Delegating in the Age of AI: Preferences for Decision Autonomy

Radosveta Ivanova-Stenzel (TU Berlin)
Michel Tolksdorf (TU Berlin)

Discussion Paper No. 558

December 22, 2025

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, Ludwig Maximilian University of Munich, 80539 Munich, Germany
info@rationality-and-competition.de

Delegating in the Age of AI: Preferences for Decision Autonomy†

December 22, 2025

Radosveta Ivanova-Stenzel[a,*], Michel Tolksdorf[a]

[a] *Technische Universität Berlin, Germany*

Abstract

Despite the documented benefits of algorithmic decision-making, individuals often prefer to retain control rather than delegate decisions to AI agents. To what extent are the aversion to and distrust of algorithms rooted in a fundamental discomfort with giving up decision authority? Using two incentivized laboratory experiments across distinct decision domains, hiring (social decision-making) and forecasting (analytical decision-making), and decision architecture (nature and number of decisions), we elicit participants' willingness to delegate decisions separately to an AI agent and a human agent. This within-subject design enables a direct comparison of delegation preferences across different agent types. We find that participants consistently underutilize both agents, even when informed of the agents' superior performance. However, participants are more willing to delegate to the AI agent than to the human agent. Our results suggest that algorithm aversion may be driven less by distrust in AI and more by a general preference for decision autonomy. This implies that efforts to increase algorithm adoption should address broader concerns about control, rather than focusing solely on trust-building interventions.

Keywords: Algorithm, Delegation, Artificial Intelligence, Trust in AI, Experiment, Preferences
JEL classifications: C72, C91, D44, D83

[*] Corresponding author: `ivanova-stenzel@tu-berlin.de`

[†] Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 "Rationality and Competition" (project number 280092119) as well as by the Berlin Centre for Consumer Policies (BCCP), Germany is gratefully acknowledged. The study received ethics approval by the TU Berlin's Faculty of Economics and Management (VII) Ethics Committee (project number 20230424). The experiments for this study were pre-registered at AsPredicted.org (#131736 & #200326). Declarations of interest: none.

1. Introduction

Individuals often resist delegating decision-making authority even when delegation would lead to superior outcomes. While delegation can reduce cognitive load, improve performance, and leverage specialized expertise, individuals frequently opt to make suboptimal decisions themselves, favoring personal agency over efficiency. For example, individuals prefer to manage their own stock portfolio instead of delegating the decisions to a financial advisor, despite making frequent emotional trades and lower returns (Barber and Odean, 2001). Examining how individuals respond to authority in a hierarchical relationship, Fehr et al. (2013) observe that principals systematically underdelegate authority, even when doing so leads to substantial efficiency losses. Prior research attributes this persistent reluctance to perceived loss of control or loss of autonomy (Wiltermuth and Neale, 2011; Bartling et al., 2014; Owens et al., 2014; Longoni et al., 2019).[1]

Recent studies suggest similar behavioral patterns when the potential delegate is an artificial intelligence (AI) system or algorithm rather than a human. Despite well-documented gains in accuracy and efficiency offered by algorithmic decision-making tools, many individuals prefer to rely on their own judgment, a phenomenon known as algorithm aversion (see, e.g., Burton et al., 2020; Jussupow et al., 2020; Chugunova and Sele, 2022; Mahmud et al., 2022, for extensive surveys on algorithm aversion). Yet, existing research stops short of answering a crucial question: Is algorithm aversion truly about mistrusting AI? We address this question by testing whether resistance to algorithmic decision-making is uniquely tied to AI or rooted in a broader discomfort with delegation, irrespective of the agent type.

Prior research offers no conclusive answer. While some studies show that individuals are less likely to delegate to AI than to humans (e.g., Dietvorst et al., 2015; Leyer and Schneider, 2019), others find a preference for AI over human agents in strategic decision settings (e.g., Logg et al., 2019; Candrian and Scherer, 2022; Germann and Merkle, 2023; Holzmeister et al., 2023). A key limitation of the existing literature is its reliance on between-subjects designs and dichotomous choice setups – participants are often asked either whether to give up control to a certain agent or to choose between two agents without the option of retaining control. This restricts the ability to observe individual-level trade-offs between autonomy and delegation across agent types.[2]

We address this gap through two incentivized laboratory experiments using a within-subject design, in which participants make separate delegation decisions for an AI agent and a human agent. Each decision is made independently and offers participants the choice to retain or relinquish control. This approach enables an intra-individual comparison of preferences for ceding control to an AI versus delegating authority to another human. In the first experiment, participants engage in a hiring task (a socially framed selection decision) where success depends on choosing the best-fitting candidate. In the second experiment, participants complete a forecasting task (an analytically framed prediction problem) where success depends on accuracy. The two experiments also differ with respect to the nature (binary versus non-binary) and number (repeated versus one-shot) of decisions.

In both experiments, we use a menu-based design that systematically varies the cost structure of decision

[1] Prior research has also evidenced a general tendency to discount external advice, a widespread phenomenon known as "egocentric advice discounting" (see Yaniv and Kleinberger (2000); Bonaccio and Dalal (2006)).

[2] In their survey, Kaufmann et al. 2023 highlight the lack of within-subject studies and the importance and impact of task variety.

errors. Participants complete two sets of simple choice problems: one involving the AI agent and one involving the human agent. For each problem, participants decide whether to retain control or delegate the decision. In particular, we vary the size of the payment reduction due to errors when the external agent's decision (AI or human) becomes payoff-relevant across the list of the choice problems and count the number of cases in which a subject delegates the decision to the agent. Hence, for each individual, we obtain two measures of their willingness to delegate decisions, one to the AI agent and one to the human agent. The difference between these two measures captures the individual's relative trust in AI decisions while accounting for their general preference for decision autonomy. This design allows us to isolate preferences for retaining control versus delegating and to directly compare attitudes toward AI versus human agents. Another advantage of our design is that it allows us to classify participants based on their behavioral patterns, facilitating a more nuanced analysis of heterogeneity in delegation preferences. We employ performance-based tasks that provide full feedback on participants' own performance and the performance of both agents, eliminating information asymmetries and isolating delegation preferences from misperceived competence.

We find that people are willing to sacrifice potential economic gains in order to maintain control over their own decisions. This holds true for interactions with both AI and human agents. Compared to performance-based predictions, participants consistently underutilize both AI and human agents, even when informed that those agents outperform them. However, they show a stronger preference for delegating to AI than to human agents. These findings suggest that the commonly observed algorithm aversion can be largely attributed to a broader reluctance to give up personal decision-making authority.

This paper makes three key contributions. First, it quantifies the extent to which aversion to giving up control, rather than algorithm-specific distrust, explains the reluctance to delegate to AI systems. Second, it offers an intra-individual comparison of delegation to AI and human agents, a dimension largely overlooked in prior work. Third, it provides experimental evidence across distinct decision domains and architectures, which enhances the generalizability of the findings.

The paper offers new evidence on the psychological and behavioral underpinnings of algorithm aversion. While individuals systematically underutilize both AI and human agents, even when aware of the agents' superior performance, they are more willing to delegate to AI than to other humans. This suggests that algorithm aversion may be better understood as a preference for decision autonomy, rather than distrust in AI per se.

The results have implications for the design of human-AI decision systems. They suggest that increasing experience with delegation and reducing general aversion to give up control may be more effective than focusing solely on increasing trust in AI. Promoting familiarity with external decision aids, including but not limited to AI, may facilitate more effective and efficient decision-making.

2. Experimental design

We conducted two experiments, where participants were asked to either make a hiring decision based on real workers' performance (Experiment 1) or to forecast real outcomes (Experiment 2). While the experiments differ in the decision domain and architecture (see Table 1), they have the same within-subject structure that enables direct comparison of delegation choices across agent types.

Table 1: Features of the experiments

	Experiment 1	Experiment 2
Task	hiring	forecasting
Domain	social	non-social
Nature of decisions	binary	non-binary
Number of decisions	repeated	one-shot

We begin by outlining the common features of the design of the two experiments before describing the specifics of each experiment. In both experiments, we implemented tasks that allow for a precise measure of performance after observing the realized outcomes as well as for the comparison of agents' performances on the same scale, regardless of the nature of the external agent, AI or human. The participants knew that (i) the AI agent's decisions were based on an algorithm created specifically for the respective task, and (ii) the other human's decisions stem from a similar experiment conducted in the past that employed the same task.

Each experiment consisted of two parts: a payoff-irrelevant training part and a payoff-relevant part, where the participant's payoff depended on the correctness of the hiring choices (Experiment 1) or, respectively, the precision of the forecasts (Experiment 2).

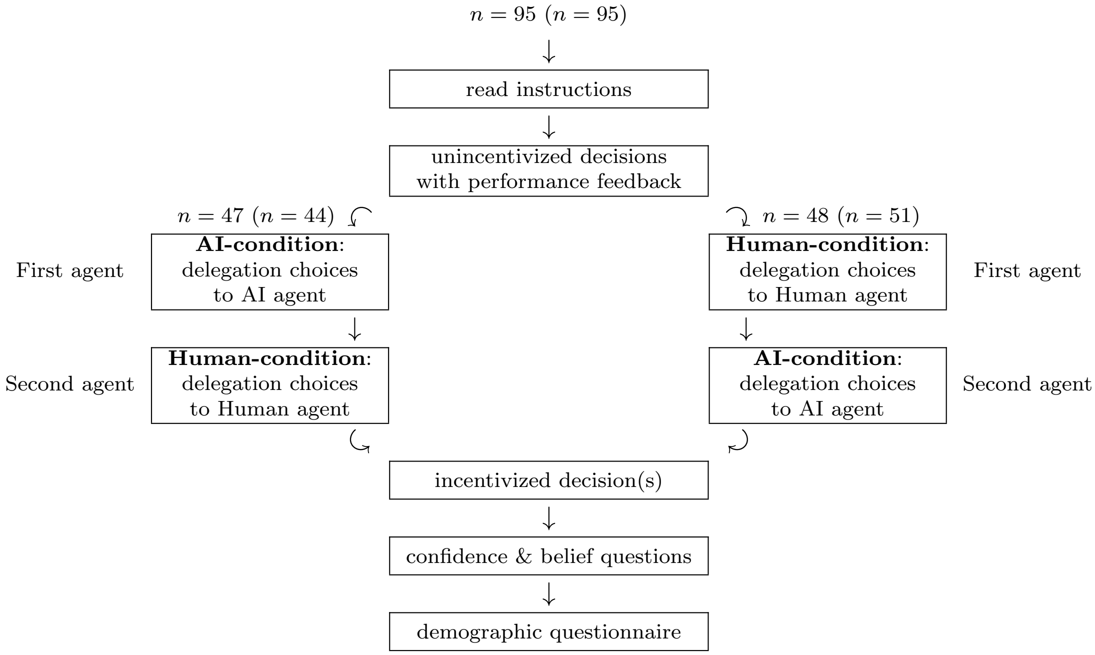
The first, payoff-irrelevant part consisted of ten rounds. After completing the task in a round, participants received feedback on their own performance, the performance of an AI agent, and the performance of another human. The idea was to allow participants to understand the task well enough and gain experience with their own decisions, as well as the decisions made by both agents (AI and human), before deciding whether to delegate decision-making authority in the second, payoff-relevant part. Given the complete feedback provided about the participants' own performance and the performance of both agent types, any misperceptions about the participants' abilities in comparison to those of the agents should not have influenced their actual delegation decisions.

In the second, payoff-relevant part, subjects were first presented with a menu of simple choice problems. We used the menu-based approach first introduced in Ivanova-Stenzel and Tolksdorf (2024) to elicit preferences for relying on external decisions when performing a task. In each problem, participants had to choose between performing the task on their own or delegating it to an agent. The problems differed with respect to the performance-based payoff when the decisions are delegated to the agent. By varying the payoff for employing the agent's decisions, the menu-based approach allows for measuring the degree of aversion to ceding decision authority.

We employed a within-subject design and implemented two conditions. The conditions differ with respect to the nature of the agent. In the "AI"-condition, subjects had to decide whether to give up control to the AI agent. In the "Human"-condition, participants had the choice between using their own decisions or the decisions of the human agent whose performance they learned during the first training part.

To control for order effects, participants in each session of each experiment were randomly assigned to one of two groups. Subjects from the first group participated in the "AI"-condition first, followed by the "Human"-condition, and vice versa for the subjects from the second group. At the end of the experiment, one of the two conditions (AI or Human) was randomly selected, and one of the choice problems from that condition was randomly selected and implemented for real payment.

