015. "Five-Factor Model Personality Traits as Predictors of Perceived and Actual Usage of Technology." *European Journal of Information Systems* 24: 374–390. <https://doi.org/10.1057/ejis.2014.10>.
- Bastani, H., O. Bastani, and W. P. Sinchaisri. 2021. "Improving Human Sequential Decision-Making With Reinforcement Learning." Working Paper, arXiv. <https://doi.org/10.48550/arXiv.2108.08454>.
- Berger, B., M. Adam, A. Rühr, and A. Benlian. 2021. "Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn." *Business & Information Systems Engineering* 63, no. 1: 55–68. <https://doi.org/10.1007/s12599-020-00678-5>.
- Bigman, Y. E., and K. Gray. 2018. "People Are Averse to Machines Making Moral Decisions." *Cognition* 181: 21–34. <https://doi.org/10.1016/J.COGNITION.2018.08.003>.
- Burrell, J. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3, no. 1: 2053951715622512. <https://doi.org/10.1177/2053951715622512>.
- Burton, J. W., M. K. Stein, and T. B. Jensen. 2020. "A Systematic Review of Algorithm Aversion in Augmented Decision Making." *Journal of Behavioral Decision Making* 33, no. 2: 220–239. <https://doi.org/10.1002/bdm.2155>.
- Burton, J. W., M. K. Stein, and T. B. Jensen. 2023. "Beyond Algorithm Aversion in Human–Machine Decision-Making." In *Judgment in Predictive Analytics*, edited by M. In Seifert, 3–26. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-30085-1\\_1](https://doi.org/10.1007/978-3-031-30085-1_1).
- Cabiddu, F., L. Moi, G. Patriotta, and D. G. Allen. 2022. "Why Do Users Trust Algorithms? A Review and Conceptualization of Initial Trust and Trust Over Time." *European Management Journal* 40, no. 5: 685–706. <https://doi.org/10.1016/j.emj.2022.06.001>.
- Caro, F., and A. S. de Tejada Cuenca. 2023. "Believing in Analytics: Managers' Adherence to Price Recommendations From a DSS." *Manufacturing & Service Operations Management* 25, no. 2: 524–542. <https://doi.org/10.1287/msom.2022.1166>.
- Castelo, N., M. W. Bos, and D. R. Lehmann. 2019. "Task-Dependent Algorithm Aversion." *Journal of Marketing Research* 56, no. 5: 809–825. <https://doi.org/10.1177/0022243719851788>.

