

Ivanova-Stenzel, Radosveta; Tolksdorf, Michel

Conference Paper

Measuring Preferences for Algorithms - Are people really algorithm averse after seeing the algorithm perform?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage"

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Ivanova-Stenzel, Radosveta; Tolksdorf, Michel (2023) : Measuring Preferences for Algorithms - Are people really algorithm averse after seeing the algorithm perform?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage", ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/277692>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Measuring Preferences for Algorithms

– Are people really algorithm averse after seeing the algorithm perform? – [†]

March 1, 2023

Radosveta Ivanova-Stenzel^{a,*}, Michel Tolksdorf^a

^a*Technische Universität Berlin, Germany*

Abstract

We develop a simple method for elicitation of individual preferences for algorithms. By altering the payoffs for employing the algorithm, the suggested menu-based approach allows measuring the degree of algorithm aversion and appreciation, respectively. As humans seem to shy away from algorithms when seeing them err, we apply the new method to elicit preferences for algorithms in a performance-based task under the presence and absence of information about the algorithm’s performance. We observe both appreciation of and aversion to algorithms dependent on the remuneration of the use of the algorithm’s advice. The presence of information about the algorithm’s performance positively affects the use of algorithms when trusting in the algorithm’s advice is more costly than trusting in the own assessment. Our results suggest that the commonly used measures of algorithm aversion based on single individual choices may insufficiently scrutinize heterogeneous human preferences for algorithms.

Keywords: Algorithm Aversion, Algorithm Appreciation, Artificial Intelligence, Experiment, Preferences

JEL classifications: C72, C91, D44, D83

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

[†]We thank Max Padubrin, Felix Bönisch and participants of the berlin behavioral economics reading group. 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. Introduction

Algorithms are omnipresent and have become a fundamental part of our daily life. When solving a problem, humans increasingly use algorithms in business, medicine, public policy as well as in everyday decisions. The research on the interaction between humans and algorithms has revealed two phenomena: algorithm appreciation and algorithm aversion. Algorithm aversion describes the human reluctance to interact with algorithms in favor of relying on their own judgment or on decisions made by human agents.¹ Algorithm appreciation is defined as the opposite phenomenon. It describes a situation where a human decision-maker prefers to take advice from an algorithm instead of a human. A large body of literature demonstrates that people dislike algorithm advice and that they prefer human advice even when algorithms outperform human decision-making (e.g., Dietvorst et al., 2015, 2018; Dietvorst and Bharti, 2020). However, empirical findings on algorithm aversion are inconclusive as several papers find the opposite effect of algorithm appreciation (e.g., Logg et al., 2019; Mahmud et al., 2022; Chugunova and Sele, 2020).

Despite the impressive number of studies documenting both phenomena, far too little is known about the strength of the preferences for algorithms. The methodology used so far in experimental studies to measure human preferences for algorithms involves the choice between the subject’s own assessment or an algorithm advice when facing a decision problem on what might be called a *ceteris paribus* basis. That is, with all things being equal, which advice would subjects prefer? While this approach is suitable for proving the existence of algorithm aversion and measuring the strength of algorithm aversion on aggregate, it does not allow getting insights into the shape of the individual preferences for algorithms. It would be useful to take the investigation a step further and measure the intensity of those preferences. Our study aims at filling this gap. We suggest a simple procedure that allows us to elicit subjects’ algorithm attitudes or find an answer to the question: “How much monetary benefit are subjects willing to give up to avoid algorithms?”

In our experiment, we present subjects with a menu of simple choice problems that may be used to determine the degree of algorithm aversion. By varying the size of payment reduction due to error when the algorithm decision becomes payoff relevant in subsequent choice problems and counting the number of choice problems in which a subject selects the algorithm advice, we obtain a measure of the degree of algorithm aversion (resp. appreciation). The smaller the number, the higher (resp. lower) the algorithm aversion (resp. appreciation). The crossover point from the choice of the algorithm advice to the own assessment determines the degree of algorithm aversion.

