Measuring Preferences for Algorithms — How willing are people to cede control to algorithms? [†]

July 29, 2024

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

[a] *Technische Universität Berlin, Germany*

Abstract

We suggest a simple method to elicit individual preferences for algorithms. By altering the monetary incentives for ceding control to the algorithm, the menu-based approach allows for measuring, in particular, the degree of algorithm aversion. Using an experiment, we elicit preferences for algorithms in an environment with measurable performance accuracy under two conditions—the absence and the presence of information about the algorithm's performance. Providing such information raises subjects' willing-ness to rely on algorithms when ceding control to the algorithm is more costly than trusting in their own assessment. However, algorithms are still underutilized.

Keywords: Algorithm Aversion, Delegation, Experiment, Preferences

JEL classifications: C91, D83, D91

1. Introduction

Algorithms are omnipresent and have become a fundamental part of our daily life. When solving problems, humans are increasingly faced with the choice of whether to delegate decisions to algorithms in business, medicine, public policy, as well as in everyday situations. Two phenomena dominate the ongoing discussion on the interaction between humans and algorithms: algorithm aversion and algorithm appreciation. Algorithm aversion describes the human reluctance to rely on decisions or recommendations made by an algorithm in favor of relying on the own judgment (Dietvorst et al., 2015). Algorithm appreciation is defined as the opposite phenomenon: a human decision-maker prefers algorithmic judgment to human judgment (Logg et al., 2019).[1]

The literature has identified a number of factors that influence the presence of algorithm aversion, e.g., performance feedback (Dietvorst et al., 2015; Alexander et al., 2018; Jung and Seiter, 2021), the characteristics of the task (Dietvorst and Bharti, 2020), response time of the algorithm (Efendić et al., 2020),

[*] E-mail: `ivanova-stenzel@tu-berlin.de`

[**] E-mail: `michel.tolksdorf@tu-berlin.de`

[†] We thank Max Padubrin, Felix Bönisch, and participants of the Berlin Behavioral Economics reading group, the Meeting of the Verein für Socialpolitik 2023, the Meeting of the CRC "Rationality and Competition," and the 2023 European Winter Meeting of the Econometric Society. 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 was pre-registered at AsPredicted.org (#112129).

[1] For a comprehensive interdisciplinary review on how humans interact with automated agents, see Chugunova and Sele (2022).

time pressure (Jung and Seiter, 2021), learning (Filiz et al., 2021), the illusion of better understanding human than algorithm decision-making (Bonezzi et al., 2022), or consequences of failure (Filiz et al., 2023). Several factors influencing algorithm appreciation have also been identified, such as the nature of the task (Logg et al., 2019), performance feedback (You et al., 2022), and incentives (Greiner et al., 2022), to name a few. An increasing number of papers focus on finding ways to overcome algorithm aversion and increase algorithm appreciation. This includes permitting subjects to modify the decisions produced by an algorithm or to select which information the algorithm processes (Dietvorst et al., 2018; Jung and Seiter, 2021; Sele and Chugunova, 2024), as well as demonstrating the algorithm's ability to learn and improve (Berger et al., 2021; Reich et al., 2023).[2]

Despite the large body of literature on algorithm aversion and algorithm appreciation, far too little is known about the strength of the preferences for algorithms, in particular the degree of algorithm aversion. [3] The methodology used so far involves the choice between one's own decision and the algorithm's decision when facing a problem on what might be called a ceteris paribus basis. That is, with all things being equal, which judgment would subjects prefer? This approach neither captures the heterogeneity of preferences for algorithms nor takes the costs of employing algorithms into account. While it allows for measuring the effects of interventions aiming to overcome algorithm aversion at the extensive margin, it misses the effects at the intensive margin. In this paper, we take the investigation a step further and suggest a simple method that allows for the measurement of the intensity of the preferences for algorithms.

Consider a payoff-relevant task that can be performed by an algorithm. The method allows for finding an answer to the question: "How many monetary benefits are subjects willing to give up to avoid the algorithm?" It involves presenting subjects with a menu of simple choice problems. In each problem, subjects can choose between performing the task on their own or delegating it to the algorithm. The problems differ with respect to the performance-based payoff when the algorithm is chosen. By varying the payoff for employing the algorithm, the menu-based approach allows for measuring the degree of algorithm aversion.

We apply the suggested fine-grained method and elicit subjects' algorithm attitudes under two conditions—the absence and the presence of information about the algorithm's performance. Our experimental results show the existence of potentially strong preferences for algorithms. In both conditions, subjects exhibit lower algorithm aversion compared to the incentive-based prediction, i.e., when taking into account only the size of the payment reduction for inaccurate forecasts. Still, subjects underutilize algorithms compared to a performance-based prediction, i.e., when taking into account feedback on the subject's own performance and the algorithm's performance. In the presence of feedback on the algorithm's performance, we find a correlation of predicted and observed algorithm choices. This correlation is largely driven by beliefs about the algorithm's and the own performance. Providing information about the algorithm's

[2] Jussupow et al. (2020) provide an overview of empirical evidence on algorithm aversion and appreciation. Mahmud et al. (2022) review the factors that affect algorithm aversion. Kaufmann et al. (2023) review the task-dependence of algorithm advice acceptance.

[3] One approach to measure preferences for algorithms is by eliciting the Weight on Advice (WOA) via the Judge Advisor System (JAS) used, e.g., in Logg et al. (2019). In the JAS, participants may revise an initial judgment after receiving advice. Participants' algorithm use is then measured as the WOA. The WOA ranges from 0% to 100% when a participant revises their judgment toward the advice. WOA is 0% when the advice is ignored or the revised judgment is in the opposite direction of the advice. Thus, the JAS allows for a conclusive measurement only of the degree of algorithm appreciation.

performance positively affects the use of algorithms.

When applying the binary measure commonly used in previous studies to our data, we are unable to detect either the positive effect on the algorithm's performance or the role of beliefs. This demonstrates that the suggested menu-based approach can uncover differences in algorithm preferences that methods which rely on ceding control to algorithms, when the payoff consequences are identical, might miss.

2. Experimental Design

We apply the suggested method to elicit preferences for algorithms in a forecasting task used in several experimental studies, which gives subjects a choice between making a forecast decision themselves or delegating it to an algorithm (e.g., Dietvorst et al., 2015; Logg et al., 2019; Jung and Seiter, 2021). Those studies typically report that participants prefer the algorithm's decision over their own in 40 to 60% of cases, providing sufficient margins for both algorithm appreciation and aversion. Moreover, performance in forecasting tasks can be precisely measured after observing the realized outcomes. Furthermore, these tasks enable the elicitation of algorithm performance on the same scale as human performance.