- Chacon, A., E. E. Kausel, and T. Reyes. 2022. "A Longitudinal Approach for Understanding Algorithm use." *Journal of Behavioral Decision Making* 35, no. 4: e2275. <https://doi.org/10.1002/bdm.2275>.
- Chiou, E. K., and J. D. Lee. 2023. "Trusting Automation: Designing for Responsivity and Resilience." *Human Factors* 65, no. 1: 137–165. <https://doi.org/10.1177/00187208211009995>.
- Chong, L., G. Zhang, K. Goucher-Lambert, K. Kotovsky, and J. Cagan. 2022. "Human Confidence in Artificial Intelligence and in Themselves: The Evolution and Impact of Confidence on Adoption of AI Advice." *Computers in Human Behavior* 127: 107018. <https://doi.org/10.1016/j.chb.2021.107018>.
- Commerford, B. P., S. A. Dennis, J. R. Joe, and J. W. Ulla. 2022. "Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence." *Journal of Accounting Research* 60, no. 1: 171–201. <https://doi.org/10.1111/1475-679X.12407>.
- Crandall, J. W., M. Oudah, I.-O. F. Tennom, et al. 2018. "Cooperating With Machines." *Nature Communications* 9, no. 1: 233. <https://doi.org/10.1038/s41467-017-02597-8>.
- De Freitas, J., S. Agarwal, B. Schmitt, and N. Haslam. 2023. "Psychological Factors Underlying Attitudes Toward AI Tools." *Nature Human Behaviour* 7, no. 11: 1845–1854. <https://doi.org/10.1038/s41562-023-01734-2>.
- Dietvorst, B. J., and S. Bharti. 2020. "People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error." *Psychological Science* 31, no. 10: 1302–1314. <https://doi.org/10.1177/0956797620948841>.
- Dietvorst, B. J., J. P. Simmons, and C. Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." *Journal of Experimental Psychology: General* 144, no. 1: 114–126. <https://psycnet.apa.org/doi/10.1037/xge0000033>.
- Dietvorst, B. J., J. P. Simmons, and C. Massey. 2018. "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms if They Can (Even Slightly) Modify Them." *Management Science* 64, no. 3: 1155–1170. <https://doi.org/10.1287/mnsc.2016.2643>.
- Dyson, B. J., J. M. P. Wilbiks, R. Sandhu, G. Papanicolaou, and J. Lintag. 2016. "Negative Outcomes Evoke Cyclic Irrational Decisions in Rock, Paper, Scissors." *Scientific Reports* 6, no. 1: 1–6. <https://doi.org/10.1038/srep20479>.
- Dzindolet, M. T., S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck. 2003. "The Role of Trust in Automation Reliance." *International Journal of Human-Computer Studies* 58, no. 6: 697–718. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7).
- Erlei, A., F. Nekdem, L. Meub, A. Anand, and U. Gadiraju. 2020. "Impact of Algorithmic Decision Making on Human Behavior: Evidence From Ultimatum Bargaining." In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (Vol. 8, pp. 43–52). <https://doi.org/10.1609/hcomp.v8i1.7462>.
- Filiz, I., J. R. Judek, M. Lorenz, and M. Spiwoks. 2021. "Reducing Algorithm Aversion Through Experience." *Journal of Behavioral and Experimental Finance* 31: 100524. <https://doi.org/10.1016/j.jbef.2021.100524>.
- Frederick, S. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19, no. 4: 25–42. <https://doi.org/10.1257/089533005775196732>.
- Glikson, E., and A. W. Woolley. 2020. "Human Trust in Artificial Intelligence: Review of Empirical Research." *Academy of Management Annals* 14, no. 2: 627–660. <https://doi.org/10.5465/annals.2018.0057>.
- Greiner, B. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments With ORSEE." *Journal of the Economic Science Association* 1, no. 1: 114–125. <https://doi.org/10.1007/s40881-015-0004-4>.
- Gunning, D., M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Z. Yang. 2019. "XAI-Explainable Artificial Intelligence." *Science Robotics* 4, no. 37: eaay7120. <https://doi.org/10.1126/scirobotics.aay7120>.
- Himmelstein, M., and D. V. Budescu. 2023. "Preference for Human or Algorithmic Forecasting Advice Does Not Predict if and How It Is Used." *Journal of Behavioral Decision Making* 36, no. 1: e2285. <https://doi.org/10.1002/bdm.2285>.
- Hoff, K. A., and M. Bashir. 2015. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." *Human Factors* 57, no. 3: 407–434. <https://doi.org/10.1177/0018720814547570>.
- Jussupow, E., I. Benbasat, and A. Heinzl. 2020. "Why Are We Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion." Working Paper, ECIS 2020 Proceedings. [https://aisel.aisnet.org/ecis2020\\_rp/168](https://aisel.aisnet.org/ecis2020_rp/168).
- Kawaguchi, K. 2021. "When Will Workers Follow an Algorithm? A Field Experiment With a Retail Business." *Management Science* 67, no. 3: 1670–1695. <https://doi.org/10.1287/mnsc.2020.3599>.
- Kim, J. H., M. Kim, D. W. Kwak, and S. Lee. 2022. "Home-Tutoring Services Assisted With Technology: Investigating the Role of Artificial Intelligence Using a Randomized Field Experiment." *Journal of Marketing Research* 59, no. 1: 79–96. <https://doi.org/10.1177/00222437211050351>.
- Lee, M. K. 2018. "Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management." *Big Data & Society* 5, no. 1: 205395171875668. <https://doi.org/10.1177/2053951718756684>.
- Lin, W., S. H. Kim, and J. Tong. 2022. "What Drives Algorithm Use? An Empirical Analysis of Algorithm Use in Type 1 Diabetes Self-Management." Working Paper, SSRN. <https://doi.org/10.2139/ssrn.3891832>.
- Littman, M. L., I. Ajunwa, G. Berger, C. Boutilier, M. Currie, F. Doshi-Velez, G. Hadfield, M. C. Horowitz, C. Isbell, H. Kitano, K. Levy, T. Lyons, M. Mitchell, J. Shah, S. Sloman, S. Vallor, and T. Walsh. 2021. "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report." Working Paper, arXiv. <https://doi.org/10.48550/arXiv.2210.15767>.
- Liu, J., W. T. Yue, A. Leung, and X. Zhang. 2022. "Find the Good. Seek the Unity: A Hidden Markov Model of Human-AI Delegation Dynamics." Working Paper, SSRN. <https://doi.org/10.2139/ssrn.4150772>.
- Liu, M., X. Tang, S. Xia, S. Zhang, Y. Zhu, and Q. Meng. 2023. "Algorithm Aversion: Evidence From Ridesharing Drivers." *Management Science*. <https://doi.org/10.1287/mnsc.2022.02475>.
- Logg, J. M., J. A. Minson, and D. A. Moore. 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>.
- Luo, X., M. S. Qin, Z. Fang, and Z. Qu. 2021. "Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions." *Journal of Marketing* 85, no. 2: 14–32. <https://doi.org/10.1177/0022242920956676>.
- Mahmud, H., A. N. Islam, S. I. Ahmed, and K. Smolander. 2022. "What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion." *Technological Forecasting and Social Change* 175: 121390. <https://doi.org/10.1016/j.techfore.2021.121390>.
- Marsh, S., and M. R. Dibben. 2003. "The Role of Trust in Information Science and Technology." *Annual Review of Information Science and Technology (ARIST)* 37: 465–498.
- Merritt, S. M., J. L. Unnerstall, D. Lee, and K. Huber. 2015. "Measuring Individual Differences in the Perfect Automation Schema." *Human Factors* 57, no. 5: 740–753. <https://doi.org/10.1177/0018720815581247>.