As performance feedback plays a crucial role in the occurrence of algorithm aversion (Dietvorst et al., 2015, 2018; Jung and Seiter, 2021), we elicit and compare subjects’ algorithm attitudes for two environments that differ with respect to the level of information about the algorithm performance, i.e., under the presence and absence of information about the algorithm performance. We study the preferences for algorithms in a forecasting environment. Forecasts allow for a precise measure of performance after observing the realized outcomes. They also allow for the elicitation of the algorithm performance on the same scale as the human performance. Moreover, algorithms have become an indispensable part of many forecasting tasks that

¹Dietvorst et al. (2015, 2018) introduce the term algorithm aversion when humans avoid using statistical algorithms in forecasting decisions.

have wide applications in economic and managerial decisions. For example, firms use algorithm forecasts when evaluating employees’ performance, in supply chain management (Kimbrough, 2002) or in hiring (Johar, 2022; Polli, 2019). Consider a human who is faced with the choice of whether to rely on their own decision or on the decision made by an algorithm. How does this human decide which decision to rely on? There are a number of factors that might affect this decision. For example, seeing the algorithm err makes people less likely to rely on the suggested decisions by the algorithm (Dietvorst et al., 2015). Other factors such as performance feedback (Arkes et al., 1986; Dietvorst et al., 2015; Jung and Seiter, 2021), time pressure (Jung and Seiter, 2021), familiarity with the task (Berger et al., 2021), the characteristics of the task (Logg et al., 2019), learning (Filiz et al., 2021), or the illusion of better understanding of human than algorithm decision making (Bonezzi et al., 2022) might influence the presence and strength of algorithm aversion.² It is possible though that subject’s ambiguity attitudes can have an impact on human preferences for algorithms. An ambiguity-averse individual prefers to avoid an alternative where the probability distribution of possible outcomes is unknown. Not having any information about the performance of an algorithm or having feedback about the imperfect performance of the algorithm may increase the uncertainty about the outcome based on the algorithm performance. It seems reasonable to expect that an ambiguity-averse individual might exhibit a higher degree of algorithm aversion. This point will also be an issue of this study.

Our experimental results show the existence of potentially strong preferences for algorithms. We observe both, appreciation of and aversion to algorithms dependent on the remuneration of the use of the algorithm advice. In general, subjects exhibit higher algorithm appreciation compared to the incentive-based prediction, i.e., the crossover point from the choice of the algorithm advice to the own assessment is higher compared to the predicted crossover point if subjects were concerned only with a monetary benefit. Feedback on the algorithm performance positively affects the use of algorithms when reliance on algorithm advice is relatively more costly than the own assessment. At the same time subjects underreact compared to a performance-based prediction, i.e., the crossover point if subjects take into account the feedback on their own and the algorithm performance. We also find that ambiguity aversion does not seem to have an impact on algorithm preferences.

The rest of the paper is organized as follows. The next section presents the design of the tasks and the elicitation procedure. Section 3 offers the analysis of our experimental data. Section 4 provides concluding remarks.

2. Experimental design

Our experimental design aims at measuring human preferences for statistical algorithms and at understanding whether and to what extent ambiguity aversion shapes those preferences.

The experiment consists of two independent parts. In part I, we elicited subjects’ algorithm attitudes in a forecasting task. Part II assessed participants’ ambiguity aversion. We informed participants at the beginning of the experiment that they would receive instructions for the second part only after the completion of the first part.

²For surveys, see Mahmud et al. (2022); Burton et al. (2015); Jussupow et al. (2020); Chugunova and Sele (2020).

Elicitation of subjects' preferences for algorithms (Part I):

In Part I, the subject's task was to make a forecast. More specifically, they 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 (number of major airports, rank of population (2010), rank of county, rank of median household income (2008), rank of expenditure on domestic travel (2009)). 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.

We chose this task since several experimental studies report that participants prefer the algorithm advice over their own assessment in 40 – 60% of the cases, providing sufficient margins for both, algorithm appreciation and aversion (Dietvorst et al., 2015; Logg et al., 2019; Jung and Seiter, 2021).³

Part I consisted of eleven rounds. In the first ten rounds, subjects made their own forecast. These rounds were unincentivized. The idea was to allow subjects to understand the task well enough before making an incentivized forecast in the final, eleventh round.

We used a between-subjects design and implemented two conditions. The conditions differ with respect to the feedback on the algorithm forecasting performance. In the “Human”-condition, subjects did not gain experience with the algorithm forecasting performance. They made their own forecasts and afterward got feedback only on their own performance by seeing their forecast and the true rank. In the “Human-Algorithm”-condition, subjects gained experience with both, the algorithm forecasting performance and their own forecasting as they additionally saw the algorithm forecast. The original algorithm was developed by Dietvorst et al. (2015). Participants got the following information about the algorithm: “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.”