Figure 1: Procedure of Experiment 1 (Experiment 2)



After participants decided which decisions, their own or the agent's, should become payoff-relevant, they were asked to perform the task again. This design choice excludes the possibility that participants could have delegated the decision to an agent solely to receive payment without exerting effort.

At the end of the second part, participants were asked questions about their beliefs regarding expected errors and their confidence in both their own and the agents' performance (see Tables 9 and 10 in the Appendix for the complete set of questions about beliefs and confidence).

Participants also completed a post-experimental questionnaire, which included, among other things, a self-assessed, hypothetical measure of risk attitude on a $0-10$ scale (SOEP), as well as questions about gender, age, and field of study. Figure 1 summarizes the experimental procedure.

In the following, we outline the specific features of each experiment.

Experiment 1: Hiring decisions

In Experiment 1, we adopted the hiring task from Dargnies et al. (2024), in which participants were required to make several binary hiring decisions.[3] More specifically, the participants had to select one of two candidates to hire based on information about their performance on two different tasks (Task 1 and Task 2), as well as their gender[4]. The goal was to choose, from each pair, the candidate who performed best in a third task (Task 3), which consisted of 7 questions of the same type as in Task 1 and 5 questions of the same type as in Task 2. Before making their hiring decisions, participants were shown information about the candidates' gender and performance across all three tasks for 20 randomly selected candidates.

[3]Dargnies et al. (2024) report a proportion of algorithm-averse managers of 54.8% in the baseline treatment and 53.4% in the transparency treatment, where managers received information about how the algorithm works.

[4]Participants were told that in each task, the candidates had two minutes to solve 12 real-effort exercises. Task 1 was the standard Raven matrices test. Task 2 involved counting the number of zeros in a 6 x 6 matrix.

The first, payoff-irrelevant part consisted of ten hiring decisions. For each hiring decision, participants were shown the number of correct answers achieved by both candidates in Tasks 1 and 2, along with their respective gender. The participants then had to choose which candidate to hire. After each decision, participants received feedback on their choice and were also informed which of the two candidates was actually better. Additionally, they were provided with information on which candidate from the same pair was selected by an AI agent and by a human agent.

Following Dargnies et al. (2024), participants were told that the AI agent's decisions stem from a computer program specifically developed for this hiring task. It was trained to predict candidates' performance in Task 3 based on their performance in Task 1, Task 2, and their gender, using data from at least 200 candidates. For each pair, the algorithm selects the candidate it predicts will perform best in Task 3.

Regarding the hiring decisions made by the human agent, the participants were told that those decisions came from another person who participated in a similar experiment conducted in the past and that the participants in that experiment had access to the same information, including the performance in Tasks 1, 2, and 3 of the same 20 randomly selected candidates, as well as their gender. The hiring decisions of the human agents used in our experiment stem from 10 participants from the experiment by Dargnies et al. (2024). The 10 participants were selected such that both the participant and the algorithm make the correct hiring decision 7 out of 10 times.[5] At the same time, the 10 participants made distinct decisions from AI agents either 4 or 6 out of 10 times.

In the second, payoff-relevant part, a participant's payoff was determined by the following payment rule:

$$\pi = 11 \text{ EUR} - X_i \text{ EUR} \cdot \text{number of } i\text{'s incorrect hiring decisions},$$

where $i = \{\text{own}, \textit{agent}\}$ and $\textit{agent} = \{\text{AI, Human}\}$. Since the payment was reduced by the penalty rate X_i for each incorrect hiring decision, the number of incorrect hiring decisions served as a measure of decision quality.

At the beginning of the second part, participants had to decide for both agents whether their own hiring decisions or the agent's hiring decisions should be used to determine their payoff. The participants made this decision in nine choice problems per agent (see Table 2, first three columns). For the case that the subject's own hiring decisions become payoff relevant, $X_{\text{own}} = 0.40$ remained constant across all choice problems. The monetary incentives for giving up control to the agent were varied: X_{agent} increases when going down the list of choice problems, taking one of the following values: {0.00; 0.10; 0.20; 0.30; 0.40; 0.50; 0.60; 0.70; 0.80}. Note that in the first choice problem, the agent's errors were payoff-irrelevant ($X_{agent} = 0$). In the fifth choice problem, the penalty rate $X_i = 0.40$ was the same regardless of whether the decision was made by the participant, the AI agent, or the human agent. As the penalty rate X_{agent} increases sufficiently, a payoff-maximizing participant should stop delegating and opt to make their own hiring decisions. For monotonic preferences, the number of delegation choices indicates a unique crossover point where participants shift from delegating decisions to the agent to relying on their own decisions. Note that a subject who only takes the size of the penalty rate into account would stop delegating when $X_{agent} > X_{\text{own}}$. Taking performance feedback from the first part into account, the predicted stopping

[5] The average number of correct hiring decisions in Dargnies et al. (2024) was 5.59 out of 10.

rule is:

$$X_{agent} > X_{\text{own}} \cdot \frac{\text{number of own incorrect hiring decisions}}{\text{number of agent's incorrect hiring decisions}}.$$

After participants decided on all choice problems, they had to make another 10 hiring decisions similar to those in the first part. However, the participants did not receive any feedback.

Experiment 2: Forecasting decisions

In Experiment 2, the task was to make non-binary forecasting decisions. We adopted the forecasting task that has been used in several experimental studies (e.g., Dietvorst et al., 2015; Logg et al., 2019; Jung and Seiter, 2021; Ivanova-Stenzel and Tolksdorf, 2024).[6] More specifically, participants were asked to determine the rank of a randomly chosen U.S. state in terms of departing air passengers in 2011 based on five pieces of information[7].

As in Experiment 1, in the first part, each participant made their own forecast in ten payoff-irrelevant rounds. Feedback consisted of the own performance, the performance of an AI agent, and the performance of another human. The participants were told that (i) the algorithm was specifically developed for this prediction task, and (ii) the forecasting decisions of the other human came from a similar experiment conducted in the past, where that person made forecasts based on the same information and also observed the algorithm's forecasts over the 10 training rounds.

The decisions of the human agent stem from 10 participants from Ivanova-Stenzel and Tolksdorf (2024), whose average performance (deviation from true rank) in their 10 training rounds was different from the algorithm's average performance by not more than .5 in absolute terms . Five out of the 10 participants slightly outperformed the algorithm and 5 participants were slightly outperformed by the algorithm. The observed average absolute deviation between the performance of the human and AI agents was between .2 and .5 (mean= .34, sd= .12).[8]

In the second, payoff-relevant part, as in Experiment 1, subjects had to decide which forecast should be used to determine the payoff in that part – their own forecast or the forecast stemming from the AI agent, respectively, the human agent whose decisions they observed in the first part. Participants made this decision in nine choice problems, see Table 2 (last three columns). As in Ivanova-Stenzel and Tolksdorf (2024), for each unit of deviation of the forecast from the true rank, the payoff was reduced by a penalty rate. More precisely, the payment rule was:

$$\pi = 7 \text{ EUR} - X_i \text{ EUR} \cdot |i\text{'s forecast} - \text{true rank}|,$$

where $i = \{\text{own}, agent\}$ and $agent = \{\text{AI, Human}\}$. Thus, the deviation of each forecast from the true rank is the measure of the precision of the forecast. As in Experiment 1, for the case that the subject's own forecast becomes payoff relevant, $X_{\text{own}} = 0.12$ remains constant across all choice problems. For the case that the forecast of either the AI or the human agent is used to determine the subject's payoff,

[6] Those studies typically find that in 40 to 60% of cases, participants choose their own decision over that of the algorithm.

[7] The information includes the number of major airports, the rank of population (2010), the rank of county, the rank of median household income (2008), the rank of expenditure on domestic travel (2009)

[8] The average performance of all participants in Ivanova-Stenzel and Tolksdorf (2024) was 6.75.

X_{agent} increases when going down the list of problems, taking one of the values {0; 0.03; 0.06; 0.09; 0.12; 0.15; 0.18; 0.21; 0.24}. In the first choice problem, the agent's forecasting precision was payoff-irrelevant ($X_{agent} = 0$). In the fifth choice problem, the penalty rate $X_i = 0.12$ was the same regardless of whether the decision was made by the participant, the AI agent, or the human agent.[9] When the agent's penalty rate X_{agent} increases enough, a person should stop delegating and cross over to opting for their own forecast. A subject who is only concerned with the size of the payment reduction would stop choosing the agent when $X_{agent} > X_{\text{own}}$. Taking performance feedback into account changes the stopping rule to:

$$X_{agent} > X_{\text{own}} \cdot \frac{|\text{own forecast}-\text{true rank}|}{|\text{agent's forecast}-\text{true rank}|},$$

where $|i\text{'s forecast} - \text{true rank}|$ is the individual average of i's performance in the first part. After the participants decided on all the choice problems, they had to make a forecast similar to those in the first part.

Table 2: The payoff structure of the nine paired choice problems.

Choice problem	Payoff in EUR					
	Experiment 1 (hiring task): $11 - X_i \cdot \#i$'s incorrect hiring decisions			Experiment 2 (forecasting task): $7 - X_i \cdot |i\text{'s forecast} - \text{true rank}|$		
	X_{own}	X_{agent}	$X_{agent} - X_{\text{own}}$	X_{own}	X_{agent}	$X_{agent} - X_{\text{own}}$
1	0.40	0.00	−0.40	0.12	0.00	−0.12
2	0.40	0.10	−0.30	0.12	0.03	−0.09
3	0.40	0.20	−0.20	0.12	0.06	−0.06
4	0.40	0.30	−0.10	0.12	0.09	−0.03
5	0.40	0.40	0	0.12	0.12	0
6	0.40	0.50	0.10	0.12	0.15	0.03
7	0.40	0.60	0.20	0.12	0.18	0.06
8	0.40	0.70	0.30	0.12	0.21	0.09
9	0.40	0.80	0.40	0.12	0.24	0.12
	Performance-based prediction†:					
	$X_{agent} > X_{\text{own}} \cdot \frac{\text{number of own incorrect hiring decisions}}{\text{number of agent's incorrect hiring decisions}}$			$X_{agent} > X_{\text{own}} \cdot \frac{|\text{own forecast}-\text{true rank}|}{|\text{agent's forecast}-\text{true rank}|}$		

†The performance-based prediction is based on the decisions in the first part; in Experiment 2, $|i\text{'s forecast} - \text{true rank}|$ is the individual average of i's performance.

Both studies were conducted at Technische Universität Berlin. In total, 95 subjects (49.5% female) participated in Experiment 1 and 95 subjects (45.3% female) participated in Experiment 2. Each experiment consisted of 5 sessions with 17 to 21 participants each. All participants were students from the local participant pool, mostly from economics, engineering, or the natural sciences. The average duration of a session was 45 minutes in both experiments. The average total earnings per participant amounted to 15.99 EUR (Experiment 1) and 12.33 EUR (Experiment 2), including a show-up fee of 6 EUR. [10]

[9]Note that the fifth choice problem corresponds to the choice subjects usually faced in previous studies, (e.g., Dietvorst et al., 2015; Jung and Seiter, 2021; Logg et al., 2019).

[10]See the Appendix for a translated version of the instructions for both experiments.

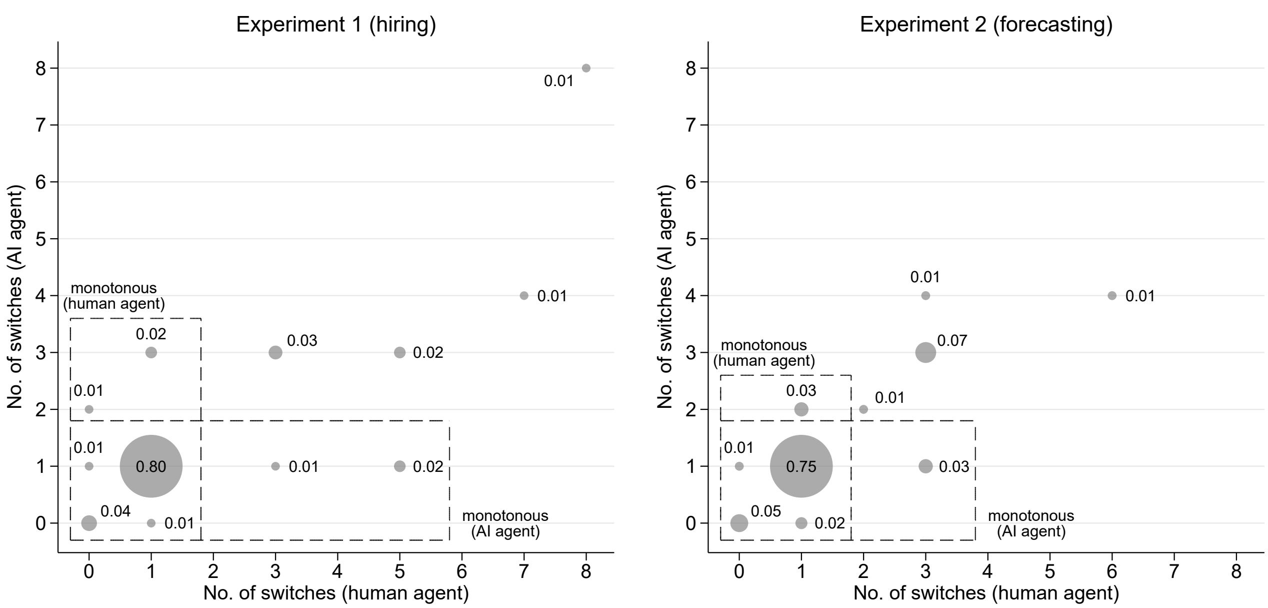
3. Results

In this section, we present the results from both experiments jointly. This approach is motivated by the shared experimental structure and core design features across the two studies, including the within-subject comparison of AI and human agents, the delegation framework, and the performance-based feedback. We begin by analyzing delegation choices, followed by confidence and belief measures, and a discussion of the robustness of our within-subject design.