- Muir, B. M., and N. Moray. 1996. "Trust in Automation. Part II. Experimental Studies of Trust and Human Intervention in a Process Control Simulation." *Ergonomics* 39, no. 3: 429–460. <https://doi.org/10.1080/00140139608964474>.
- Neumann, M., A. S. M. Niessen, and R. R. Meijer. 2023. "Predicting Decision-Makers' Algorithm Use." *Computers in Human Behavior* 145: 107759. <https://doi.org/10.1016/j.chb.2023.107759>.
- Normann, H. T., and M. Sternberg. 2023. "Human–Algorithm Interaction: Algorithmic Pricing in Hybrid Laboratory Markets." *European Economic Review* 152: 104347. <https://doi.org/10.1016/j.eurocorev.2022.104347>.
- Önköl, D., P. Goodwin, M. Thomson, S. Gönöl, and A. Pollock. 2009. "The Relative Influence of Advice From Human Experts and Statistical Methods on Forecast Adjustments." *Journal of Behavioral Decision Making* 22, no. 4: 390–409. <https://doi.org/10.1002/bdm.637>.
- Pirlott, A. G., and D. P. MacKinnon. 2016. "Design Approaches to Experimental Mediation." *Journal of Experimental Social Psychology* 66: 29–38. <https://doi.org/10.1016/j.jesp.2015.09.012>.
- Prahl, A., and L. Van Swol. 2017. "Understanding Algorithm Aversion: When Is Advice From Automation Discounted?" *Journal of Forecasting* 36, no. 6: 691–702. <https://doi.org/10.1002/for.2464>.
- Prahl, A., and L. Van Swol. 2021. "Out With the Humans, in With the Machines?: Investigating the Behavioral and Psychological Effects of Replacing Human Advisors With a Machine." *Human-Machine Communication* 2, no. 1: 11–234. <https://doi.org/10.30658/hmc.2.11>.
- Rammstedt, B., and O. P. John. 2007. "Measuring Personality in one Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German." *Journal of Research in Personality* 41, no. 1: 203–212. <https://doi.org/10.1016/j.jrp.2006.02.001>.
- Schauer, A., and D. Schnurr. 2023. "Competition and Collaboration Between Human and Artificial Intelligence in Digital Markets." Working Paper, SSRN. <https://doi.org/10.2139/ssrn.4565678>.
- Sieck, W. R., and H. R. Arkes. 2005. "The Recalcitrance of Overconfidence and Its Contribution to Decision Aid Neglect." *Journal of Behavioral Decision Making* 18, no. 1: 29–53. <https://doi.org/10.1002/bdm.486>.
- Snyder, C., S. Keppler, and S. Leider. 2022. "Algorithm Reliance Under Pressure: The Effect of Customer Load on Service Workers." Working Paper, SSRN. <https://doi.org/10.2139/ssrn.4066823>.
- Spencer, S. J., M. P. Zanna, and G. T. Fong. 2005. "Establishing a Causal Chain: Why Experiments Are Often More Effective Than Mediational Analyses in Examining Psychological Processes." *Journal of Personality and Social Psychology* 89, no. 6: 845–851. <https://doi.org/10.1037/0022-3514.89.6.845>.
- Venkatesh, V., and F. D. Davis. 2000. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies." *Management Science* 46, no. 2: 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>.
- Wang, Z., B. Xu, and H. J. Zhou. 2014. "Social Cycling and Conditional Responses in the Rock–Paper–Scissors Game." *Scientific Reports* 4, no. 1: 1–7. <https://doi.org/10.1038/srep05830>.
- Whitecotton, S. M. 1996. "The Effects of Experience and a Decision Aid on the Slope, Scatter, and Bias of Earnings Forecasts." *Organizational Behavior and Human Decision Processes* 66, no. 1: 111–121. <https://doi.org/10.1006/obhd.1996.0042>.
- Xu, B., H. J. Zhou, and Z. Wang. 2013. "Cycle Frequency in Standard Rock–Paper–Scissors Games: Evidence From Experimental Economics." *Physica a: Statistical Mechanics and Its Applications* 392, no. 20: 4997–5005. <https://doi.org/10.1016/j.physa.2013.06.039>.
- Yeomans, M., A. Shah, S. Mullainathan, and J. Kleinberg. 2019. "Making Sense of Recommendations." *Journal of Behavioral Decision Making* 32, no. 4: 403–414. <https://doi.org/10.1002/bdm.2118>.
- Yin, M., J. Wortman Vaughan, and H. Wallach. 2019. "Understanding the Effect of Accuracy on Trust in Machine Learning Models." In Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems (pp. 1–12). <https://doi.org/10.1145/3290605.3300509>.
- Zhang, P. 2023. "Taking Advice from ChatGPT." Working Paper, arXiv. <https://doi.org/10.48550/arXiv.2305.11888>.