As performance feedback plays a crucial role in the occurrence of algorithm aversion (see surveys by Jussupow et al., 2020; Mahmud et al., 2022, and the studies cited therein), we used a between-subjects design and implemented two conditions, which differed with respect to the feedback participants received. In the "Human"-condition, participants only got feedback on their performance. In the "Human-Algorithm"-condition, they saw the algorithm's performance in addition to their own performance. By keeping the experience with the task constant (i.e., providing feedback on participants' performance) between conditions, we can isolate and measure the effect of feedback on the algorithm's performance.

Following Dietvorst et al. (2015), Logg et al. (2019), and Jung and Seiter (2021), our experiment consisted of 11 rounds. In the first 10 unincentivized rounds, participants were asked to forecast the rank of a randomly chosen U.S. state in terms of departing air passengers in 2011 based on five pieces of information (number of major airports, rank of population, rank of county, rank of median household income, rank of expenditure on domestic travel). In both conditions, participants got feedback comprising this information, their forecast, and the true rank. In the Human-Algorithm condition, they additionally received information on what rank an algorithm predicted for this task.[4]

In the final payoff-relevant round, subjects first had to decide whether their own forecast or the forecast of the algorithm should be used to determine their payoff in that round. For each unit of deviation of the forecast from the true rank, the payoff was reduced by the penalty rate X. More precisely, the payment rule was

$$\text{payoff} = 7 - X * |\text{forecast} \ - \ \text{true rank}|$$

in EUR. Thus, the unit of deviation in a forecast is the measure for forecasting inaccuracy.

In contrast to the previous studies, where subjects made a single choice between using their own forecast or the algorithm's forecast to determine their payoff, participants in our experiment made decisions in nine choice problems, see Table 1. For the case in which the subject's own forecast becomes payoff-relevant, the penalty rate X_{own} remains constant across all nine problems. For the case that the algorithm's forecast

[4] We used the original algorithm employed in Experiment 3A/B by Dietvorst et al. (2015).

is used to determine the subject's payoff, the penalty rate X_{alg} increases when going down the list of choice problems. In the first choice problem, the algorithm forecast is payoff-irrelevant ($X_{\text{alg}} = 0$). In the fifth choice problem, the penalty rate is the same regardless of the type of forecast, human or algorithm ($X_{\text{own}} = X_{\text{alg}} = 0.12$). Hence, this choice problem corresponds to the single choice subjects usually faced in previous studies (e.g., Dietvorst et al., 2015; Logg et al., 2019; Jung and Seiter, 2021).

Table 1: The payoff structure of the nine paired choice problems.

Choice problem	Payoff (own forecast) in EUR $7 - X_{\text{own}} * \lvert\text{forecast} - \text{true rank}\rvert$	Payoff (algorithm forecast) in EUR $7 - X_{\text{alg}} * \lvert\text{forecast} - \text{true rank}\rvert$	Penalty rate difference $(X_{\text{alg}} - X_{\text{own}})$
1	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0 * \lvert$ forecast $-$ true rank$\rvert$	-0.12
2	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.03 * \lvert$ forecast $-$ true rank$\rvert$	-0.09
3	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.06 * \lvert$ forecast $-$ true rank$\rvert$	-0.06
4	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.09 * \lvert$ forecast $-$ true rank$\rvert$	-0.03
5	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	0
6	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.15 * \lvert$ forecast $-$ true rank$\rvert$	0.03
7	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.18 * \lvert$ forecast $-$ true rank$\rvert$	0.06
8	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.21 * \lvert$ forecast $-$ true rank$\rvert$	0.09
9	$7 - 0.12 * \lvert$ forecast $-$ true rank$\rvert$	$7 - 0.24 * \lvert$ forecast $-$ true rank$\rvert$	0.12

A subject who is only concerned with the size of the payment reduction would stop choosing the algorithm when $X_{\text{alg}} > X_{\text{own}}$ (incentive-based prediction).

Taking performance feedback into account, it would be optimal to stop choosing the algorithm for

$$X_{\text{alg}} > X_{\text{own}} * \frac{\lvert\text{own forecast}-\text{true rank}\rvert}{\lvert\text{algorithm's forecast}-\text{true rank}\rvert},$$

where |own forecast − true rank| and |algorithm's forecast − true rank| are the individual averages of the performances in the 10 unincentivized rounds (performance-based prediction).

After subjects decided on all nine choice problems, they had to make a forecast once again. At the end of the round, subjects were asked questions that dealt with their confidence regarding their own and the algorithm's forecasting performance. Furthermore, they were asked about their beliefs regarding the expected error in their own forecast and in the algorithm's forecast (see Table A1 in the Appendix for the set of questions).

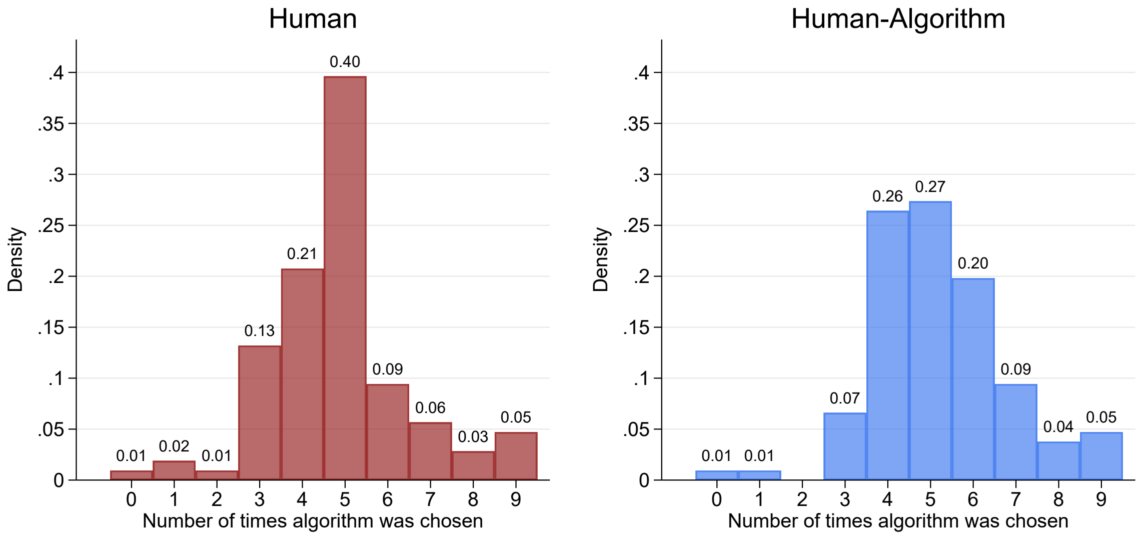
At the end of the experiment, one of the nine choice problems was randomly selected and implemented for real payment. Participants had to fill in a post-experimental questionnaire that contained, among other things, a self-assessed, hypothetical measure of risk attitude on a 0–10 scale (SOEP). We also elicit subjects' attitudes toward ambiguity. Altogether, the experiment encompassed a total of 212 subjects (106 per condition). The experiment was conducted with students at the Technische Universität Berlin (47%/52% female in the Human/Human-Algorithm condition). Participants were invited to the experiment with ORSEE (Greiner, 2015). All experimental sessions were conducted using z-Tree (Fischbacher, 2007). The average length of a session was 45 minutes. The average total earnings per participant was 13.68 EUR, including a show-up fee of 6 EUR.[5]

[5] See Section 4.3 in the Appendix for a translated version of the instructions.