Delegation choices

First, we assess the monotonicity of participants' delegation patterns. The overwhelming majority of participants (85% in Experiment 1 and 82% in Experiment 2) switched at most once in their delegation decisions to each agent type.[11] That is, they exhibited at most one switch between delegating and deciding for themselves for both the AI agent and the human agent.

Figure 2: Number of switches.



Given this high degree of monotonicity, we focus the main analysis on those participants who exhibit monotonic preferences. This allows for a straightforward interpretation of the number of delegation choices as the crossover point. [12]

Next, we examine the correlations of delegation choices. Kendall's tau coefficients show that delegation choices to the two agent types are highly positively correlated (Experiment 1: 0.44; Experiment 2: 0.46),

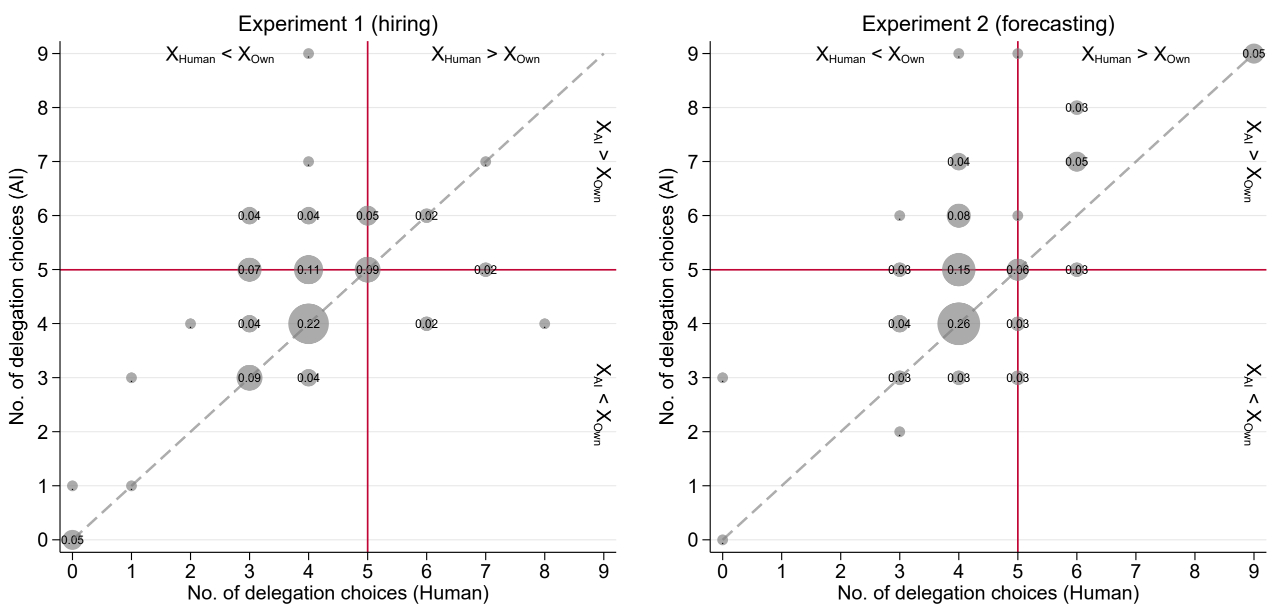
[11] 80% (75%) in Experiment 1 (2) switch exactly once. The largest fraction of the remaining participants, 3% (7%) in Experiment 1 (2), switch 3 times for each type of agent, which could indicate some uncertainty in the switching decision. See Figure 2 for the distributions of the number of switches.

[12] An alternative way to deal with non-monotonic preferences is to consider intervals of delegation choices, where the lower (upper) bound of an interval is determined by the first (last) switching point. The results remain qualitatively unchanged when using all observations in interval regressions as in Ivanova-Stenzel and Tolksdorf (2024).

and that delegation to AI and human agents is positively correlated with the respective performance-based predictions.[13] We find no correlation between delegation choices and the order in which agent types were encountered. Nonetheless, this variable is included as a control in the regression analyses. Delegation choices to each agent type are also negatively correlated with participants' confidence in their own performance. In Experiment 2 (forecasting), the delegation choices are positively correlated with the respective agent-specific confidence and belief measures.[14]

Figure 3 plots the relationship of the number of delegations (crossover point) by agent type for both experiments. In Experiment 1 (Experiment 2), 48% (38%) of the participants are in the lower-left quadrant, i.e., stopped choosing the agent for some $X_{agent} \leq X_{own}$. The distribution of delegation choices clusters above and along the 45-degree line, indicating that, across both experiments, participants generally favored AI over human agents when relinquishing control. In Experiment 1 (Experiment 2), 41% (47%) delegated more frequently to the AI agent, whereas only 10% (12%) delegated more frequently to the human agent. However, the distribution is somewhat more spread out in the forecasting task (Experiment 2), which may reflect greater variability of delegation to AI in analytical contexts.

Figure 3: Delegation choices (crossover points).



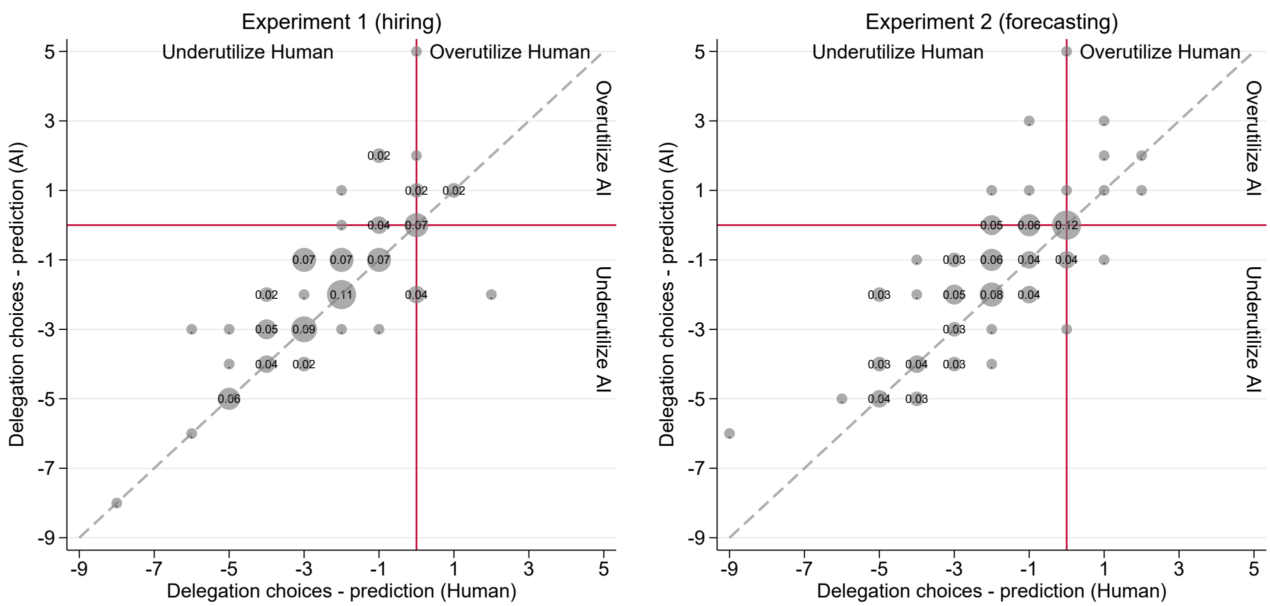
In Figure 4, we plot the difference between the observed crossover points and the performance-based predictions defined as the crossover points that would be payoff-optimal given the performance in the training rounds of both the participant and the AI/human agent (see the bottom row of Table 2). In both experiments, the AI/human agent significantly outperform the participants; see Table 5 in the Appendix.Observations in the lower left (upper right) quadrant represent participants who underutilize

[13]For Experiment 1, we also calculate an alternative performance-based prediction based on the 10 hiring decisions in the second payoff-relevant part. This measure is not correlated with delegation choices to either agent.

[14]See Tables 7 and 8 in the Appendix for the full Kendall's tau correlation matrices.

(overutilize) both types of agents. Observations in the upper left quadrant correspond to underutilization of the human agent and overutilization of the AI agent, while those in the lower right quadrant reflect the reverse pattern. We draw two main conclusions, both of which are consistent across experiments and independent of whether the decision environment is social or analytical. First, the majority of participants underutilize both types of agents. Second, there is a clear positive correlation in the agents' utilization. These results suggest that participants tend to exhibit consistent delegation behavior toward both agent types (AI and human), indicating that their choices are driven less by agent-specific characteristics and more by broader individual preferences regarding delegation and control.

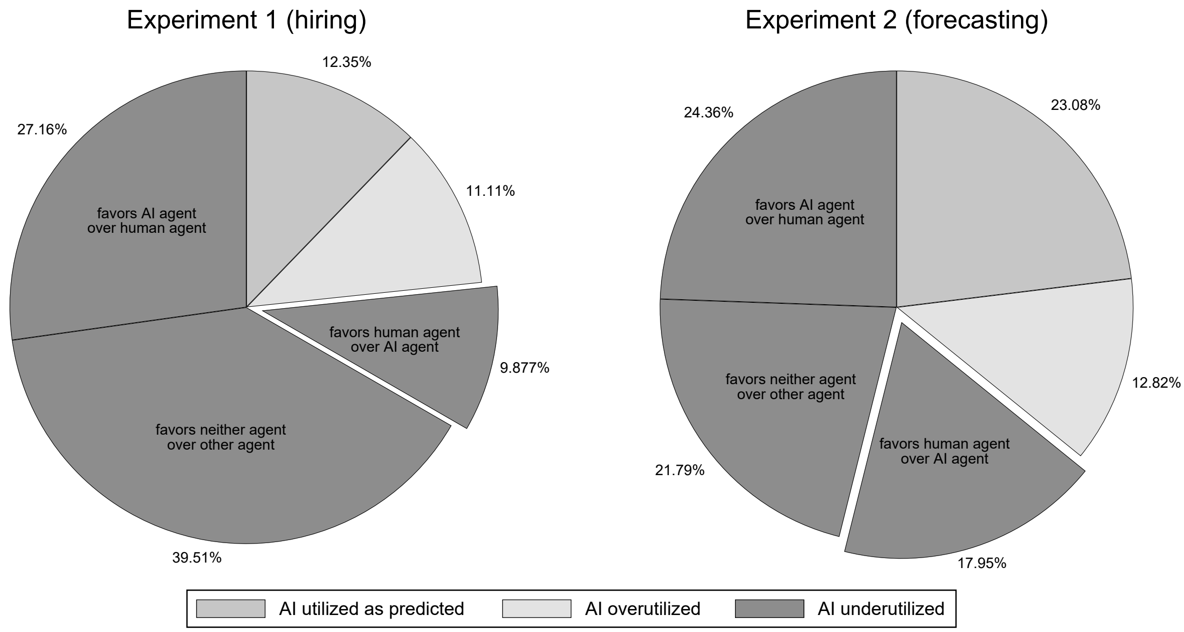
Figure 4: Delegation choices in relation to performance-based prediction.