## Appendix A: Postexperimental Questionnaire Items

**TABLE A1** | Postexperimental questionnaire items.

(N=176)	Mean	SD	p25 p50 p75
<b>Prior experience with algorithms and strategy games, assessment of skills</b>			
Knowledge of programming languages (JavaScript, PHP, and C++ ...) (1 to 4)	1.65	0.73	1 2 2
Usefulness of recommendations in online shops (e.g., Amazon) (1 to 5)	2.64	0.96	2 3 3
Usefulness of recommendations on streaming platforms (e.g., Netflix) (1 to 5)	3.39	0.99	3 3 4
Use of digital assistants (e.g., Alexa, Siri) (1 = never, 5 = daily)	1.72	1.04	1 1 3
Evaluation of strategy games engines (agreement, 1 to 5)	3.68	1.07	3 4 4
("Whether chess, Go or poker: I am convinced that computer programs are already better than the best human players today, or will be in the near future")			
Online strategy games (e.g., Dota, League of Legends, and StarCraft) (agreement, 1 to 5)			
I regularly watch players stream their play	1.57	1.06	1 1 2
I regularly play them myself	1.73	1.15	1 1 2.5
When I play them, I am successful	1.91	1.24	1 1 3
Evaluation of own strategy play skills (agreement, 1 to 5)	3.70	0.94	3 4 4
("In many games (like card or board games) you need not only luck, but also a good strategy to win. I am pretty good in such games")			
Evaluation of own RPS skills (agreement, 1 to 5)	3.29	0.78	3 3 4
("If I had to play Rock–Paper–Scissors against everyone else in this room, I would win more often than lose the game")			
<b>Demographics</b>			
age (years)	22.9	3.85	20 22 25
work experience (month part-time or full-time employment)	21.3	32.8	1 8 30
gender: 50.6% male, 49.4% female (other: 0%)			
course of studies: business or economics: 62%, other programs: 27%, no student: 11%			
<b>Personality traits</b> ("R" indicates that the original items were reversed): I see myself as ...			
... reserved, quiet (extraversion) [R]	3.13	1.17	2 3 4
... extraverted, enthusiastic (extraversion)	3.47	1.07	3 4 4
... sympathetic, warm (agreeableness)	3.38	1.13	3 4 4
... critical, quarrelsome (agreeableness) [R]	2.84	1.05	2 3 4
... disorganized, careless (conscientiousness) [R]	2.67	1.09	2 3 3
... dependable, self-disciplined (conscientiousness)	3.91	0.87	3 4 4.5
... calm, emotionally stable (neuroticism) [R]	2.95	1.23	2 3 4
... anxious, easily upset (neuroticism)	3.05	1.17	2 3 4
... conventional, uncreative (openness) [R]	3.09	1.35	2 3 4
... open to new experiences, (openness)	3.57	1.11	3 4 4
... who likes to prove their skills in competitions with others (competitiveness)	3.53	1.13	3 4 4
<b>Other</b>			
Dictator proposal (pie size: 4 Euros): 41% give 0, 25% give 1, 32% give 2, 2.2% give 3 or 4			
Correct answers to cognitive reflection test: bat and ball	67%		
machines, minutes, widgets	90%		
lily pads	69%		