3. Results

Figure 1 illustrates the heterogeneity of participants' preferences for algorithms by showing the number of times participants chose the algorithm for each condition. The number of algorithm choices is an indicator of the degree of algorithm aversion. In Human, the number of times the algorithm was chosen peaks at five, while it peaks at both four and five in Human-Algorithm. In Human, 17% of the participants choose the algorithm zero to three times, opposed to 9% in Human-Algorithm. In contrast, in Human-Algorithm 38% of the participants choose the algorithm six to nine times, opposed to 23% in Human. The majority of participants (94% in each condition) are neither completely reliant on the algorithm nor do they fully cede control to the algorithm.

Figure 1: Relative share of the number of times the algorithm was chosen per condition.



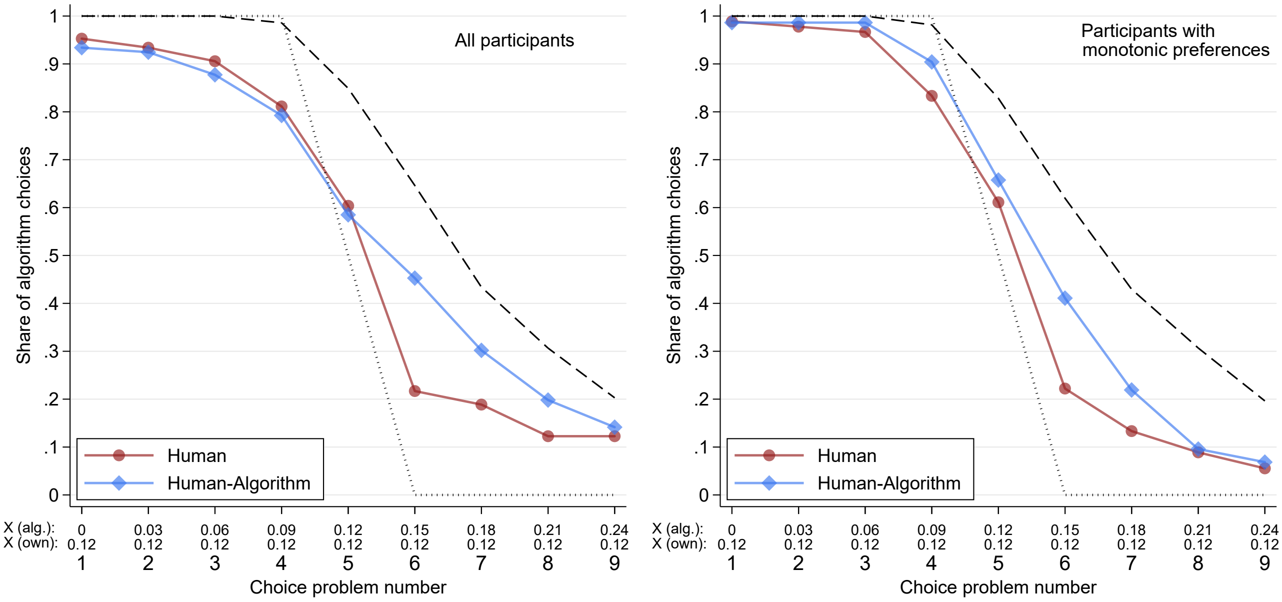
Next, we check whether the participants' choices are consistent and purposeful. In the first choice problem, where the penalty rate is zero for the algorithm's forecast, 95%, respectively 93%, of the participants choose the algorithm in Human, respectively Human-Algorithm. We consider participants as having monotonic preferences when they switch at most once from choosing the algorithm to choosing oneself. 85% of the participants in Human and 69% of the participants in Human-Algorithm exhibit monotonic preferences. Note that for monotonic preferences the number of algorithm choices designates a unique crossover point from choosing the algorithm to choosing oneself.[6]

Figure 2 presents the proportion of algorithm choices for each of the nine choice problems per condition for all participants (left graph) and those who exhibit monotonic preferences (right graph). The figure in-

[6] See Figure A1 in the Appendix for the distributions of the number of switches. The mode and median number of switches are one in both conditions. 85% of the participants in Human and 72% of the participants in Human-Algorithm switch at most once. A higher number of switches might indicate uncertainty about the point of indifference. Instead of using the number of algorithm choices, an alternative way to deal with such non-monotonic preferences is to consider intervals of penalty rates. The lower (upper) bound of an interval is determined by the first (last) switching point. Results do not change when using those intervals as the dependent variable in the regression analysis.

cludes reference lines representing the incentive-based prediction (dotted line) and the performance-based prediction (dashed line) based on data from both conditions. For the derivation of both predictions, see section 2. In Human, participants behave close to the incentive-based prediction: the share of algorithm choices drops from 81% in problem 4 to 22% in problem 6 (83 to 22% in the case of monotonic preferences). In Human-Algorithm, the proportion of algorithm choices is close to both (aligned) predictions in problems 1 to 4, particularly in the case of monotonic preferences. In problems 6 to 9, where both predictions are not aligned, we observe a higher usage of the algorithm compared to the incentive-based prediction.

Figure 2: Share of algorithm choices by choice problem number (and corresponding X). The incentive-based prediction (dotted line) and the performance-based prediction (dashed line) are computed using data from both conditions.



Compared to the performance-based prediction participants underutilize the algorithm in both conditions. Yet, we find a positive correlation between the observed and the optimal number of times the algorithm was chosen in Human-Algorithm (Spearman's rank correlation, all participants: $r(105) = 0.4056$, $p < 0.01$, participants with monotonic preferences: $r(72) = 0.4414$, $p < 0.01$). This is not the case in Human, where participants do not receive feedback on the algorithm's performance (all participants: $r(105) = 0.1154$, $p = 0.24$, participants with monotonic preferences: $r(89) = 0.1349$, $p = 0.20$).[7]

In both conditions, participants behave very similarly up to choice problem 5 ($X_{\text{own}} = X_{\text{alg}}$). This changes substantially in choice problems 6 to 9, where $X_{\text{own}} < X_{\text{alg}}$. The proportion of algorithm choices in the Human-Algorithm condition is larger compared to the Human condition, indicating that algorithm aversion decreases in the presence of feedback on the algorithm's performance. Indeed, based on a Wilcoxon rank-sum test, we find significant differences between the number of algorithm choices in

[7] For the corresponding plots, see Figure A2 in the Appendix.