Another way of demonstrating the heterogeneity in delegation preferences is the classification of participants based on their utilization of the AI agent (overutilization, utilization as predicted, underutilization). Figure 5 presents the results of this classification. The majority of participants underutilize the AI agent (77% in Experiment 1; 64% in Experiment 2). To check whether resistance to rely on algorithmic decision-making reflects distrust in AI per se, we categorize underutilization more finely by comparing how participants treat the AI agent versus the human agent. In particular, we distinguish between participants who favor the human agent over the AI agent, participants who favor the AI agent over the human agent, and participants who show no clear preference. The smallest proportion of those participants who underutilized the AI agent actually preferred the human agent over the AI agent. Thus, reluctance to delegate to AI agents in general does not necessarily indicate a specific distrust in AI. Consequently, individuals who decide not to rely on the decision of an AI agent must not necessarily be considered algorithm-averse. This finding is robust across different decision domains (social and analytical) and decision architectures (nature and number of decisions).

In Table 3, we show the main regression results that corroborate this finding. In all specifications, our main variable of interest is the treatment effect, i.e., whether the agent is an AI or a human (captured

Figure 5: Classification of AI utilization.



by the dummy variable *AI agent*, which takes the value 1 for the AI agent and 0 for the Human agent), while controlling for order (indicating whether the first agent was the AI agent), gender and risk aversion. The even specifications (2), (4), and (6) also include the performance-based prediction. The results in specifications (1) and (2) reveal that participants on average delegated around one half (Experiment 1) to two third (Experiment 2) more times to the AI agent compared to the Human agent. This trend is consistent when focusing on those participants who underutilized the AI agent (dark gray areas in Figure 3) in specifications (3) and (4), with around one third more delegations to AI agents compared with human agents in both experiments. Specifications (5) and (6) serve as robustness checks, restricting the sample to delegation choices to the agent encountered first (in a between-subjects manner). The direction and magnitude of the effects remain consistent. The significance of the performance-based prediction variable indicates that participants respond meaningfully to the performance feedback they received in the first part.[15]

To sum up, participants consistently delegate more to AI agents than to human agents, including those who might have been considered algorithm-averse in previous studies where participants could only delegate to an AI agent (e.g., Dietvorst et al., 2015, 2018; Jung and Seiter, 2021; Dargnies et al., 2024; Ivanova-Stenzel and Tolksdorf, 2024). This effect holds for both within-subject and between-subject comparisons. Moreover, delegation is meaningfully influenced by performance feedback.

[15] The regression results remain qualitatively the same when including further controls (specific to each experiment): whether a participant made more correct hires in the second part than in the first part (Experiment 1), and whether a participant observes that the Human agent made on average more accurate forecasting decisions that the AI agent (Experiment 2).

Table 3: Regression analysis of number of delegation choices.

	No. of delegation choices		No. of delegation choices (AI underutilized only)		No. of delegation choices (first agent only)	
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment 1 (hiring)						
AI agent	0.506***	0.506***	0.290*	0.290*	0.551*	0.625**
	(0.147)	(0.147)	(0.154)	(0.155)	(0.329)	(0.296)
Performance-based prediction		0.305***		0.485***		0.355***
		(0.094)		(0.114)		(0.096)
Order (AI first/Human second)	-0.065	0.025	0.153	0.021		
	(0.318)	(0.296)	(0.350)	(0.315)		
Constant	3.998***	2.240***	3.949***	1.012	4.235***	2.120**
	(0.640)	(0.836)	(0.714)	(0.826)	(0.659)	(0.823)
Gender and risk aversion	✓	✓	✓	✓	✓	✓
Observations	162	162	124	124	85	85
Experiment 2 (forecasting)						
AI agent	0.679***	0.653***	0.420***	0.386**	0.892**	0.873**
	(0.146)	(0.150)	(0.163)	(0.165)	(0.390)	(0.389)
Performance-based prediction		0.187*		0.186*		0.206*
		(0.096)		(0.098)		(0.108)
Order (AI first/Human second)	0.177	0.180	-0.388	-0.373		
	(0.346)	(0.344)	(0.330)	(0.320)		
Constant	3.474***	2.413***	3.415***	2.202***	3.837***	2.661***
	(0.661)	(0.757)	(0.529)	(0.759)	(0.698)	(0.885)
Gender and risk aversion	✓	✓	✓	✓	✓	✓
Observations	156	156	100	100	79	79

The dependent variable is the number of delegation choices. In specifications (5) and (6), all participants with monotonic preferences with respect to the agent type (AI/human) included. Reference category is Human agent. Estimation by random effects regression with standard errors clustered on subject level in specifications (1) to (4) and by OLS regression with robust standard errors in specifications (5) and (6). Standard errors in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Confidence and Beliefs

In Table 4, we explore the participants' confidence and beliefs as potential drivers for the observed delegation choices. In particular, we examine whether, and to what extent, participants differentiate between agents based on their nature (AI versus human).

Consistent over both experiments, that is, regardless of whether the decision environment is social or non-social and of the decision architecture (binary versus non-binary and repeated versus one-shot decisions), we observe that participants are less confident in the human agent both compared to the AI agent and to oneself (see specifications (1) and (2)). Moreover, participants believe that it is more likely that the AI agent might make perfect choices, both compared to oneself and compared to the human agent (see specifications (3) and (4)). We observe differences in the beliefs about the number of errors between the two experiments: In Experiment 1, participants believe that the AI agent makes significantly fewer incorrect hiring decisions than the human agent, but not compared to themselves. In Experiment 2, participants believe that the AI agent deviates significantly less from the true rank than both the participants themselves and the human agent (see specifications (5) and (6)). The observed different perception of the agents' error-proneness between the two experiments could be a reflection of the differences in domains (social vs. analytical) and nature of decisions (binary vs. non-binary; repeated vs. one-shot).[16]

Table 4: Regression analysis of confidence and belief measures.

	Confidence in choices		Belief about making no errors		Belief about error-proneness	
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment 1 (hiring)						
AI agent	-0.185	-0.185	0.457**	0.457**	-0.160	-0.160
	(0.147)	(0.147)	(0.199)	(0.200)	(0.144)	(0.144)
Human agent	-0.593***	-0.593***	-0.111	-0.111	0.185	0.185
	(0.121)	(0.122)	(0.136)	(0.137)	(0.133)	(0.133)
Performance-based prediction		-0.014		-0.300***		0.146***
		(0.038)		(0.093)		(0.050)
Order (AI first/Human second)	-0.137	-0.141	-0.146	-0.233	-0.122	-0.080
	(0.133)	(0.133)	(0.374)	(0.359)	(0.201)	(0.195)
Constant	2.445***	2.528***	2.732***	4.462***	3.766***	2.925***
	(0.206)	(0.337)	(0.516)	(0.732)	(0.259)	(0.381)
Gender and risk aversion	✓	✓	✓	✓	✓	✓
Observations	243	243	243	243	243	243
Postestimation Wald tests:						
H_0: AI agent = Human agent	$p < 0.01$	$p < 0.01$	$p = 0.01$	$p = 0.01$	$p < 0.01$	$p < 0.01$
Experiment 2 (forecasting)						
AI agent	0.167	0.167	0.731***	0.731***	-0.808***	-0.808***
	(0.146)	(0.146)	(0.198)	(0.199)	(0.254)	(0.254)
Human agent	-0.321**	-0.321**	0.064	0.064	0.013	0.013
	(0.128)	(0.129)	(0.138)	(0.139)	(0.256)	(0.257)
Performance-based prediction		-0.005		-0.035		0.022
		(0.037)		(0.113)		(0.132)
Order (AI first/Human second)	0.086	0.086	0.465	0.465	-0.210	-0.210
	(0.115)	(0.115)	(0.333)	(0.332)	(0.401)	(0.402)
Constant	2.057***	2.086***	3.695***	3.893***	5.554***	5.431***
	(0.208)	(0.305)	(0.542)	(0.871)	(0.711)	(1.026)
Gender and risk aversion	✓	✓	✓	✓	✓	✓
Observations	234	234	234	234	234	234
Postestimation Wald tests:						
H_0: AI agent = Human agent	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$

The dependent variable is the confidence in specifications (1) and (2), the belief about perfect hiring decisions in Experiment 1/a perfect forecast in Experiment 2 in specifications (3) and (4), the belief about the no. of wrong hires in Experiment 1/the deviation from the true rank in Experiment 2 in specifications (5) and (6). Only participants with monotonic preferences included. Reference category is the participants' self evaluation. Estimation by random effects regression with standard errors clustered on subject level in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

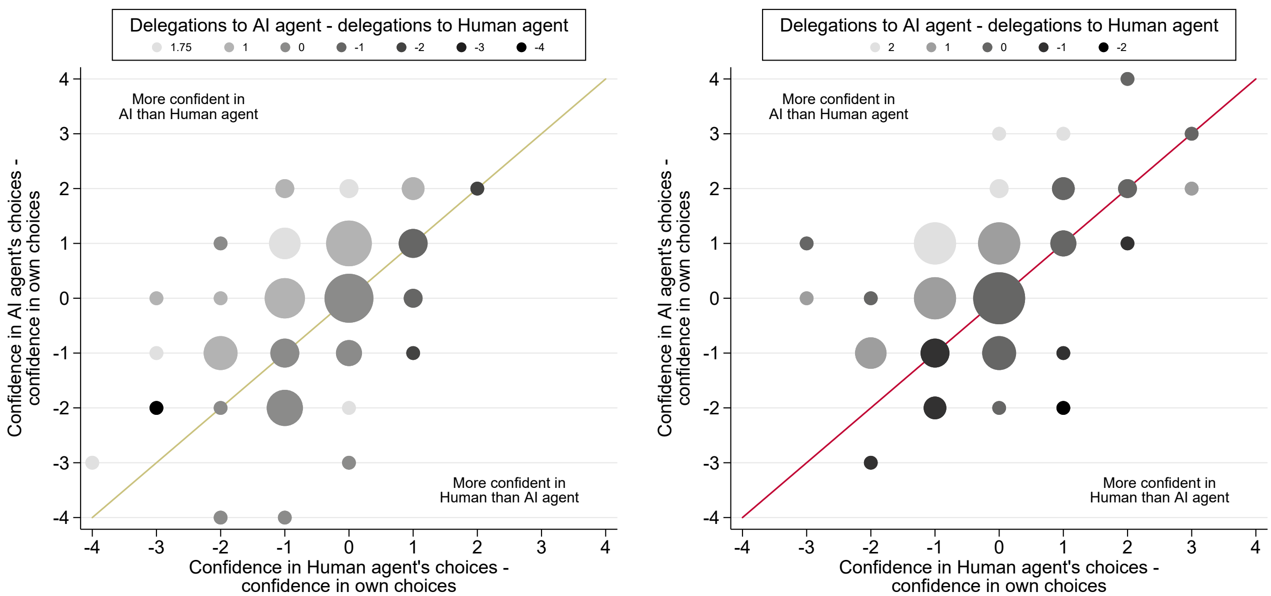
Figure 6 illustrates how participants' confidence relates to their delegation behavior. Specifically, it plots participants' confidence in the AI agent relative to themselves (y-axis) against their confidence in the human agent relative to themselves (x-axis), with the circles' size representing the number of participants in each combination of confidence levels, and the circles' colors indicating the difference in the number of delegation choices to the AI versus the human agent.

The plot reveals that participants who were more confident in the AI than the human agent relative to

[16] In Experiment 1 the measure of error-proneness captures the number of incorrect hiring decisions out of 10. In Experiment 2, the measure of inaccuracy is based on a single forecast's deviation from the true rank. These metrics may carry different cognitive weights or levels of interpretability for participants, potentially influencing how they assessed the agents' performance.

themselves (above the 45-degree line) were also more likely to delegate more to the AI than to the human agent. Those with higher confidence in the human agent tended to delegate more to the human, though fewer such cases are observed. These patterns were consistent across the hiring and forecasting task, across delegations to either agent type individually, and across all belief measures (the corresponding plots can be found in the Appendix). This suggests that delegation behavior is not arbitrary but systematically reflects participants' subjective assessments of agent competence and reliability, specific to each agent type (AI or human).

Figure 6: Relation of delegation preference and relative confidence.