*Note:* The table shows descriptive statistics for the postexperimental questionnaire items. Familiarity with algorithms is measured as the average of the four items knowledge of programming languages, usefulness of recommendations in online shops, usefulness of recommendations, and evaluation of strategy games engines. A participant's self-assessment of her/his strategy game skills is measured as the average of the two items evaluation of own strategy play skills and evaluation of own RPS skills. The proxy for experience with online strategy games is the average of the two items online strategy games: I regularly watch players stream their play and I regularly play them myself.

## Appendix B: Data Analyses for the HAvA Treatment

**TABLE B1** | Descriptive data on algorithm utilization in the HAvA treatment.

Panel A: Game outcome and advice rejection frequencies					
Average score of participants having algorithmic decision support ( $N=88$ )			49.27		
Participants that win tie lose the 50-round game:			38.6% 8.0% 53.4%		
Frequencies of advice rejection	All rounds	Rounds 1–25		Rounds 26–50	
all advice	43.9%	43.5%		44.4%	
contradicting advice	61.3%	60.6%		62.1%	
supporting advice	10.3%	10.1%		10.3%	
Panel B: Early experience and subsequent advice rejections: Correlations					
Number of rounds from 1 to $t$ , in which ...	$t=5$	$t=10$	$t=15$	$t=20$	$t=25$
... the algorithm's advice would lose	0.03	0.12	0.14*	0.12	0.22***
... the algorithm's advice would win	−0.19**	−0.26***	−0.17**	−0.17**	−0.23***
... the participant's initial choice would lose	−0.07	−0.11	−0.09	−0.02	−0.07
... the participant's initial choice would win	0.05	0.08	0.07	0.03	0.08

*Note:* Panel A of the table shows, for the  $N=176$  participants receiving algorithmic decision support in the HAvA treatment, the average score and frequencies with which they win, tie, or lose the 50 rounds game, and the frequencies with which they reject advice over all 50 rounds, over the first 25 rounds, and over the last 25 rounds. Panel B of the Table shows Spearman rank correlations between four proxies of the game's history up to round  $t$  and the frequencies with which participants reject advice after round  $t$ . The histories span the first 5, 10, 15, 20, and 25 rounds, respectively, and the corresponding rejection frequencies are calculated over rounds 6–50, 11–50, 16–50, 21–50, and 26–50, respectively. Panel B includes only cases of contradicting advice.

– score is the sum of all 50 rounds' outcomes, with a win (draw, loss) giving 2 (1, 0) points.

– supporting (contradicting) advice is a recommendation given by the algorithm that is equal (unequal) to the participant's initial choice

– advice rejection is an indicator variable equal to 1 if the participant's final choice is not the recommended choice.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .



**TABLE B2** | HAvA: Decision to reject advice as a response to prior rounds outcome.

Panel A: Descriptive statistics: Frequencies of rejecting advice in round $t + 1$				
Reject advice in round $t + 1$ : Frequency [observations groups]	All advice		Contradicting advice only	
decision in round $t$ :	not reject	reject	not reject	reject
outcome of round $t$ : win	31.4% [1676 176]	63.4% [1166 164]	26.0% [516 143]	86.1% [722 159]
draw	36.9% [1516 176]	54.6% [1263 167]	37.2% [486 145]	75.0% [756 163]
loss	39.3% [1611 176]	46.4% [1352 169]	38.9% [509 143]	66.6% [820 164]
total	35.8% [4803 176]	54.4% [3781 172]	34.0% [1511 162]	75.5% [2298 168]