In the final payoff relevant round of part I in each condition, subjects first had to decide whether their own forecast or the algorithm forecast should be used to determine the payoff in that round. Following Dietvorst et al. (2015) and Jung and Seiter (2021), for each unit of deviation of the forecast from the true rank, the payoff was reduced by X EUR. More precisely, the payment rule was $\pi = 7 \text{ EUR} - X \text{ EUR} \cdot |\text{forecast} - \text{true rank}|$. Thus, the unit of deviation in each forecast is the measure for forecasting accuracy.

In Dietvorst et al. (2015) and Jung and Seiter (2021) subjects made a single choice between using their own forecast or the algorithm forecast to determine their payoff. In contrast, participants in our experiment made nine choices. Thereby, the penalty rate X increases when going down the list of choices for the case that the algorithm forecast is used to determine the subject's payoff, while X remains constant across all nine choices for the case that the subject's own forecast becomes payoff relevant (see Table 1).

In the first choice problem, $X = 0$. Obviously, a subject should always choose the algorithm in that choice problem, as the algorithm forecast accuracy is payoff irrelevant. In the fifth choice problem, the penalty rate X was the same regardless of the type of forecast, human or algorithm. 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). An earlier crossover point indicates a higher level of algorithm aversion. Moreover, the wide range of payoffs used allows us to measure both, algorithm aversion and

³The task was first used by Dietvorst et al. (2015), experiment 3A/B in their paper.

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

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

algorithm appreciation. Note, further, that the penalty rate $\tilde{X} = 0.12 \times \frac{|\text{own forecast} - \text{true rank}|}{|\text{algorithm forecast} - \text{true rank}|}$, where $|\text{own forecast} - \text{true rank}|$ and $|\text{algorithm forecast} - \text{true rank}|$ are the averages of the training round performances, at which participants should expect similar payoffs, can be used to determine the optimal number of algorithm choices.⁴

After subjects decided on all nine choice problems, once again they had to submit their own forecast based on the five pieces of information. At the end of this round, subjects were asked several questions that dealt with their confidence regarding their own and the algorithm forecasting performance. Furthermore, they were asked for their belief about the expected error of their own and the algorithm forecast (see Table A2 in the Appendix for the full set of questions on beliefs and confidence).

At the end of part I, one of the nine choice items was randomly selected and implemented for real payment. Performance feedback on the incentivized final round was provided only after the entire experiment to avoid behavioral complementarities in performance between the two parts or behavioral effects from wealth accumulation.

Elicitation of subjects' ambiguity attitudes (Part II):

Part II provided a systematic assessment of the subjects' ambiguity attitudes. In that part, participants made choices between an ambiguous prospect and nine different risky prospects. Similar to Kocher et al. (2018), we assume the exchangeability of events for each of the ambiguous prospects. That is, an ambiguity-neutral decision maker is indifferent between betting on a color drawn from a two-color ambiguous prospect, and a risky prospect with an ambiguity-neutral probability of 50%. We elicited probability equivalents for the ambiguous prospect: we looked for the risky prospect where a subject is indifferent between this prospect and the ambiguous prospect. Note that differences in the chosen risky prospects across subjects allow us to identify ambiguity attitudes more precisely. While the ambiguity prospect remains constant across all nine choices, the known probability of the risky prospect increases when going down the list of choices (see Table A3 in the Appendix). At the top of the list, the known-risk prospect is very unattractive and subsequently becomes more attractive. If the first risky prospect is less attractive than the ambiguous prospect and the ninth risky prospect is more attractive than the ambiguous prospect, there will be a probability at which the decision maker is indifferent between the

⁴We define algorithm choices as optimal when they yield strictly (not weakly) higher payoffs compared to the own assessment. This can be interpreted as a lower bound of the optimal algorithm choices.

two prospects. Following the approach of Kocher et al. (2018), we use the choice item at which the participant switched from a preference for the ambiguous prospect to a preference for the risky prospect to approximate the participant’s probability equivalent.

At the end of the experiment, 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), gender, age, field of study, and the number of semesters at university.

Altogether, the experiment encompasses a total of 212 subjects (106 in the Human condition; 106 participants in the Human-Algorithm condition). The experiment was conducted at Technische Universität Berlin. Participants were students (47%/52% female in the Human/Human-Algorithm condition) from the local participant pool, mostly from economics, engineering, or the natural sciences. 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 amounted to 13.68 EUR, including show-up fee of 6 EUR. Earnings were paid out in cash directly after the end of the experiment.