Robustness of the Within-Subject Design

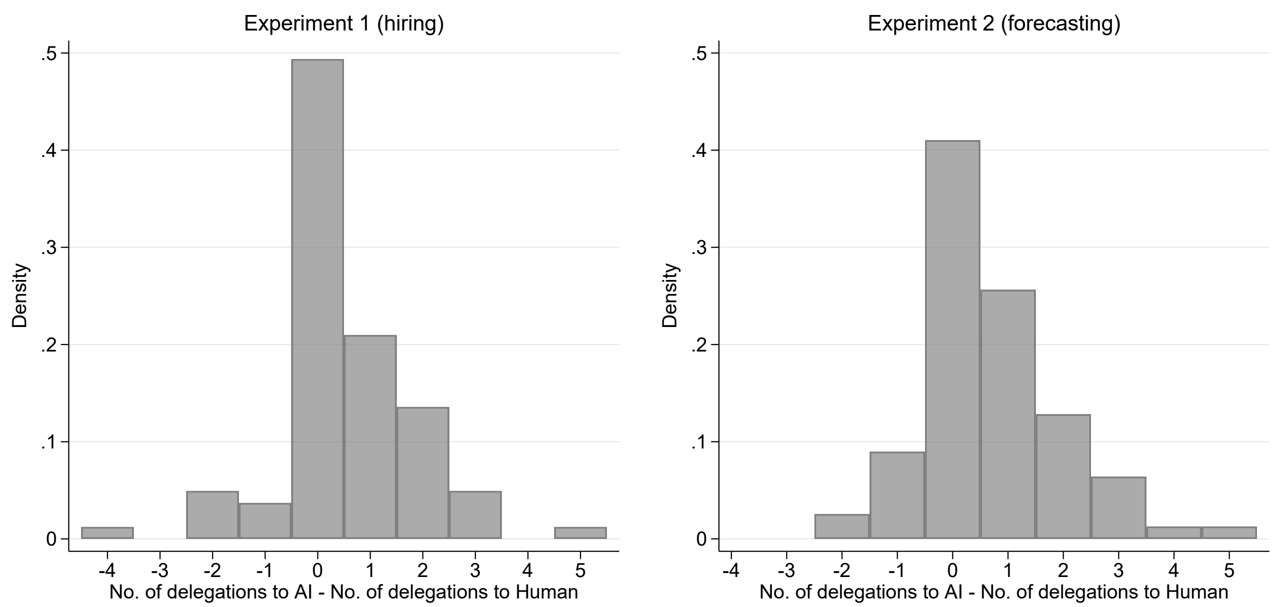
Our within-subject design enables a direct comparison of delegation behavior toward AI and human agents while holding all task features constant. It also allows for the precise identification of individual delegation preferences and the measurement of relative trust in AI decisions, conditional on general aversion to giving up control.

According to List (2025), to ensure internally valid estimates, a within-subject design should satisfy three assumptions that are not required in a between-subject design: balanced panel, temporal stability, and causal transience. Our design achieves a fully balanced panel with no attrition: all participants completed both conditions. The two conditions were implemented consecutively within the same session, with no break and no feedback, minimizing the risk of time-related shocks or learning effects. To assess the causal transience assumption, i.e., that delegation decisions to a given agent are unaffected by prior exposure to the other agent type, we test for agent-specific order effects. For each agent type (AI or human), we test whether it matters if participants encounter that agent first or second. We find no significant agent-specific order effects, suggesting the absence of carryover or anticipation effects (for the regression results, see Table 6 in the Appendix).

The absence of threats to internal validity supports the interpretability of between-subject comparisons derived from participants' responses to the agent encountered first.[17] This allows us to treat those observations as if they stem from a conventional between-subject setup, thereby validating between-subject inferences within our experiment. Recall that the regression results in Table 3 show that the direction and magnitude of the effects observed in specifications (1) and (2) for the full sample persist even when the analysis is restricted to a between-subject comparison, as in specifications (5) and (6).

A key advantage of the within-subject design is its ability to identify individual treatment effects. Figure 7 illustrates the distribution of the individual treatment effects for both experiments. For nearly half of the participants, the number of delegations to both AI and human agents is identical (Experiment 1: 49%; Experiment 2: 41%). The distributions are centered around zero but skewed slightly to the right, indicating a general tendency to favor AI over human agents in delegation. The similarity of these distributions across the different decision domains (social and non-social) and decision architectures (binary vs. non-binary and repeated vs. one-shot decisions) reinforces the robustness and generalizability of our findings.

Figure 7: Difference of delegations to agent types for Experiment 1 & Experiment 2.



4. Concluding Remarks

This paper examines whether reluctance to delegate to AI stems from a specific distrust in algorithms or reflects a broader resistance to ceding decision-making authority. Employing a within-subject design and varying the decision architecture (nature and number of decisions), we conduct two incentivized laboratory experiments, one focused on hiring (a social task) and the other on forecasting (an analytical

[17] In particular, anticipation effects may arise if awareness of encountering the agents in a sequence influences delegation decisions to the agent encountered first.

task). In the experiments, participants have the choice between making decisions themselves or delegate to an AI agent, respectively, a human agent. We find that participants systematically underutilize both AI and human agents, even when those agents outperform them. Despite a general hesitancy to delegate, we observe a clear preference for delegating to AI rather than human agents, a behavioral pattern that remains consistent across both decision domains and architectures.

This consistent tendency to prefer algorithmic delegation over human delegation, despite an overarching desire to retain control, extends previous evidence based on average treatment effects that individuals are more likely to delegate to AI than to other humans (Candrian and Scherer, 2022; Germann and Merkle, 2023; Holzmeister et al., 2023) by scrutinizing the individual-level trade-offs between autonomy and delegation to AI, respectively, other humans.

Recent comprehensive reviews of human-algorithm interactions highlight a consistent behavioral pattern. People are often willing to consider algorithmic advice, yet remain reluctant to fully cede control to automated systems (Chugunova and Sele, 2022). Our results not only corroborate this pattern but take it one step further by showing that reluctance to rely on algorithms may not primarily reflect distrust in algorithmic judgment but rather a fundamental discomfort with giving up decision-making authority.

Dietvorst et al. (2018) demonstrate that giving individuals partial control over algorithmic forecasts, such as allowing them to modify algorithmic output, significantly reduces algorithm aversion. Similarly, Fink et al. (2024) report that granting decision makers greater autonomy, such as control over the number or structure of recommendations, increases the reliance on algorithmic recommendation systems. Our findings offer a behavioral explanation for these observed effects, suggesting that algorithm aversion stems primarily from a broader aversion to relinquishing control rather than from specific distrust towards AI. Thus, the effectiveness of restoring control likely arises from addressing a fundamental discomfort with ceding decision-making authority, rather than focusing solely on enhancing trust in algorithmic predictions.

This broader perspective may also explain why increasing algorithm transparency has shown limited effectiveness in reducing algorithm aversion, for example, in hiring contexts (Dargnies et al., 2024) and in contexts involving (un)ethical judgments (Leib et al., 2024). If individuals are driven primarily by general reluctance to relinquish control rather than specific distrust in AI, then transparency alone, focused narrowly on increasing trust in AI, will likely fall short of overcoming this barrier. Consequently, our findings suggest that interventions aimed at enhancing algorithm appreciation might achieve greater success if they explicitly target the underlying autonomy and control concerns. This interpretation aligns with research on algorithmic advice utilization, where human decision makers maintain ultimate decision-making authority (e.g., Logg et al., 2019; Chacon et al., 2025; Greiner et al., 2025).

Last but not least, our findings align with and extend prior research on delegation and control preferences, as well as intrinsic preferences for autonomy. For example, Fehr et al. (2013) demonstrate that principals systematically underdelegate authority due to a preference for control and an aversion to being overruled, even when doing so would objectively improve outcomes. We extend this insight to AI contexts, highlighting that reluctance to relinquish decision-making authority is driven by intrinsic autonomy preferences and is largely independent of whether authority is delegated to a human or an AI. Similarly, Freundt et al. (2023) provide a theoretical framework and empirical evidence for intrinsic preferences for autonomy. Our findings can be interpreted as behavioral evidence of such preferences: individuals may resist delegating to AI not because they distrust algorithms but because they value making decisions themselves.

Our results have important implications for managers, policymakers, and designers of decision-support systems: Organizations seeking to increase AI adoption should focus not only on building trust in technology but also on addressing individuals' deeper reluctance to relinquish control. Educational interventions that normalize delegation may be more effective than simply highlighting algorithmic accuracy. Providing opportunities for decision-makers to experience the benefits of delegating to AI, such as interactive training or feedback-rich simulations, could help reduce reluctance to give up control over time. Rather than requiring a binary choice between human and algorithmic decisions, systems could enable flexible collaboration, such as human-AI hybrid models, override options, or adjustable levels of autonomy, that ease the psychological cost of delegation. Emphasizing how delegation can conserve cognitive resources, reduce error rates, and align with broader goals (e.g., fairness, efficiency) may help increase acceptance. Framing decisions as strategic choices about resource allocation, not as a loss of control, could shift perceptions.

Future research could investigate the conditions under which delegation aversion becomes more or less salient, and when algorithm aversion predominates. It would be particularly valuable to examine whether experience, training, or targeted framing interventions can mitigate resistance to external control and enhance the effective utilization of decision aids. It would also be valuable to examine how these patterns evolve among professional decision-makers, where delegation has more tangible consequences, and whether interventions that increase familiarity or reduce perceived personal accountability can sustainably shift delegation behavior. Additionally, exploring how team dynamics or hierarchical roles moderate these effects would offer valuable extensions.

Overall, our results not only advance our understanding of human-AI collaboration but also highlight the importance of autonomy and control considerations in designing effective decision-support systems. Future research and practical efforts aiming to increase the adoption of algorithmic tools should thus focus not merely on boosting trust in AI, but also and perhaps more critically, on addressing people's inherent preference for decision-making autonomy.

References

Barber, B. M. and Odean, T. (2001). The internet and the investor. *Journal of Economic Perspectives*, 15(1):41–54.

Bartling, B., Fehr, E., and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, 82(6):2005–2039.

Bonaccio, S. and Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2):127–151.

Burton, J. W., Stein, M.-K., and Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2):220–239.

Candrian, C. and Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous ai. *Computers in Human Behavior*, 134:107308.

Chacon, A., Kausel, E. E., Reyes, T., and Trautmann, S. (2025). Preventing algorithm aversion: People are willing to use algorithms with a learning label. *Journal of Business Research*, 187:115032.

Chugunova, M. and Sele, D. (2022). We and it: an interdisciplinary review of the experimental evidence on how humans interact with machines. *Journal of Behavioral and Experimental Economics*, 99:101897.

Dargnies, M.-P., Hakimov, R., and Kübler, D. (2024). Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence. *Management Science*.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114.

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

Fehr, E., Herz, H., and Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review*, 103(4):1325–1359.

Fink, L., Newman, L., and Haran, U. (2024). Let me decide: Increasing user autonomy increases recommendation acceptance. *Computers in Human Behavior*, 156:108244.

Freundt, J., Herz, H., and Kopp, L. (2023). Intrinsic preferences for autonomy.

Germann, M. and Merkle, C. (2023). Algorithm aversion in delegated investing. *Journal of Business Economics*, 93(9):1691–1727.

Greiner, B., Grünwald, P., Lindner, T., Lintner, G., and Wiernsperger, M. (2025). Incentives, framing, and reliance on algorithmic advice: An experimental study. *Management Science*.

Holzmeister, F., Holmén, M., Kirchler, M., Stefan, M., and Wengström, E. (2023). Delegation decisions in finance. *Management Science*, 69(8):4828–4844.

Ivanova-Stenzel, R. and Tolksdorf, M. (2024). Measuring preferences for algorithms–how willing are people to cede control to algorithms? *Journal of Behavioral and Experimental Economics*, 112:102270.

Jung, M. and Seiter, M. (2021). Towards a better understanding on mitigating algorithm aversion in forecasting: an experimental study. *Journal of Management Control*, 32(4):495–516.

Jussupow, E., Benbasat, I., and Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *ECIS Proceedings*, 168.

Leib, M., Köbis, N., Rilke, R. M., Hagens, M., and Irlenbusch, B. (2024). Corrupted by algorithms? how ai-generated and human-written advice shape (dis) honesty. *The Economic Journal*, 134(658):766–784.

Leyer, M. and Schneider, S. (2019). Me, you or ai? how do we feel about delegation.

List, J. A. (2025). The experimentalist looks within: Toward an understanding of within-subject experimental designs. Technical report, National Bureau of Economic Research.

Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103.

Longoni, C., Bonezzi, A., and Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4):629–650.

Mahmud, H., Islam, A. N., Ahmed, S. I., and Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting & Social Change*, 175.

Owens, D., Grossman, Z., and Fackler, R. (2014). The control premium: A preference for payoff autonomy. *American Economic Journal: Microeconomics*, 6(4):138–161.

Wiltermuth, S. S. and Neale, M. A. (2011). Too much information: The perils of nondiagnostic information in negotiations. *Journal of Applied Psychology*, 96(1):192.

Yaniv, I. and Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision processes*, 83(2):260–281.

Appendix

Table 5: Mean performance in Experiment 1 and Experiment 2 (standard deviations in parentheses).