Panel B: Logistic regressions			
Dependent variable: reject advice in round $t + 1$	All advice	Contradicting advice only	
Independent variables:	(1)	(2)	(3)
prior round $t$ : not reject $\cap$ win	-0.270***	-0.589***	-0.529***
not reject $\cap$ win $\times$ potential loss			-0.114
not reject $\cap$ loss	0.085	0.020	0.080
not reject $\cap$ loss $\times$ potential win			-0.122
reject $\cap$ draw	0.096	0.210*	-0.020
reject $\cap$ draw $\times$ potential loss			0.457***
reject $\cap$ win	0.434***	0.588***	0.617***
reject $\cap$ win $\times$ potential loss			0.055
reject $\cap$ loss	-0.235***	-0.032	0.171
reject $\cap$ loss $\times$ potential win			-0.366***
$N$ (groups)	8473 (172)	5522 (169)	5522 (169)

Note: Panel A shows the frequencies with which participants having algorithm decision support in the HAvA treatment rejected advice in round  $t + 1$ , classified by the six potential events of round  $t$  ( $\{\text{reject|not reject}\} \times \{\text{win|draw|lose}\}$ ). Panel B of the table presents fixed effects logistic regressions estimating the decision to reject advice in round  $t + 1$  based on the same round  $t$  events, with not reject  $\times$  draw being the baseline in each respective regression. All regressions include round as control variable. Wald tests of regression coefficients:

model (1),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{win})$ ,  $\chi^2 = 15.23$ ,  $p < .001$

model (1),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{loss})$ ,  $\chi^2 = 16.50$ ,  $p < .001$

model (2),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{win})$ ,  $\chi^2 = 17.78$ ,  $p < .001$

model (2),  $H_0$ :  $b(\text{reject} \cap \text{draw}) = b(\text{reject} \cap \text{loss})$ ,  $\chi^2 = 8.16$ ,  $p = .004$

– reject (not reject) is an indicator variable equal to 1 if the participant's final choice is not (equal to) the recommended choice.

– win, draw, and loss are indicator variables for the round's outcome.

– not reject  $\cap$  win etc. are indicator variables for the respective events, that is, a win after not rejecting advice, etc.

– potential win and potential loss are indicator variables for hypothetical outcomes of a round: In case the participant rejected the advice, potential loss (potential win) is equal to one if the actual outcome of the round was no loss (no win), but would have been a loss (a win) if the participant had not rejected the advice. In case the participant did not reject advice, potential loss (potential win) is equal to one if the actual outcome of the round was no loss (no win), but would have been a loss (a win) if the participant had rejected advice.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .

**TABLE B3** | HAvA: Random effects logistic regressions.

<i>Dependent variable:</i>							
<i>reject advice in round <math>t + 1</math></i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Independent variables:</i>							
<i>round <math>t</math>: not reject <math>\cap</math> win</i>	−0.271***	−0.271***	−0.271***	−0.271***	−0.271***	−0.271***	−0.271***
<i>round <math>t</math>: not reject <math>\cap</math> loss</i>	0.089	0.089	0.089	0.089	0.089	0.089	0.089
<i>round <math>t</math>: reject <math>\cap</math> draw</i>	0.188**	0.188**	0.188**	0.188**	0.188**	0.187**	0.187**
<i>round <math>t</math>: reject <math>\cap</math> win</i>	0.537***	0.537***	0.537***	0.537***	0.536***	0.536***	0.537***
<i>round <math>t</math>: reject <math>\cap</math> loss</i>	−0.151*	−0.151*	−0.151*	−0.151*	−0.151*	−0.150*	−0.151*
<i>familiarity with algorithms</i>		−0.070					−0.038
<i>experience with online strategy games</i>			−0.074				−0.067
<i>self-assessment of strategy game skills</i>				−0.017			0.001
<i>score HvH</i>					−0.027		−0.015
<i>score HvA</i>						0.155**	0.151**
<i>N (groups)</i>	8473 (172)	8473 (172)	8473 (172)	8473 (172)	8473 (172)	8473 (172)	8473 (172)

*Note:* The table shows results from random effects logistic regressions estimating the likelihood of rejecting advice in round  $t + 1$  based on the events in round  $t$  and proxies for the experience made prior to or within the experiment. For all measures representing events, *not reject*  $\times$  *draw* is the baseline in each respective regression. As four participants never rejected the advice, the group size is 172 in all three regressions. All regressions include round as control variable.