Human ($M = 4.86$, $SD = 1.64$) and in Human-Algorithm ($M = 5.21$, $SD = 1.60$), with $p = 0.09$.[8]

Table 2: OLS regressions within and between conditions.

	Within conditions				Between conditions	
	Human		Human-Algorithm		Reference category: Human	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance-based prediction	0.003	0.019	0.194**	0.064	0.091	0.050
	(0.076)	(0.082)	(0.080)	(0.111)	(0.055)	(0.063)
Human-Algorithm					0.474**	0.516**
					(0.188)	(0.209)
Confidence in						
forecast (own)	-0.476**	-0.538**	-0.266	-0.365	-0.395**	-0.467**
	(0.208)	(0.243)	(0.246)	(0.286)	(0.155)	(0.188)
forecast (alg.)	0.556***	0.673***	0.416	0.469	0.519***	0.601***
	(0.195)	(0.217)	(0.265)	(0.353)	(0.156)	(0.185)
Belief about						
deviation from true rank (own)	0.011	0.006	0.016	0.178**	0.011	0.036
	(0.034)	(0.037)	(0.019)	(0.081)	(0.015)	(0.037)
deviation from true rank (alg.)	0.004	0.020	-0.016	-0.209**	0.006	-0.006
	(0.042)	(0.053)	(0.033)	(0.104)	(0.025)	(0.048)
perfect prediction (own)	-0.394*	-0.400	-0.305*	-0.140	-0.379***	-0.396**
	(0.214)	(0.242)	(0.179)	(0.207)	(0.136)	(0.158)
perfect prediction (alg.)	0.389*	0.318	0.477***	0.399**	0.473***	0.491***
	(0.233)	(0.286)	(0.161)	(0.172)	(0.129)	(0.149)
Constant	4.522***	4.426***	2.720*	2.679*	2.935***	2.506**
	(1.255)	(1.440)	(1.439)	(1.358)	(1.013)	(1.030)
Gender	Yes	Yes	Yes	Yes	Yes	Yes
Ambiguity and risk aversion	Yes	Yes	Yes	Yes	Yes	Yes
Only monotonic	No	Yes	No	Yes	No	Yes
Observations	106	90	106	73	212	163
R^2	0.380	0.422	0.389	0.526	0.373	0.451

Standard errors in parentheses. The dependent variable is the number of algorithm choices. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 2 presents the OLS regression results on algorithm choices within conditions (specifications 1 to 4) and between conditions (specifications 5 and 6). In the even-numbered specifications, we consider only participants with monotonic preferences. In all specifications, we utilize belief and confidence measures while controlling for gender, risk, and ambiguity attitudes. As shown in specifications (1) and (2), the confidence in both the own and the algorithm's forecast explain the algorithm choices in the Human condition. An increased (decreased) confidence in the algorithm's (own) performance by one step on the 5-point Likert scale increases (decreases) the number of algorithm choices by around 0.5. The OLS regression results in (3) confirm the positive correlation between the observed and predicted number of algorithm choices in the Human-Algorithm condition. In the case of monotonic preferences, as apparent in (4), this correlation seems to be driven by beliefs about the own and the algorithm's accuracy of forecasts. The larger the belief about the own (algorithm's) forecasting error, the larger (lower) the number of algorithm choices. In (5) and (6), the effect of the presence of information on the algorithm's performance is captured by the "Human-Algorithm" dummy variable. Providing information about the

[8] $p = 0.07$, Human (M=4.88, SD=1.67), Human-Algorithm (M=5.32, SD=1.70) in the case of monotonic preferences.

algorithm's performance increases the number of algorithm choices by 0.5 on average.[9]

To assess the degree of algorithm aversion in terms of the methodology used so far in previous studies, we reduce our data to a binary form in two distinct ways. First, we classify participants with less than five algorithm choices as not choosing the algorithm, and participants with at least five algorithm choices as choosing the algorithm. The shares of participants who chose the algorithm obtained in this way are 62% in Human and 65% in Human-Algorithm. The difference between the two conditions is not statistically significant (two-sided Fisher's exact test, $p = 0.78$).[10] Second, we limit our data only to the single choice, where the penalty rate of the own and the algorithm's forecasting errors is equal (choice problem 5). This corresponds to the choices subjects usually faced in previous studies (e.g., Dietvorst et al., 2015; Logg et al., 2019; Jung and Seiter, 2021). The shares of algorithm choices obtained in this way are 60% in Human and 58% in Human-Algorithm.[11] Again, the difference is not statistically significant (two-sided Fisher's exact test, $p = 0.89$). Using the binary measures as dependent variables in our regressions, we cannot confirm the results obtained with our fine-grained measure (see Table A4 in the Appendix). Providing feedback on the algorithm's performance does not affect the tendency to choose the algorithm. Moreover, we are not able to detect the role of confidence in the own forecast in the Human condition and the role of beliefs in the Human-Algorithm condition.

This exercise shows that the suggested menu-based approach is able to reveal differences in algorithm preferences, where the commonly used binary measures might fail.

4. Conclusion

We suggest a simple method for the fine-grained elicitation of human preferences for algorithms. We apply it to measure how willing people are to cede control to algorithms under two conditions, the absence and the presence of information on the algorithm's performance.

We find that providing information about the performance of the algorithm reduces human reluctance to rely on algorithms. This effect would have been missed when applying the commonly used binary method. Our results reveal that the latter may be insufficient given the heterogeneity of human preferences and the complexity of the underlying decision problem.

The proposed method allows for inspecting the consistency of choices, distinguishing between monotonic and non-monotonic preferences, and capturing the variety of the degree of algorithm aversion among individuals. This heterogeneity could be an explanatory factor for economically consequential decisions on the use of algorithms in several domains, e.g., delegation of hiring decisions to algorithms (Dargnies et al., 2024), relying on algorithm advice in price estimation tasks (Greiner et al., 2022), seeking advice of algorithms in moral dilemmas (Leib et al., 2024), or delegation to algorithms in financial decision-making (Germann and Merkle, 2023).

Measuring the strength of people's preferences for algorithms is the first step toward creating a standardized assessment of algorithm aversion. A standardized measurement can confirm existing research

[9] Results remain qualitatively unchanged when *i)* employing interval regression (Andersen et al., 2006), see Table A3 in the Appendix, and *ii)* controlling for learning, i.e., by only using the second half of the 10 unincentivized rounds to determine the performance-based prediction.

[10] See Figure 1 for the number of times the algorithm was chosen per condition.