	Mean performance (all participants)	Mean performance (only monotonic)
Experiment 1 (hiring)		
Payoff-irrelevant decisions		
# of own correct hiring decisions	5.63 (1.45)	5.64 (1.49)
# of AI agent's correct hiring decisions	7 (0)	7 (0)
# of human agent's correct hiring decisions	7 (0)	7 (0)
Payoff-relevant decisions		
# of own correct hiring decisions	6.47 (1.26)	6.60 (1.24)
# of AI agent's correct hiring decisions	7 (0)	7 (0)
# of human agent's correct hiring decisions	7 (0)	7 (0)
Experiment 2 (forecasting)		
Payoff-irrelevant forecasts		
$\lvert$own forecast − true rank$\rvert$	6.43 (2.96)	6.40 (3.04)
$\lvert$AI agent's forecast − true rank$\rvert$	3.99 (0.86)	3.97 (0.87)
$\lvert$human agent's forecast − true rank$\rvert$	4.05 (0.75)	4.02 (0.75)
Payoff-relevant forecast		
$\lvert$own forecast − true rank$\rvert$	7.11 (6.10)	7.59 (6.38)
$\lvert$AI agent's forecast − true rank$\rvert$	3.88 (3.81)	3.97 (3.85)
$\lvert$human agent's forecast − true rank$\rvert$	5.24 (4.04)	5.45 (4.08)

Performance between participants and agent (AI/human) significantly different ($p < 0.01$) based on two-sided t-test for all comparisons. Peformance between agents (AI vs. human) not significantly different ($p > 0.1$) for all comparisons.

Table 6: Test for violation of causal transience.

	No. of delegation choices (AI agent only)		No. of delegation choices (Human agent only)	
	(1)	(2)	(3)	(4)
Experiment 1 (hiring)				
Order (AI agent first/Human agent second)	-0.119	-0.044	0.067	0.144
	(0.346)	(0.336)	(0.354)	(0.322)
Performance-based prediction		0.269**		0.371***
		(0.111)		(0.089)
Constant	4.763***	3.204***	3.844***	1.632**
	(0.677)	(0.997)	(0.654)	(0.775)
Gender and risk aversion	✓	✓	✓	✓
Observations	84	84	84	84
Experiment 2 (forecasting)				
Order (AI agent first/Human agent second)	0.408	0.389	0.007	0.011
	(0.408)	(0.409)	(0.319)	(0.320)
Performance-based prediction		0.237*		0.138
		(0.120)		(0.104)
Constant	4.051***	2.707***	3.502***	2.712***
	(0.690)	(0.886)	(0.668)	(0.764)
Gender and risk aversion	✓	✓	✓	✓
Observations	82	82	81	81

The dependent variable is the number of delegation choices. Only participants with monotonic preferences with respect to the agent type (AI/human) included (see Figure 2). Reference category is Human agent. Estimation by OLS regression with robust standard errors in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 7: Correlation Table Experiment 1 (Hiring task)

Variable	1	2	3	4	5	6	7	8A	8B	8C	9A	9B	9C	9D	9E	9F
1. Delegation to AI agent	1.00															
2. Delegation to human agent	0.44^{c}	1.00														
3. Performance-based prediction (first part)†	0.23^{b}	0.33^{c}	1.00													
4. Performance-based prediction (second part)	-0.01	0.04	-0.01	1.00												
5. Order (human agent first)	-0.05	0.00	-0.06	-0.09	1.00											
6. Female participant	-0.05	0.11	0.11	0.07	-0.01	1.00										
7. Risk aversion	-0.00	-0.04	0.02	-0.19^{b}	-0.01	-0.19^{b}	1.00									
<u>8. Confidence in</u>																
A. forecast (own)	-0.30^{c}	-0.30^{c}	-0.18^{a}	-0.11	-0.03	-0.02	0.22^{b}	1.00								
B. forecast (AI agent)	0.22^{b}	0.05	0.14	-0.15	0.09	0.07	-0.16^{a}	0.07	1.00							
C. forecast (human agent)	-0.13	0.12	0.11	0.08	-0.27^{c}	0.25^{b}	-0.00	0.22^{b}	0.14	1.00						
<u>9. Belief about</u>																
A. # incorrect hiring decisions (own)	0.20^{b}	0.26^{c}	0.26^{c}	0.28^{c}	-0.05	0.01	-0.26^{c}	-0.50^{c}	0.02	-0.09	1.00					
B. # incorrect hiring decisions (AI agent)	-0.10	0.00	0.06	0.20^{b}	-0.24^{b}	0.10	0.00	-0.06	-0.36^{c}	0.08	0.33^{c}	1.00				
C. # incorrect hiring decisions (human agent)	0.08	0.03	0.12	0.24^{c}	0.15	-0.06	-0.09	-0.18^{a}	0.02	-0.35^{c}	0.47^{c}	0.25^{c}	1.00			
D. perfect prediction (own)	-0.09	-0.25^{c}	-0.28^{c}	0.03	-0.04	-0.15	0.10	0.44^{c}	-0.05	0.14	-0.33^{c}	-0.15^{a}	-0.19^{b}	1.00		
E. perfect prediction (AI agent)	0.10	-0.02	-0.13	0.02	0.01	-0.06	-0.02	0.19^{b}	0.22^{b}	0.11	-0.22^{b}	-0.38^{c}	-0.18^{b}	0.54^{c}	1.00	
F. perfect prediction (human agent)	-0.02	-0.01	-0.17^{a}	0.03	-0.05	-0.11	0.08	0.28^{c}	0.00	0.34^{c}	-0.16^{a}	-0.10	-0.30^{c}	0.73^{c}	0.51^{c}	1.00

Kendall's tau correlation of experimental variables. Only monotonic observations considered. ^{a}, ^{b} and ^{c} denote significance at the 10%, 5% and 1% level, respectively.
†Both agents' number of correct hiring decisions were the same.

Table 8: Correlation Table Experiment 2 (Forecasting task)

Variable	1	2	3	4	5	6	7	8A	8B	8C	9A	9B	9C	9D	9E	9F
1. Delegation to AI agent	1.00															
2. Delegation to human agent	0.46^{c}	1.00														
3. Performance-based prediction (AI agent)	0.22^{b}	0.18^{b}	1.00													
4. Performance-based prediction (human agent)	0.20^{b}	0.19^{b}	0.87^{c}	1.00												
5. Order (human agent first)	0.13	0.12	0.02	0.00	1.00											
6. Female participant	0.10	0.27^{c}	0.17^{a}	0.17^{a}	0.09	1.00										
7. Risk aversion	0.05	-0.03	0.09	0.04	0.01	-0.08	1.00									
<u>8. Confidence in</u>																
A. forecast (own)	-0.40^{c}	-0.40^{c}	-0.21^{b}	-0.25^{c}	-0.13	-0.19^{a}	0.17^{a}	1.00								
B. forecast (AI agent)	0.40^{c}	0.10	0.17^{a}	0.12	0.10	-0.05	0.12	-0.01	1.00							
C. forecast (human agent)	0.07	0.25^{b}	0.10	0.05	0.08	0.07	0.12	0.16	0.02	1.00						
<u>9. Belief about</u>																
A. deviation from true rank (own)	0.17^{a}	0.07	0.07	0.12	-0.03	-0.10	-0.18^{b}	-0.42^{c}	0.01	-0.08	1.00					
B. deviation from true rank (AI agent)	-0.41^{c}	-0.15	-0.16^{a}	-0.07	-0.08	-0.07	-0.16^{a}	0.04	-0.50^{c}	-0.01	0.38^{c}	1.00				
C. deviation from true rank (human agent)	-0.10	-0.26^{c}	-0.05	-0.01	-0.12	-0.07	-0.00	0.00	-0.06	-0.31^{c}	0.49^{c}	0.43^{c}	1.00			
D. perfect prediction (own)	-0.04	0.00	-0.12	-0.12	0.11	-0.00	0.11	0.47^{c}	0.15	0.26^{c}	-0.44^{c}	-0.10	-0.19^{b}	1.00		
E. perfect prediction (AI agent)	0.34^{c}	0.14	0.11	0.05	0.07	-0.04	0.04	-0.01	0.59^{c}	0.05	-0.06	-0.37^{c}	-0.13	0.41^{c}	1.00	
F. perfect prediction (human agent)	0.08	0.24^{b}	0.01	-0.01	0.14	0.08	0.05	0.19^{a}	0.08	0.52^{c}	-0.25^{c}	-0.07	-0.35^{c}	0.64^{c}	0.43^{c}	1.00

Kendall's tau correlation of experimental variables. Only monotonic observations considered. ^{a}, ^{b} and ^{c} denote significance at the 10%, 5% and 1% level, respectively.

Figure 8: Beliefs about making no errors and delegation choices.

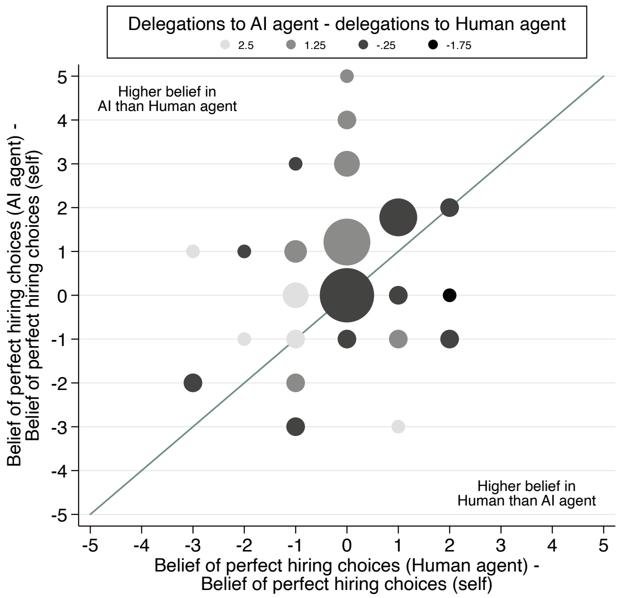
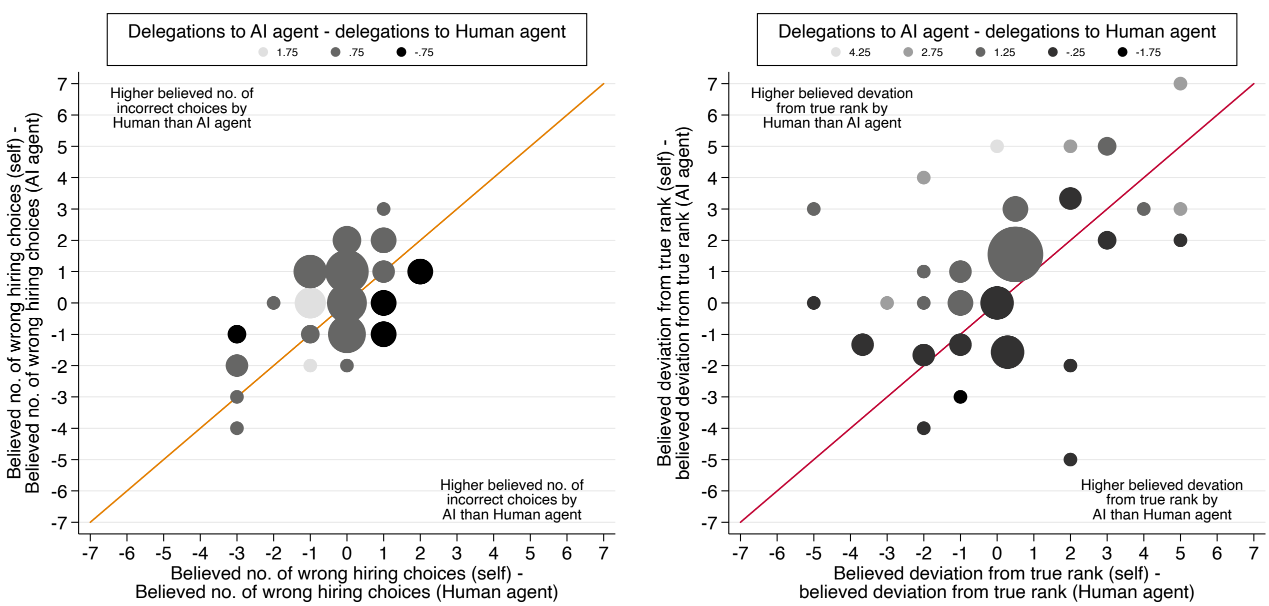


Figure 9: Beliefs about error-proneness and delegation choices.