– *reject (not reject)* is an indicator variable equal to 1 if the participant's final choice is not (equal to) the recommended choice.

– *win, draw, and loss* are indicator variables for the round's outcome.

– *not reject  $\cap$  win* etc. are indicator variables for the respective events, that is, a win after not rejecting advice, etc.

– *familiarity with algorithms* is the average of four items, scaled to lie in (0, 5). Items ask for assessments of (i) the helpfulness of recommendations in online shopping and (ii) on streaming platforms, of (iii) the strengths of strategy game (chess, Go, and poker) engines, and for a self-assessment of programming know-how.

– *experience with online strategy games* is a two-item average. Items ask for experience with both playing online strategy games and watching others playing such games on streaming platforms (like Twitch)

– *self-assessment of strategy game skills*: is a two-item average, scaled to lie in (0, 5). Items ask for self-assessments of participants' skills in playing strategy games (card games, board gamers) in general as well as Rock–Paper–Scissors.

– *score HvH* is the standardized score in the HvH stage, and *score HvA* is the standardized score in the HvA stage. Standardization is based on the mean and standard deviation of scores from the  $N = 88$  participants having algorithm support.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .

**TABLE B4** | HAvA: Personal characteristics as moderators of human–algorithm–interaction.

Panel B: Logistic regressions: Decision to reject advice in round $t + 1$						
<i>Dependent variable: reject advice in round <math>t + 1</math></i>	Referencemodel	Familiarity with algorithms	Experience with online strategy games	Strategy game skills	Score HvH	Score HvA
<i>Indep. variables (round <math>t</math>):</i>						
<i>not reject</i> $\cap$ <i>win</i>	−0.270***	−0.268***	−0.269***	−0.267***	−0.270***	−0.269***
<i>not reject</i> $\cap$ <i>win interacted</i>		0.118	0.056	0.169**	−0.031	0.005
<i>not reject</i> $\cap$ <i>loss</i>	0.085	0.088	0.085	0.086	0.083	0.085
<i>not reject</i> $\cap$ <i>loss interacted</i>		0.113	0.036	0.036	0.064	0.078
<i>reject</i> $\cap$ <i>draw</i>	0.096	0.097	0.097	0.095	0.096	0.103
<i>reject</i> $\cap$ <i>draw interacted</i>		0.021	0.043	−0.043	−0.009	−0.136*
<i>reject</i> $\cap$ <i>win</i>	0.434***	0.437***	0.433***	0.435***	0.435***	0.440***
<i>reject</i> $\cap$ <i>win interacted</i>		0.099	0.049	−0.060	0.042	−0.165*
<i>reject</i> $\cap$ <i>loss</i>	−0.235***	−0.244***	−0.238***	−0.235***	−0.236***	−0.234***
<i>reject</i> $\cap$ <i>loss interacted</i>		−0.067	−0.003	0.012	0.035	−0.129
<i>N (groups)</i>	8473 (172)	8473 (172)	8473 (172)	8473 (172)	8473 (172)	8473 (172)

*Note:* The table shows results from fixed effects logistic regressions estimating the likelihood of rejecting advice in round  $t + 1$  based on the events in round  $t$ . In each model (except for the reference model), the indicator variable for the respective event is included both in isolation and interacted with the measure that gives the model its name. All measures that are interacted with the indicator variables are z-standardized. The event *not reject*  $\cap$  *draw* is the baseline in each respective regression. As four participants never rejected the advice, the group size is 172 in all regressions. All regressions include round as control variable. – *reject (not reject)* is an indicator variable equal to 1 if the participant's final choice is not (equal to) the recommended choice.

– *win, draw, and loss* are indicator variables for the round's outcome.

– *not reject*  $\cap$  *win* etc. are indicator variables for the respective events, that is, a win after not rejecting advice, etc.

– *familiarity with algorithms* is the average of four items; *experience with online strategy games* is a two-item average; *self-assessment of strategy game skills* is a two-item average; see Table 3 for details.

– *score HvH* is the standardized score in the HvH stage, and *score HvA* is the standardized score in the HvA stage.

\*, \*\*, and \*\*\* indicate significance levels of  $p < .10$ ,  $p < .05$ , and  $p < .01$ .