[11] Note that previous studies report that subjects choose the algorithm in 40 to 60% of the cases.

findings and facilitate the development and validation of simple survey items to evaluate algorithm aversion. Furthermore, the suggested method can be employed to distinguish preferences for algorithms from preferences for external advice in general. For example, one could measure the willingness to cede control to another human and compare it to the willingness to cede control to an algorithm.[12] We plan to investigate this matter in future work by using data from the present study as the performance of another human in the task.

References

Alexander, V., Blinder, C., Zak, P. J., 2018. Why trust an algorithm? Performance, cognition, and neurophysiology. Computers in Human Behavior 89, 279–288.

Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E., 2006. Elicitation using multiple price list formats. Experimental Economics 9, 383–405.

Berger, B., Adam, M., Rühr, A., Benlian, A., 2021. Watch me improve – algorithm aversion and demonstrating the ability to learn. Business & Information Systems Engineering 63 (1), 55–68.

Bonezzi, A., Ostinelli, M., Melzner, J., 2022. The human black-box: The illusion of understanding human better than algorithmic decision-making. Journal of Experimental Psychology: General 151 (9), 2250–2258.

Chugunova, M., 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., Kübler, D., 2024. Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence. Management Science 0 (0).

Dietvorst, B. J., Bharti, S., 2020. People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. Psychological Science 31 (10), 1302–1314.

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

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

Efendić, E., Van de Calseyde, P. P., Evans, A. M., 2020. Slow response times undermine trust in algorithmic (but not human) predictions. Organizational Behavior and Human Decision Processes 157, 103–114.

Filiz, I., Judek, J. R., Lorenz, M., Spiwoks, M., 2021. Reducing algorithm aversion through experience. Journal of Behavioral and Experimental Finance 31.

[12] In a similar vein, Dietvorst et al. (2015) and Logg et al. (2019) identified the role of authority when comparing human to algorithm choices. They report that algorithm uptake is lower in the choice between oneself and an algorithm compared to the choice between another human and an algorithm.

Filiz, I., Judek, J. R., Lorenz, M., Spiwoks, M., 2023. The extent of algorithm aversion in decision-making situations with varying gravity. Plos one 18 (2), e0278751.

Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2), 171–178.

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

Greiner, B., 2015. Subject pool recruitment procedures: Organizing experiments with ORSEE. Journal of the Economic Science Association 1 (1), 114–125.

Greiner, B., Grünwald, P., Lindner, T., Lintner, G., Wiernsperger, M., 2022. Incentives, framing, and trust in algorithmic advice: An experimental study. Working paper.

Jung, M., 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., Heinzl, A., 2020. Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. ECIS Proceedings 168.

Kaufmann, E., Chacon, A., Kausel, E. E., Herrera, N., Reyes, T., 2023. Task-specific algorithm advice acceptance: A review and directions for future research. Data and Information Management, 100040.

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

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

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

Reich, T., Kaju, A., Maglio, S. J., 2023. How to overcome algorithm aversion: Learning from mistakes. Journal of Consumer Psychology 33 (2), 285–302.

Sele, D., Chugunova, M., 2024. Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision-making. Plos one 19 (2), e0298037.

You, S., Yang, C. L., Li, X., 2022. Algorithmic versus human advice: Does presenting prediction performance matter for algorithm appreciation? Journal of Management Information Systems 39 (2), 336–365.

Appendix

4.1. Additional figures

Figure A1: Relative share of number of switches between oneself and the algorithm in the choice problems by condition.

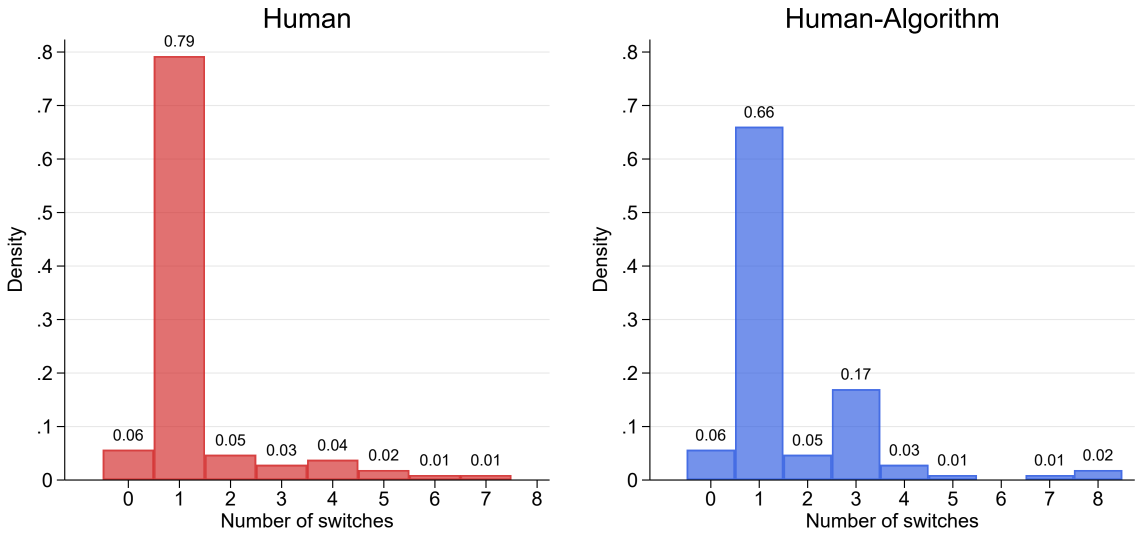
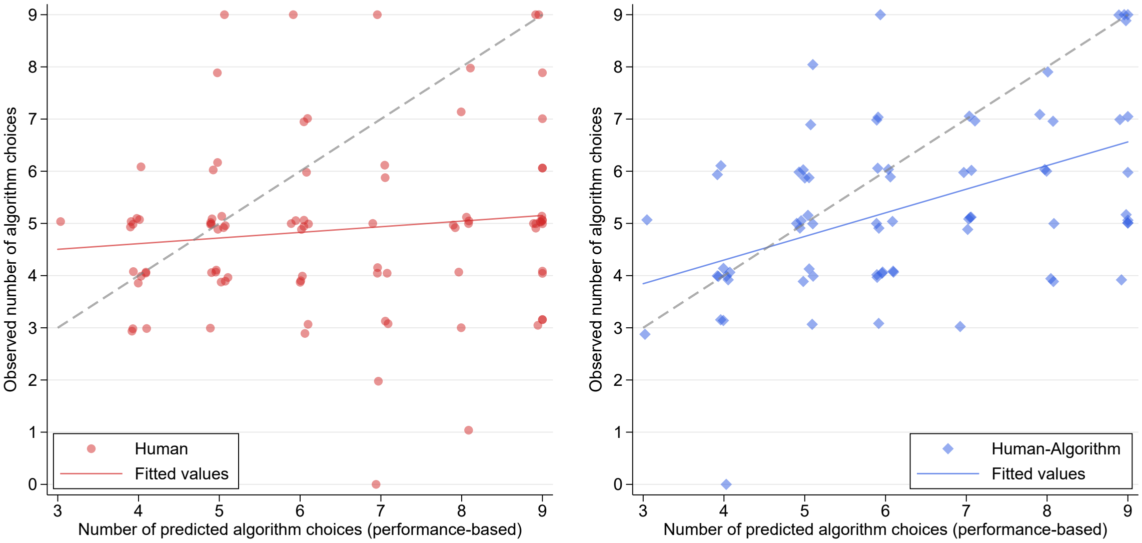


Figure A2: Observed and performance-based predicted number of algorithm choices per condition with fitted values (solid) and 45° line (dashed).