Online-Appendix

Confidence and belief questions

Table 9: Confidence and belief questions in Experiment 1 (hiring)

How many incorrect hiring decisions do you think you have made? (0–10)
How many incorrect hiring decisions do you think the algorithm has made? (0–10)
How many incorrect hiring decisions do you think you the other person has made? (0–10)
How much confidence do you have in your own hiring decisions? (1 = *none*; 5 = *a lot*)
How much confidence do you have in the algorithm's hiring decisions? (1 = *none*; 5 = *a lot*)
How much confidence do you have in the other person's hiring decisions? (1 = *none*; 5 = *a lot*)
How likely do you think it is that you have made almost no incorrect hiring decisions? (1= *extremely unlikely*; 8= *extremely likely*)
How likely do you think it is that the algorithm has made almost no incorrect hiring decisions? (1= *extremely unlikely*; 8= *extremely likely*)
How likely do you think it is that the other person has made almost no incorrect hiring decisions? (1= *extremely unlikely*; 8= *extremely likely*)

Table 10: Confidence and belief questions in Experiment 2 (forecasting)

How many ranks do you think your estimate is away from the state's true rank? (0–50)
How many ranks do you think the other participant's estimate is away from the state's true rank? (0–50)
How many ranks do you think the AI's estimate is away from the state's true rank? (0–50)
How much confidence do you have in your estimate? (1 = *none*; 5 = *a lot*)
How much confidence do you have in the other participant's estimate? (1 = *none*; 5 = *a lot*)
How much confidence do you have in the AI's estimate? (1 = *none*; 5 = *a lot*)
How likely is it that you predicted the state's rank almost perfectly? (1= *extremely unlikely*; 8= *extremely likely*)
How likely is it that the other participant predicted the state's rank almost perfectly? (1= *extremely unlikely*; 8= *extremely likely*)
How likely is it that the AI predicted the state's rank almost perfectly? (1= *extremely unlikely*; 8= *extremely likely*)

Instructions

The following instructions were translated from German. The original versions are available from the authors upon request. Below, we provide one set of instructions, where we indicate the differences between the two conditions and other information not visible to participants in brackets with cursive font.
The instructions for the two participants' groups differ with respect to the order of the two conditions. Subjects in the first group participated in the "AI"-condition first followed by the "Human"-condition. Subjects in the second group participated in the "Human"-condition first followed by the "AI"-condition.

In the following, we present the instructions for the first group ("AI"-condition first).

Welcome to the experiment and thank you for participating!
General information

Please read these instructions carefully. If there is something you do not understand, please raise your hand. We will then come to you and answer your questions privately.

You will make your decisions at the computer.

All decisions will remain anonymous. That means you will not know the identity of the other participants and no participant will know your identity.

For simplification, the instructions are given in the masculine form.

Your earnings in this experiment (i.e., the sum of your earnings from both parts) will be paid to you privately and in cash at the end of the experiment.

You will receive 6 EUR for showing up on time.

Experiment 1 (hiring):

Your task is to make several hiring decisions. For each decision, you will select one of two candidates to hire. The candidates completed three tasks as participants in another experiment. You will receive information about the candidates' performance in Task 1 and Task 2, as well as their gender. Your goal is to select, from each pair, the candidate who performed best in Task 3. Task 3 consists of 7 questions of the same type as Task 1 and 5 questions of the same type as Task 2. For each task, the candidates had 1.5 minutes to answer as many questions as possible.

First, you will be shown all the questions from Tasks 1, 2, and 3 in sequence. For each task, you will have up to 30 seconds to familiarize yourself with the questions. Before starting the hiring decisions, we will present you with 20 randomly selected candidates, including information about their gender and performance in Tasks 1, 2, and 3. This will allow you to understand which candidates you may want to hire and which characteristics are relevant for the performance in Task 3. You will have up to 2 minutes to review this information.

The experiment consists of two parts, each with 10 hiring decisions.

In the first, payoff-irrelevant part, you will gain experience with the hiring task. For each of the 10 hiring decisions, you will choose between two candidates. You will first be shown the number of correct answers both candidates achieved in Tasks 1 and 2, as well as their respective gender. Then, you will select the candidate you want to hire. The goal is to hire the candidate who performed best in Task 3. After each decision, you will receive feedback on the choice you made and which of the two candidates was actually better. Additionally, you will see information about which candidate was selected by an algorithm and which was selected by another person for the same candidate pair.

The algorithm was specifically developed for this hiring task. It was trained to predict the candidates' performance in Task 3 based on their performance in Task 1, Task 2, and their gender. The algorithm was trained using the performance data of at least 200 candidates. For each pair, the algorithm selects the candidate it predicts will perform best in Task 3.

The hiring decisions made by the other person come from a similar experiment conducted in the past. This person had access to the same information as you, including the performance of the same 20 randomly selected candidates in Tasks 1, 2, and 3, as well as their gender.

In the second, payoff-relevant part, before selecting candidates, you will make decisions for two situations (A and B) across 9 decision problems each:

- Situation A (*"AI"-condition*): For each of the 9 decision problems, you decide whether you want to make the 10 hiring decisions yourself or use the hiring decisions made by the algorithm.
- Situation B (*"Human"-condition*): For each of the 9 decision problems, you decide whether you want to make the 10 hiring decisions yourself or use the hiring decisions made by the other person.

If you make the 10 hiring decisions yourself, your own hiring decisions will determine your payment. If you delegate the 10 hiring decisions to the algorithm or the other person, the hiring decisions made by the algorithm or the other person will determine your payment.

The payment formula in both situations, A and B, is as follows:

$$11€ - \mathbf{X} \times \text{Number of incorrect hiring decisions}$$

This means that for each incorrect hiring decision, your payment will be reduced by **X**.

- **X** = 0.40€ for all 9 decision problems if you decide to make the hiring decisions yourself.
- **X** takes one of the values {0€; 0.10€; 0.20€; 0.30€; 0.40€; 0.50€; 0.60€; 0.70€; 0.80€} in the 9 decision problems if you delegate the hiring decisions to the algorithm (Situation A) or the other person (Situation B).

The screen displaying all 9 decision problems in Situation A is as follows:

Situation A: im Folgenden treffen Sie für 9 Entscheidungsprobleme die Wahl, ob **Ihre eigenen Einstellungsentscheidungen** **oder** **die Einstellungsentscheidungen des Algorithmus** für Ihre Auszahlung relevant sind.

Die Auszahlungsformel lautet: 11€ - **X** * *Anzahl falscher Einstellungsentscheidungen*. Das heißt, für jede falsche Einstellungsentscheidung wird die Auszahlung um **X** reduziert.

Nehmen Sie sich genügend Zeit für Ihre Entscheidungen. Da Sie nicht wissen, welches der 9 Entscheidungsprobleme für Ihre Auszahlung in diesem Teil relevant sein wird, ist es optimal für Sie, sich so zu entscheiden, als ob jedes Entscheidungsproblem Ihre Auszahlung bestimmt.

	Auszahlung bei Wahl **Ihrer eigenen Einstellungsentscheidungen**	**Ihre Wahl:**			Auszahlung bei Wahl **der Einstellungsentscheidungen des Algorithmus**
1.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0€** * *Anzahl falscher Einstellungsentscheidungen*
2.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.10€** * *Anzahl falscher Einstellungsentscheidungen*
3.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.20€** * *Anzahl falscher Einstellungsentscheidungen*
4.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.30€** * *Anzahl falscher Einstellungsentscheidungen*
5.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*
6.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.50€** * *Anzahl falscher Einstellungsentscheidungen*
7.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.60€** * *Anzahl falscher Einstellungsentscheidungen*
8.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.70€** * *Anzahl falscher Einstellungsentscheidungen*
9.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen des Algorithmus	**11€ - 0.80€** * *Anzahl falscher Einstellungsentscheidungen*

The screen displaying all 9 decision problems in Situation B is as follows:

<u>Situation B:</u> im Folgenden treffen Sie für 9 Entscheidungsprobleme die Wahl, ob **Ihre eigenen Einstellungsentscheidungen <u>oder</u> die Einstellungsentscheidungen der anderen Person** für Ihre Auszahlung relevant sind.

Die Auszahlungsformel lautet: 11€ - **X** * *Anzahl falscher Einstellungsentscheidungen*. Das heißt, für jede falsche Einstellungsentscheidung wird die Auszahlung um **X** reduziert.

Nehmen Sie sich genügend Zeit für Ihre Entscheidungen. Da Sie nicht wissen, welches der 9 Entscheidungsprobleme für Ihre Auszahlung in diesem Teil relevant sein wird, ist es optimal für Sie, sich so zu entscheiden, als ob jedes Entscheidungsproblem Ihre Auszahlung bestimmt.

	Auszahlung bei Wahl **Ihrer eigenen Einstellungsentscheidungen**	**Ihre Wahl:**			Auszahlung bei Wahl **der Einstellungsentscheidungen der anderen Person**
1.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0€** * *Anzahl falscher Einstellungsentscheidungen*
2.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.10€** * *Anzahl falscher Einstellungsentscheidungen*
3.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.20€** * *Anzahl falscher Einstellungsentscheidungen*
4.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.30€** * *Anzahl falscher Einstellungsentscheidungen*
5.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*
6.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.50€** * *Anzahl falscher Einstellungsentscheidungen*
7.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.60€** * *Anzahl falscher Einstellungsentscheidungen*
8.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.70€** * *Anzahl falscher Einstellungsentscheidungen*
9.	**11€ - 0.40€** * *Anzahl falscher Einstellungsentscheidungen*	Ihre eigenen Einstellungsentscheidungen	○ ○	die Einstellungsentscheidungen der anderen Person	**11€ - 0.80€** * *Anzahl falscher Einstellungsentscheidungen*

Your decision in each situation is only valid once you have made a selection for all 9 decision problems (i.e., for each row) of that situation and then clicked the "Confirm" button at the bottom of the screen.

Afterward, you will make the 10 hiring decisions. As before, for each decision, you will see the number of correct answers of both candidates in Tasks 1 and 2, as well as their gender, and you must select one of the two candidates. Unlike before, you will not receive feedback about which candidate was actually better, nor will you know what decision the algorithm or the other person made for the same pair of candidates.

Your earnings will be determined as follows: First, the computer will randomly choose whether Situation A or Situation B is relevant for payment. Then, the computer will randomly draw a number between 1 and 9. This random number determines the row and thus the payment-relevant decision problem for the situation you chose:

- Situation A: If you decided in this row to use the algorithm's hiring decisions, the algorithm's decisions will determine your payment. If you decided to make the decisions yourself, your own hiring decisions will be used for the payment.
- Situation B: If you decided in this row to use the other person's hiring decisions, their decisions will determine your payment. If you decided to make the decisions yourself, your own decisions will determine the payment.

You will only make your decisions **once**. The random selection of the payoff-relevant situation and the drawing of the random number, which determines the payoff-relevant decision problem for this situation, will happen at the end of the experiment.

Take your time with your decisions. Since you do not know which situation or which of the 9 decision problems will be relevant for your payment, it is optimal for you to make decisions as though each decision problem in each situation will determine your payment.

Example 1:
Suppose the computer randomly selected Situation A and number 3. The number of incorrect hiring decisions made by you is 5, and the number of incorrect hiring decisions made by the algorithm is also 5.

Case 1:

You decided in row 3 that you will make the hiring decisions yourself. In this case, you will be deducted $0.40€ * 5 = 2€$. Your earnings will be 9€.

Case 2:

You decided in row 3 to use the algorithm's hiring decisions. In this case, you will be deducted $0.20€ * 5 = 1€$, and your earnings will be 10€.

Example 2:

Suppose the computer randomly selected Situation A and number 7. The number of incorrect hiring decisions made by you is 5, and the number of incorrect hiring decisions made by the algorithm is also 5.

Case 1:

You decided in row 7 that you will make the hiring decisions yourself. In this case, you will be deducted $0.40€ * 5 = 2€$. Your earnings will be 9€.

Case 2:

You decided in row 7 to use the algorithm's hiring decisions. In this case, you will be deducted $0.60€ * 5 = 3€$, and your earnings will be 8€.

Example 3:

Suppose the computer randomly selected Situation B and number 3. The number of incorrect hiring decisions made by you is 5, and the number of incorrect hiring decisions made by the other person is also 5.

Case 1:

You decided in row 3 that you will make the hiring decisions yourself. In this case, you will be deducted $0.40€ * 5 = 2€$. Your earnings will be 9€.

Case 2:

You decided in row 3 to use the other person's hiring decisions. In this case, you will be deducted $€0.20 * 5 = €1$, and your earnings will be 10€.

Example 3:

Suppose the computer randomly selected Situation B and number 7. The number of incorrect hiring decisions made by you is 5, and the number of incorrect hiring decisions made by the other person is also 5.