3. Results

As a first step, we check whether the choices of the participants are consistent and purposeful. In the first choice problem, where the penalty rate is zero for the algorithm forecast, 95% (93%) of the participants choose the algorithm in Human (Human-Algorithm). The majority switch at most once in 85% (72%) of the cases in Human (Human-Algorithm), see Figure A1 in the Appendix. For the rest of the participants in Human-Algorithm, three switches are most common (17%), which might indicate some uncertainty about the point of indifference. However, even for those who switch more than once, there is a clear purpose in their choices. Participants choose the algorithm strictly more often in problems 1 to 4, where the algorithm has a lower penalty rate, compared to problems 6 to 9, where the algorithm has a higher penalty rate, in 89% (86%) and weakly more often in 98% (96%) of the cases in Human (Human-Algorithm).⁵ Therefore, the number of times a participant chose the algorithm will be used as an indicator for algorithm preferences.

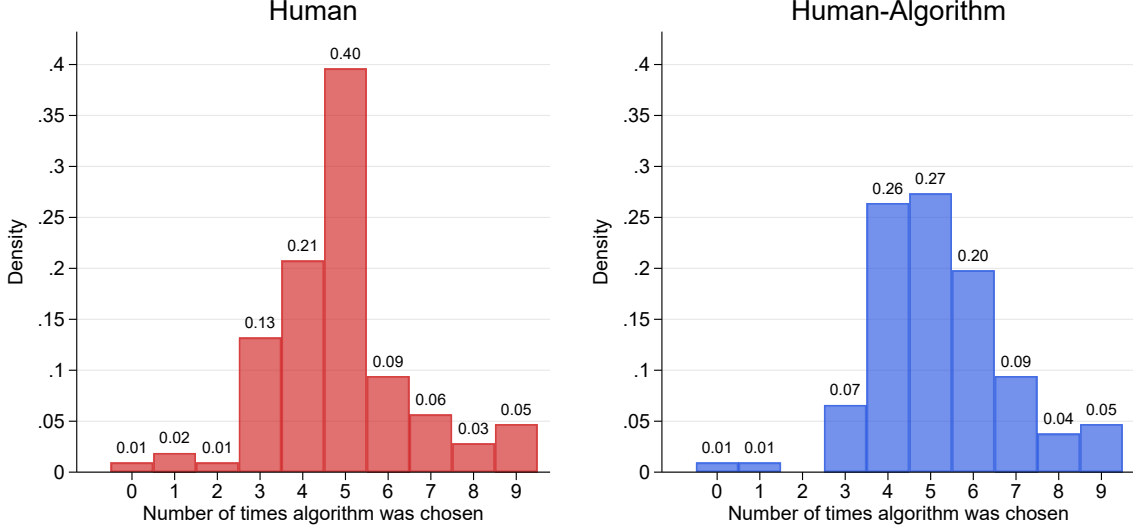
In Figure 1 we show the number of times participants chose the algorithm for each condition. In Human, the number of times the algorithm was chosen peaks at 5, while it peaks at both 4 and 5 in Human-Algorithm. Based on the choice problem structure, we can classify algorithm preferences in a strict sense with 0 to 3 algorithm choices as ‘algorithm averse’, 4 and 5 algorithm choices as ‘algorithm indifferent’ and 6 to 9 algorithm choices as ‘algorithm appreciative’.⁶ In Human, we observe more algorithm aversion with a share of 17% than in Human-Algorithm, where the share is 9%. On the contrary, we observe more algorithm appreciation in Human-Algorithm with a share of 38%, compared to a share of 23% in

⁵This leaves 2 (4) out of 106 participants who might have misunderstood the instructions, as they chose the algorithm more often at higher penalty rates in Human (Human-Algorithm). Excluding these observations, observations of participants choosing the own forecast when the algorithm has a penalty rate of 0 or observations of participants who switch more than once does not alter our results going forward.

⁶With 4 algorithm choices participants use their own forecast in the fifth choice problem and with 5 algorithm choices participants use the algorithm forecast in the fifth choice problem. Note that at the fifth choice problem the penalty rate is the same regardless of who makes the forecast.

Human. That is, nearly half of our participants express a preference for or against algorithms, which strictly exceeds indifference.