4.2. Additional tables

Table A1: Belief and confidence questions

How many ranks do you think the algorithm's estimate is away from the state's true rank? (0–50)
How many ranks do you think your 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 algorithm'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 algorithm predicted the state's rank almost perfectly? (1= *extremely unlikely*; 8= *extremely likely*)

Table A2: Mean forecasting performance of participants and algorithm by condition with std. dev. in brackets.

	|own forecast − true rank|	|algorithm forecast − true rank|	Paired t-test
Unincentivized forecasts			
Human	6.61 (2.63)	4.3 (1.12)	$t(953) = 25.40,\ p < 0.01$
Human-Algorithm	6.88 (3.32)	4.36 (1.18)	$t(953) = 23.42,\ p < 0.01$
Incentivized forecast			
Human	6.26 (5.11)	3.84 (3.77)	$t(105) = 4.51,\ p < 0.01$
Human-Algorithm	7.08 (5.48)	3.74 (3.62)	$t(105) = 7.03,\ p < 0.01$

Table A3: Interval regressions within and between conditions.

	Within conditions				Between conditions	
					Reference category:	
	Human		Human-Algorithm		Human	
	(1)	(2)	(3)	(4)	(5)	(6)
Performance-based prediction	-0.044	-0.003	0.353	0.070	0.132	0.065
	(0.197)	(0.203)	(0.222)	(0.276)	(0.153)	(0.161)
Human-Algorithm					1.434***	1.410**
					(0.526)	(0.568)
Confidence in						
Confidence in forecast (own)	-1.132**	-1.124*	-0.518	-0.555	-0.875**	-0.923*
	(0.541)	(0.580)	(0.651)	(0.776)	(0.428)	(0.477)
Confidence in forecast (algorithm)	1.787***	1.974***	1.118	0.863	1.522***	1.604***
	(0.521)	(0.557)	(0.838)	(1.001)	(0.458)	(0.500)
Belief of						
Belief of deviation (own)	0.066	0.056	0.167	0.596**	0.086	0.129
	(0.091)	(0.092)	(0.159)	(0.248)	(0.084)	(0.096)
Belief of deviation (algorithm)	-0.036	-0.005	-0.119	-0.682**	-0.014	-0.055
	(0.130)	(0.133)	(0.195)	(0.298)	(0.111)	(0.123)
Perfect prediction (own)	-0.670	-0.708	-0.782	-0.411	-0.872**	-0.868**
	(0.464)	(0.514)	(0.523)	(0.512)	(0.356)	(0.390)
Perfect prediction (algorithm)	0.425	0.312	1.366***	1.075**	1.106***	1.063***
	(0.574)	(0.660)	(0.425)	(0.450)	(0.341)	(0.388)
Constant	13.276***	13.050***	6.338*	7.868**	7.234**	7.340**
	(3.494)	(3.542)	(3.756)	(3.617)	(2.860)	(2.864)
lnsigma	1.176***	1.189***	1.153***	1.144***	1.187***	1.201***
	(0.090)	(0.091)	(0.085)	(0.089)	(0.059)	(0.060)
Gender	Yes	Yes	Yes	Yes	Yes	Yes
Ambiguity and risk aversion	Yes	Yes	Yes	Yes	Yes	Yes
Only monotonic	No	Yes	No	Yes	No	Yes
Observations	100	85	98	68	198	153

Standard errors in parentheses. Estimation by interval regression. The dependent variable is the interval of penalty rate (X). The lower (upper) bound of the interval is determined by the first (last) switching point. For example, for a participant with three switches who switches from algorithm to own forecast in choice problem 3, then back to algorithm in choice problem 4, and then back to own forecast in choice problem 5, the interval of X is $[0.04, 0.12]$. Effect sizes are expressed in euro cents. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Table A4: OLS regressions within and between conditions (binary classification and fifth choice).

	Dependent variable: Binary classification			Dependent variable: Choice in problem 5		
	Human	Human-Algorithm	Reference category: Human	Human	Human-Algorithm	Reference category: Human
	(1)	(2)	(3)	(4)	(5)	(6)
Performance-based prediction	0.106	0.432***	0.255***	0.022	0.366***	0.174**
	(0.127)	(0.117)	(0.090)	(0.121)	(0.121)	(0.086)
Human-Algorithm			0.048			0.005
			(0.062)			(0.065)
Confidence in						
Confidence in forecast (own)	-0.090	-0.165*	-0.095*	-0.042	-0.066	-0.043
	(0.064)	(0.090)	(0.052)	(0.064)	(0.101)	(0.052)
Confidence in forecast (algorithm)	0.129*	0.172**	0.132**	0.152**	0.139*	0.136***
	(0.074)	(0.083)	(0.054)	(0.065)	(0.080)	(0.049)
Belief of						
Belief of deviation (own)	0.009	0.003	0.006	0.014	0.007	0.010*
	(0.013)	(0.006)	(0.005)	(0.014)	(0.007)	(0.006)
Belief of deviation (algorithm)	-0.014	0.018	0.001	-0.003	-0.007	-0.002
	(0.017)	(0.018)	(0.011)	(0.017)	(0.020)	(0.010)
Perfect prediction (own)	-0.055	0.010	-0.056	-0.097	-0.028	-0.078*
	(0.061)	(0.066)	(0.043)	(0.061)	(0.071)	(0.043)
Perfect prediction (algorithm)	0.042	0.045	0.076	0.089	0.052	0.085*
	(0.083)	(0.057)	(0.047)	(0.085)	(0.056)	(0.047)
Constant	0.774*	0.025	0.314	0.832*	0.175	0.472
	(0.400)	(0.325)	(0.261)	(0.443)	(0.355)	(0.291)
Gender	Yes	Yes	Yes	Yes	Yes	Yes
Ambiguity and risk aversion	Yes	Yes	Yes	Yes	Yes	Yes
Within/Between conditions	Within	Within	Between	Within	Within	Between
Observations	106	106	212	106	106	212
R^2	0.226	0.291	0.224	0.236	0.216	0.199

Standard errors in parentheses. Estimation by OLS regression. The dependent variable is the binary classification of subjects' choices of the algorithm (0 corresponds to choosing the algorithm at most 4 times, 1 corresponds to choosing the algorithm at least 5 times) in specifications (1) to (3) and the subject's choice of the algorithm in the fifth choice problem in specification (4) to (6). *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

4.3. 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.