Case 1:

You decided in row 7 that you will make the hiring decisions yourself. In this case, you will be deducted $0.40€ * 5 = 2€$. Your earnings will be 9€.

Case 2:

You decided in row 7 to use the other person's hiring decisions. In this case, you will be deducted $0.60€ * 5 = 3€$, and your earnings will be 8€.

Control questions

Please answer the following questions. Raise your hand when you are finished. An experimenter will come to you and check your answers.

Determine a person's payout for the following example situations.

Example situation	**Payoff**
Suppose the computer has selected situation A and the number 1. In row 1, the person has decided to use the hiring decisions of the algorithm. The number of wrong hiring decisions is 3.	€
Suppose the computer has selected situation A and the number 5. In row 5, the person has decided to use the hiring decisions of the algorithm. The number of wrong hiring decisions is 3.	€
Suppose the computer has selected situation B and the number 5. In row 5, the person has decided to use the hiring decisions of the other person. The number of wrong hiring decisions is 3.	€

Experiment 2 (forecasting):

Your task is to make a forecast. More specifically, you need to determine the rank of a randomly chosen U.S. state in terms of the number of airline passengers that departed from that state in 2011. This rank can range from 1 to 50. Rank 1 means that most passengers have departed from this state. Rank 50 means that the fewest passengers have departed from this state.

There are five different pieces of information available to help you.

Number of major airports:
Airports that had a share of annual departing passengers of at least 1% or more of the United States.
Example: If a total of 1,000,000 passengers departed the U.S. in a year, then at least 10,000 passengers departed from that airport.

Rank in census population count in 2010:
The state with the largest population is ranked 1st, and the state with the lowest population ranks 50th.

Rank in number of counties:
States are sorted by the number of counties. A county is a territorial unit below the state. Rank 1 means the state has the highest number of counties. Rank 50 means the state has the lowest number of counties.

Rank in median household income in 2008:
States are sorted by the amount of median household income. Median household income in a state is defined so that the value is exactly in the middle of the series, ordered by size. This means that 50% have a lower household income and 50% have a higher household income. Rank 1 means that households in this state have the highest median household income. Rank 50 means that households in this state have the lowest median household income.

Rank in domestic travel expenditure in 2009:
States are sorted by spending on domestic travel. Rank 1 means that in this state spending on domestic travel was the highest. Rank 50 means that in this state spending on domestic travel was the lowest.

Part 1 consists of 11 rounds.

In the first 10, payoff-irrelevant, rounds, you gain experience with the forecasting task. In each round, for a randomly selected U.S. state, you will first be shown the five pieces of information mentioned above. Then, you will forecast of the rank of the selected U.S. state in terms of the number of airline passengers that departed from that state in 2011. You will then receive feedback comprising this information, your forecast, and the true rank of the state. In addition, you will receive information about the ranking predicted by an algorithm as well as by another person for this task. The algorithm was specifically developed for this prediction task and uses only the five pieces of information mentioned above, which are also available to you. The other person also used only the five pieces of information mentioned above for their prediction. The predictions from the other person come from a similar experiment conducted in the past.

In the **eleventh**, payoff-relevant round, you have to make a decision for two situations (A and B) for nine decision problems each:

- Situation A (*"AI"-condition*): For each of the nine decision problems, you decide whether you want to use the algorithm's prediction or make the prediction yourself.
- Situation B (*"Human"-condition*): For each of the nine decision problems, you decide whether you want to use the other person's prediction or make the prediction yourself.

If you decide to make the prediction yourself, your own prediction will be used to determine your payment. If you choose the prediction of the algorithm or the other person, the algorithm's prediction or the other person's prediction will be used to determine your payment.
The payment in both situations is calculated by the formula:

$$7€ - \mathbf{X} * |\textit{forecast} - \textit{true rank}|$$

That is, for each unit that the forecast deviates from the true rank, the payment is reduced by X.

- The *forecast* is your chosen rank or the selected rank of the algorithm (situation A) respectively the selected rank of the other person (station B) for the state in the forecasting task.
- The *true rank* is the actual rank of the state.
- $\mathbf{X} = 0.12€$ in all nine choice problems, if you decide to make the forecast yourself.
- $\mathbf{X}$ takes one of the values $\{0€; 0.03€; 0.06€; 0.09€; 0.12€; 0.15€; 0.18€; 0.21€; 0.24€\}$ in the nine choice problems if you let the algorithm (situation A) respectively the other person (situation B) make the forecast.

The screen with all nine choice problems in situation A looks like this:

Runde 11 von 11

Situation A: im Folgenden treffen Sie für 9 Entscheidungsprobleme die Wahl, ob **Ihre eigene Vorhersage** oder die **Vorhersage des Algorithmus** für Ihre Auszahlung relevant ist.

Die Auszahlungsformel lautet: 7€ – **X** * |*Prognose – wahrer Rang*|. Das heißt, für jede Einheit, die die Prognose vom wahren Rang abweicht, wird die Auszahlung um **X** reduziert.

Nehmen Sie sich genügend Zeit für Ihre Entscheidungen. Da Sie nicht wissen, welches der 9 Entscheidungsprobleme für Ihre Auszahlung in diesem Teil relevant sein wird, ist es optimal für Sie, sich so zu entscheiden, als ob jedes Entscheidungsproblem Ihre Auszahlung bestimmt.

	Auszahlung bei Wahl **Ihrer eigenen Vorhersage**	**Bitte treffen Sie Ihre Entscheidungen:**		Auszahlung bei Wahl der **Vorhersage des Algorithmus**
1.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0€** * |*Prognose – wahrer Rang*|
2.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.03€** * |*Prognose – wahrer Rang*|
3.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.06€** * |*Prognose – wahrer Rang*|
4.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.09€** * |*Prognose – wahrer Rang*|
5.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.12€** * |*Prognose – wahrer Rang*|
6.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.15€** * |*Prognose – wahrer Rang*|
7.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.18€** * |*Prognose – wahrer Rang*|
8.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.21€** * |*Prognose – wahrer Rang*|
9.	7€ – **0.12€** * |*Prognose – wahrer Rang*|	Ihre eigene Vorhersage	Vorhersage des Algorithmus	**7€ – 0.24€** * |*Prognose – wahrer Rang*|

Bestätigen

The screen with all nine choice problems in situation B looks like this:

Runde 11 von 11

Situation B: Im Folgenden treffen Sie für 9 Entscheidungsprobleme die Wahl, ob **Ihre eigene Vorhersage** oder die **Vorhersage der anderen Person** für Ihre Auszahlung relevant ist.

Die Auszahlungsformel lautet: 7€ − **X** * |*Prognose − wahrer Rang*|. Das heißt, für jede Einheit, die die Prognose vom wahren Rang abweicht, wird die Auszahlung um **X** reduziert.

Nehmen Sie sich genügend Zeit für Ihre Entscheidungen. Da Sie nicht wissen, welches der 9 Entscheidungsprobleme für Ihre Auszahlung in diesem Teil relevant sein wird, ist es optimal für Sie, sich so zu entscheiden, als ob jedes Entscheidungsproblem Ihre Auszahlung bestimmt.

	Auszahlung bei Wahl **Ihrer eigenen Vorhersage**	**Bitte treffen Sie Ihre Entscheidungen:**			Auszahlung bei Wahl der **Vorhersage der anderen Person**
1.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0€** * |*Prognose − wahrer Rang*|
2.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.03€** * |*Prognose − wahrer Rang*|
3.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.06€** * |*Prognose − wahrer Rang*|
4.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.09€** * |*Prognose − wahrer Rang*|
5.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.12€** * |*Prognose − wahrer Rang*|
6.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.15€** * |*Prognose − wahrer Rang*|
7.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.18€** * |*Prognose − wahrer Rang*|
8.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.21€** * |*Prognose − wahrer Rang*|
9.	7€ − **0.12€** * |*Prognose − wahrer Rang*|	Ihre eigene Vorhersage	◯ ◯	Vorhersage der anderen Person	7€ − **0.24€** * |*Prognose − wahrer Rang*|

Bestätigen

Your decision in each situation is only valid once you have made a selection for all choice problems (i.e., for each row) in that situation and then clicked on the "Confirm" button at the bottom of the screen.

After that, as in the previous rounds, you will receive the five pieces of information and must make your own forecast for the given information.

Your earnings are determined as follows:

First, the computer randomly selects whether situation A or situation B is payoff-relevant. After that, the computer randomly draws a number between 1 and 9. This random number determines the row and thus the payout-relevant choice problem from the previously selected situation:

- Situation A: If you decided in this row to use the algorithm's prediction, the algorithm's prediction will determine your payment. If you decided to make the prediction yourself, your own prediction will be used for the payment.
- Situation B: If you decided in this row to use the other person's prediction, the other person's prediction will determine your payment. If you decided to make the prediction yourself, your own prediction will be used for the payment.

You make your decisions only once. The random selection of the payoff-relevant situation and the drawing of the random number, which determines the payoff-relevant decision problem for that situation, takes place at the end of the experiment.

Take your time making your decisions. Since you do not know which of the two situations and which of the nine decision problems will be relevant for your payment, it is optimal for you to decide as if each decision problem in each situation determines your payment."
Below are some examples of how to determine your payment. The numbers chosen are fictional.

Example 1:
Suppose the computer randomly selected situation A and the number 3, which is the choice problem in the third row of the table. The prediction (yours or that of the algorithm) is rank 10, the true rank is 5. The deviation of the forecast from the true rank, i.e., the absolute value of *forecast−true rank*, is thus 5.
Case 1:
You decide in line 3 to make the forecast yourself. In that case, you will be deducted 0.12€·5 = 0.60€. Your earnings are 6.40€.
Case 2:
You decide in line 3 that the algorithm will make the forecast. In that case, you will be deducted 0.06€·5 = 0.30€. Your earnings are 6.70€.

Example 2:
Suppose the computer randomly selected situation A and the number 7, which is the choice problem in the 7th row of the table. The prediction (yours or that of the algorithm) is rank 5, the true rank is 10. The deviation of the forecast from the true rank, i.e., the absolute value of *forecast−true rank*, is thus 5.
Case 1:
You decide in line 7 to make the forecast yourself. In that case, you will be deducted $0.12€\cdot 5 = 0.60€$. Your earnings are 6.40€.
Case 2:
You decide in line 7 that the algorithm will make the forecast. In that case, you will be deducted $0.18€\cdot 5 = 0.90€$. Your earnings are 6.10€.

Example 3:
Suppose the computer randomly selected situation B and the number 3, which is the choice problem in the third row of the table. The prediction (yours or that of the other person) is rank 10, the true rank is 5. The deviation of the forecast from the true rank, i.e., the absolute value of *forecast−true rank*, is thus 5.
Case 1:
You decide in line 3 to make the forecast yourself. In that case, you will be deducted $0.12€\cdot 5 = 0.60€$. Your earnings are 6.40€.
Case 2:
You decide in line 3 that the algorithm will make the forecast. In that case, you will be deducted $0.06€\cdot 5 = 0.30€$. Your earnings are 6.70€.

Example 4:
Suppose the computer randomly selected situation B and the number 7, which is the choice problem in the 7th row of the table. The prediction (yours or that of the other person) is rank 5, the true rank is 10. The deviation of the forecast from the true rank, i.e., the absolute value of *forecast−true rank*, is thus 5.
Case 1:
You decide in line 7 to make the forecast yourself. In that case, you will be deducted $0.12€\cdot 5 = 0.60€$. Your earnings are 6.40€.
Case 2:
You decide in line 7 that the algorithm will make the forecast. In that case, you will be deducted $0.18€\cdot 5 = 0.90€$. Your earnings are 6.10€.

Review questions

Please answer the following questions. Raise your hand as soon as you have finished answering the questions. An experimenter will come to you and check your answers.

1. Indicate whether the statements are true or false.

Statement	**True**	**False**
In each round, a state's rank in terms of departing passengers in 2011 must be predicted.		
Only the forecast (your own or that of the algorithm resp. of the other person) of the 11th round and a randomly drawn choice problem are relevant for payoff.		
The more the prediction (your own or that of the algorithm resp. of the other person) deviates from the true rank, the higher your payoff.		

2. Determine a person's payout for the following example situations.

Example situation	**Payoff**
For example, suppose the computer has selected situation A and the number 1. The person has decided to use the algorithm's prediction in row 1. The absolute value of *forecast−true rank* is 10.	€
For example, suppose the computer has selected situation A and the number 5. The person has decided in row 5 that they will make the forecast themselves. The absolute value of *forecast−true rank* is 10.	€
For example, suppose the computer has selected situation B and the number 5. The person decided to use the other person's prediction in row 5. The absolute value of *forecast−true rank* is 10.	€