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



In line with the literature (e.g., Dietvorst et al., 2015; Jung and Seiter, 2021), we find that the algorithm significantly outperforms the human participants, both in the unincentivized forecasting tasks in the first ten training rounds and in the incentivized forecasting task in the last round. The average deviation of the participants' forecast from the true rank ranges from 6.26 to 7.08, while the average deviation of the algorithm forecasts ranges from 3.74 to 4.36 (see Table A1 in the Appendix). There are no significant differences in participants' forecasting performance between unincentivized and incentivized rounds: $t(1164) = 0.56, p = 0.58$ in Human and $t(1164) = -0.31, p = 0.76$ in Human-Algorithm (based on t -tests).⁷

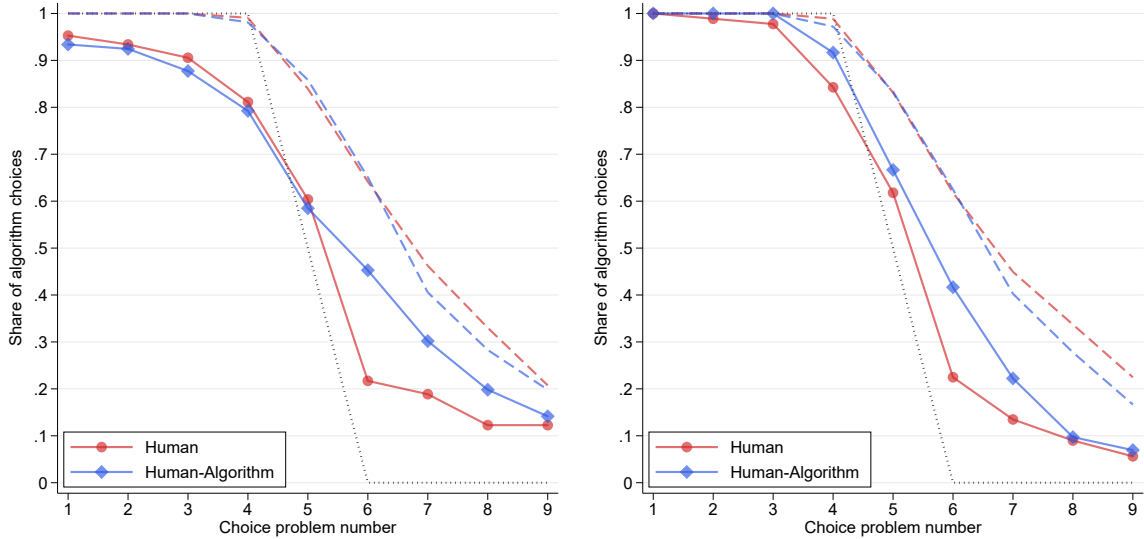
Figure 2 presents the proportion of algorithm choices per condition. The left graph in the figure shows the results for the full sample, the right graph for participants who switch at most once and choose the algorithm in the first problem (where there is no penalty if the algorithm forecast is inaccurate), i.e., for participants whose behavior can be considered rational and consistent. The graphs in the figure also include reference lines representing two different types of predictions: an incentive-based prediction (dotted line) and a performance-based prediction (dashed lines).⁸ A purely incentive-based choice would predict 4.5 algorithm choices. In both conditions, participants show a tendency towards algorithm appreciation: the mean number of algorithm choices is 4.86 in Human and 5.21 in Human-Algorithm. At the same time, algorithm choices are throughout below the performance-based predictions. When considering all participants' choices (see left graph), we observe that participants behave very similar in both conditions up until the 5th decision, where the penalty rate of the own and the algorithm forecasts are equal. The

⁷All reported tests throughout the paper are two-sided.

⁸For the derivation of the performance-based prediction, see section 2.

proportion of algorithm choices of those participants who behave rationally and consistent (see right graph) is close to the predictions in choice problems 1 to 4, where both predictions, incentive-based and performance-based, are aligned. In problems 6 to 9, the penalty rate of the algorithm is strictly larger than the penalty rate for the own forecast. In this area, we observe a higher proportion of algorithm choices in the Human-Algorithm condition compared to the Human condition. Based on a Wilcoxon rank-sum test, $p = 0.09$, we find significant differences between the number of algorithm choices in Human ($M = 4.86$, $SD = 1.64$) and in Human-Algorithm ($M = 5.21$, $SD = 1.60$).⁹

Figure 2: Share of algorithm choices by choice problem number with incentive-based prediction (dotted line) and performance-based prediction (dashed lines). The left figure includes the full sample. The right figure includes only participants who switch at most once and choose the algorithm in the first problem ($X = 0$).