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.

The experiment consists of two parts. At the beginning of each part, you will receive detailed instructions. The parts are independent, i.e., your decisions in one part will not affect the results in the other part. At the beginning of each part, you will receive detailed instructions. On the following pages you will find the instructions for part 1. The instructions for Part 2 will appear on your screen after Part 1. In every part of the experiment you will earn money. How exactly you can earn money will be described in the instructions.

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.

Part 1

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.

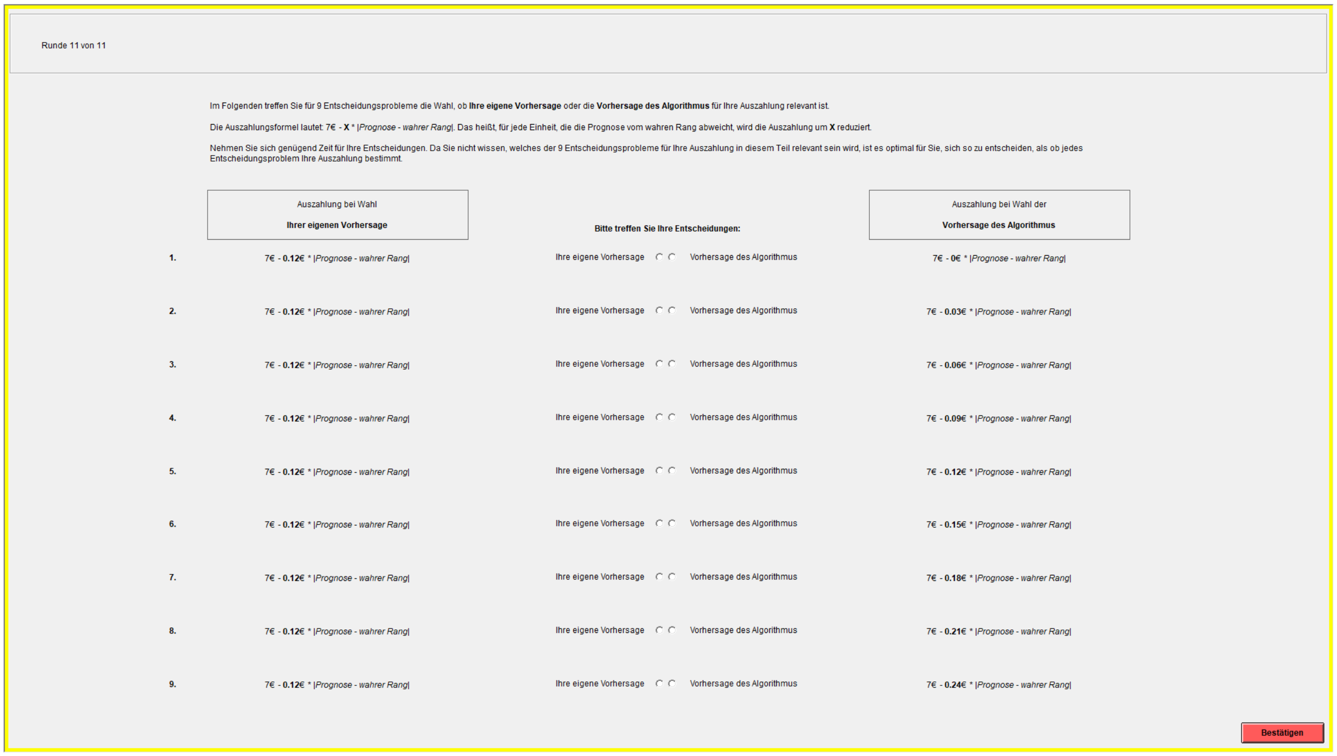
Rank in domestic travel expenditure in 2009:
States are sorted by spending on domestic travel. Rank 1 means that in this state spending on domestic travel was the highest. Rank 50 means that in this state spending on domestic travel was the lowest.

Part 1 consists of 11 rounds.

In the first 10, non-payoff-relevant, 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.

[*Human-Algorithm condition*:
In addition, you will receive information about what rank an algorithm predicted for this task. The algorithm was developed for this forecasting task. The algorithm uses only the five pieces of information mentioned above for the forecast, which is also available to you.]

In the **11th**, payoff-relevant round, you have to make a decision for nine choice problems. The screen with all nine choice problems looks like this:



Runde 11 von 11

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.

<table>
<tr><th></th><th>Auszahlung bei Wahl
<b>Ihrer eigenen Vorhersage</b></th><th><b>Bitte treffen Sie Ihre Entscheidungen:</b></th><th>Auszahlung bei Wahl der
<b>Vorhersage des Algorithmus</b></th></tr>
<tr><td><b>1.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>2.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.03€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>3.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.06€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>4.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.09€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>5.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>6.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.15€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>7.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.18€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>8.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.21€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
<tr><td><b>9.</b></td><td>7€ - <b>0.12€</b> * |<i>Prognose - wahrer Rang</i>|</td><td>Ihre eigene Vorhersage ○ ○ Vorhersage des Algorithmus</td><td>7€ - <b>0.24€</b> * |<i>Prognose - wahrer Rang</i>|</td></tr>
</table>

In each of these nine choice problems, you have to decide whether you want to let an algorithm make the forecast or make the forecast yourself.

[<u>*Human condition*</u>:
The algorithm was developed for this forecasting task. The algorithm uses only the five pieces of information mentioned above for this forecast, which is also available to you.]

If you decide to make the forecast yourself, your own prediction will be used for your payment. If you choose the algorithm, the algorithm's prediction will determine your payment.

The payment is calculated by the formula:

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

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 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 choose the algorithm.

Your decision is only valid once you have made a selection for all choice problems (i.e., for each row) 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 in Part 1 are determined as follows:

The computer randomly draws a number between 1 and 9. This random number determines the row and thus the payoff-relevant choice problem from the table shown above. If you chose the algorithm in this line, the algorithm's prediction will determine your payout. If you decided to make the forecast yourself, your own forecast will be used for the payout.

Take your time making decisions. Since you don't know which of the nine choice problems will be relevant to your payout in this part, it's optimal for you to decide as if each choice problem determines your payout.

You only make your decisions once. The drawing of the random number takes place at the end of Part 2. The result of the 11th round is displayed at the end of the experiment, i.e., after Part 2.

Below are two examples of how to determine your payment. The numbers chosen are fictional.

Example 1:
Suppose the computer randomly chooses 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 1:
Suppose the computer randomly chooses 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€·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€·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) of the 11th round and a randomly drawn choice problem are relevant for payouts.		
The more the prediction (your own or that of the algorithm) deviates from the true rank, the higher your payout.		

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

Example situation	**Payout**
For example, suppose the computer has selected the choice problem on line 1. The person has decided in line 1 that the algorithm will make the forecast. The absolute value of *forecast−true rank* is 10.	€
For example, suppose the computer has selected the choice problem on line 5. The person has decided in line 5 that they will make the forecast themselves. The absolute value of *forecast−true rank* is 10.	€
For example, suppose the computer has selected the choice problem on line 5. The person has decided in line 5 that the algorithm will make the forecast. The absolute value of *forecast−true rank* is 10.	€

Part 2

[*First screen*]

In this part, you must first choose a personal color. On the next screen, you will learn how your personal color is relevant to your payout.

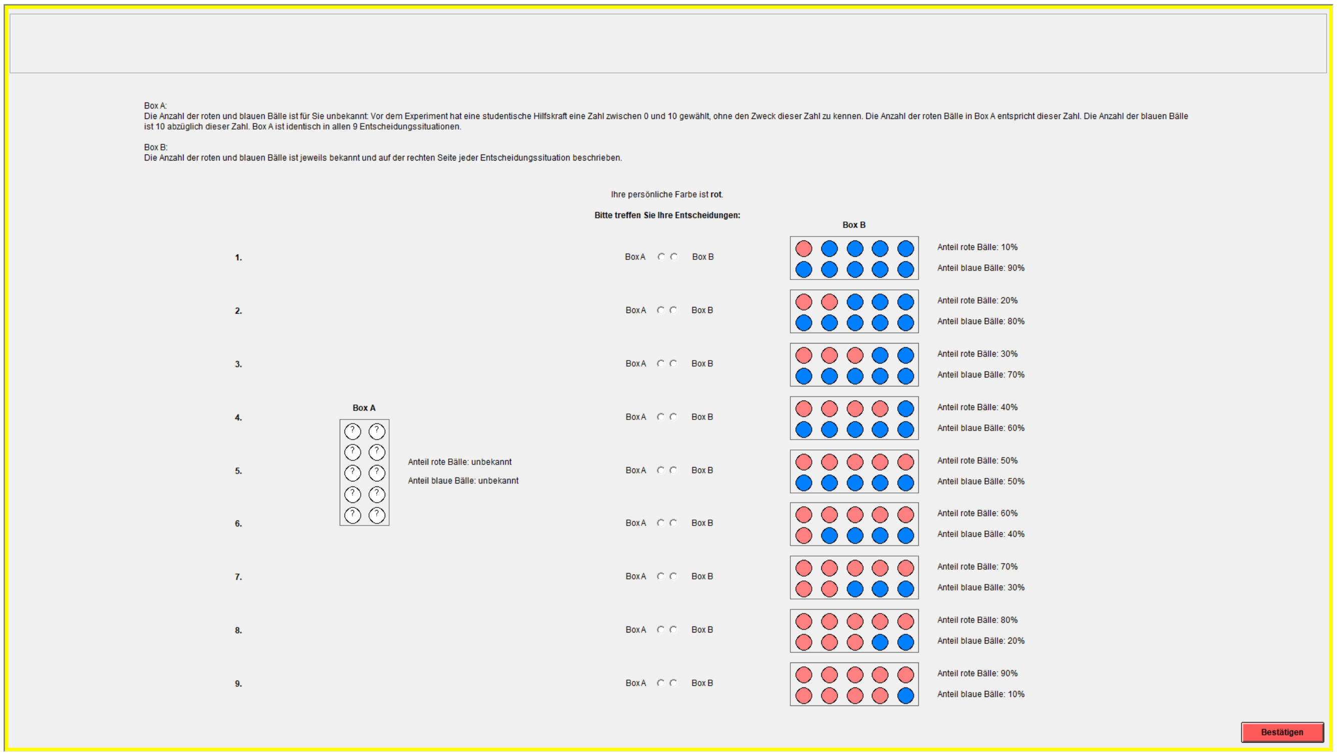
Please choose your personal color:
○ Red
○ Blue

[*Second screen*]
The next screen shows nine decision situations with two boxes each, A and B. Each box contains 10 balls. The balls can be either red or blue.

The number of red and blue balls is unknown to you: Before the experiment, a student assistant chose a number between 0 and 10 without knowing the purpose of this number. The number of red balls in box A corresponds to this number. The number of blue balls is 10 minus this number. Box A is identical in all nine decision situations.

You have to choose one box at a time. Subsequently, one of the nine decision situations is randomly selected, whereby each decision situation is equally likely. Then, a ball is randomly drawn from the box you chose in the drawn decision situation. Each of the 10 balls in the box can be drawn with equal probability. You get 2€ if the color of the randomly drawn ball matches your personal color. If the ball is the other color, you get 0€. Since you do not know which of the nine decision situations will be relevant to your payout in this part, it is optimal for you to decide as if each decision situation determines your payout.

[*Third screen*]



Box A:
Die Anzahl der roten und blauen Bälle ist für Sie unbekannt: Vor dem Experiment hat eine studentische Hilfskraft eine Zahl zwischen 0 und 10 gewählt, ohne den Zweck dieser Zahl zu kennen. Die Anzahl der roten Bälle in Box A entspricht dieser Zahl. Die Anzahl der blauen Bälle ist 10 abzüglich dieser Zahl. Box A ist identisch in allen 9 Entscheidungssituationen.

Box B:
Die Anzahl der roten und blauen Bälle ist jeweils bekannt und auf der rechten Seite jeder Entscheidungssituation beschrieben.

Ihre persönliche Farbe ist **rot**.

Bitte treffen Sie Ihre Entscheidungen:

Box A

Anteil rote Bälle: unbekannt

Anteil blaue Bälle: unbekannt

Box B

1.	Box A ○ ○ Box B	Anteil rote Bälle: 10%	Anteil blaue Bälle: 90%
2.	Box A ○ ○ Box B	Anteil rote Bälle: 20%	Anteil blaue Bälle: 80%
3.	Box A ○ ○ Box B	Anteil rote Bälle: 30%	Anteil blaue Bälle: 70%
4.	Box A ○ ○ Box B	Anteil rote Bälle: 40%	Anteil blaue Bälle: 60%
5.	Box A ○ ○ Box B	Anteil rote Bälle: 50%	Anteil blaue Bälle: 50%
6.	Box A ○ ○ Box B	Anteil rote Bälle: 60%	Anteil blaue Bälle: 40%
7.	Box A ○ ○ Box B	Anteil rote Bälle: 70%	Anteil blaue Bälle: 30%
8.	Box A ○ ○ Box B	Anteil rote Bälle: 80%	Anteil blaue Bälle: 20%
9.	Box A ○ ○ Box B	Anteil rote Bälle: 90%	Anteil blaue Bälle: 10%