We expect that participants in the Human-Algorithm condition react to the feedback about the algorithm performance and make their choices accordingly. Based on Spearman's rank correlation, we do find a positive correlation between the observed and the optimal number of times the algorithm was chosen, $r(105) = 0.4056$, $p < 0.01$. In the Human condition, participants don't receive feedback on the algorithm performance. Indeed, we do not find a correlation between the observed and the optimal number of times the algorithm was chosen, $r(105) = 0.1154$, $p = 0.24$ (for the corresponding plots, see Figure A2 in the Appendix).

In specifications (1) and (2) of Table 2, we explore which factors drive the number of algorithm choices within each condition. The OLS-regression results prove the robustness of the finding that the performance-based prediction is positive and significant in the Human-Algorithm condition, but not in the Human condition. That is, our earlier finding is robust when introducing additional explanatory variables. As documented in the literature, we do not find effects for gender and risk aversion. However, we also do not find effects for ambiguity aversion. While beliefs about the deviation of the own and the algorithm's

⁹Considering only participants who choose the algorithm in the first choice problem and switch at most once, there is also a significant difference between Human ($M = 4.93$, $SD = 1.59$) and Human-Algorithm ($M = 5.39$, $SD = 1.59$), based on a Wilcoxon rank-sum test, $p = 0.06$.

Table 2: OLS regressions within and between conditions.

	Within conditions		Between conditions
	Human (1)	Human-Algorithm (2)	Reference category: Human (3)
Human-Algorithm			0.482** (0.196)
Performance-based prediction	0.015 (0.086)	0.229** (0.087)	
<u>Demographics</u>			
Female	-0.497 (0.303)	0.008 (0.269)	-0.278 (0.206)
Diverse		-4.098 (2.548)	-3.354** (1.623)
Age	0.009 (0.030)	0.057* (0.034)	0.025 (0.021)
<u>Ambiguity and risk aversion</u>			
Ambiguity aversion	-0.048 (0.206)	0.047 (0.194)	0.006 (0.139)
Risk aversion	-0.035 (0.064)	-0.061 (0.078)	-0.043 (0.045)
<u>Belief and confidence measures</u>			
Belief of deviation (own)	0.012 (0.033)	0.025 (0.020)	0.027* (0.016)
Belief of deviation (algorithm)	-0.013 (0.046)	-0.003 (0.036)	-0.008 (0.029)
Confidence in forecast (own)	-0.413* (0.213)	-0.263 (0.239)	-0.430*** (0.159)
Confidence in forecast (algorithm)	0.569** (0.218)	0.538* (0.293)	0.513*** (0.160)
Perfect prediction (own)	-0.439** (0.219)	-0.295 (0.179)	-0.371** (0.145)
Perfect prediction (algorithm)	0.346 (0.250)	0.410** (0.175)	0.496*** (0.135)
Constant	4.014** (1.640)	1.515 (1.665)	2.905** (1.148)
Field of study	Yes	Yes	Yes
Observations	106	106	212
R^2	0.475	0.480	0.397

Standard errors in parantheses. Estimation by OLS regression. The dependent variable is the number of algorithm choices $\in \{0, 9\}$. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

forecasts have the expected signs, they are both insignificant in both conditions. The confidence in the algorithm forecast is positive and significant in both conditions, suggesting, that participants are more likely to choose the algorithm when they put higher trust in it. The belief about the algorithm making a near-perfect prediction is positive in both conditions, but only significant in the Human-Algorithm condition. Note that in contrast to the general confidence in the algorithm forecast, the belief about the algorithm making a near-perfect prediction is more specific and hard to judge without any further knowledge or feedback about the algorithm. The confidence in the own forecast and the belief about making a near-perfect prediction oneself is significant in the Human condition, but not in the Human-Algorithm condition. Overall, the results suggest that confidence in the own capabilities in the Human condition is replaced by feedback, and subsequently reliance on the algorithm performance in the Human-Algorithm

condition. In specification (3) of Table 2, we confirm these observations as there is a positive effect of the presence of information about the algorithm performance when controlling for demographics, ambiguity, and risk attitudes as well as belief and confidence measures.

To assess the degree of algorithm aversion (respectively appreciation) applying the methodology used so far in previous studies (e.g., Dietvorst et al., 2015; Jung and Seiter, 2021), we reduce our data to a binary form. For this purpose, we classify the number of algorithm choices 0 to 4 as weakly leaning towards algorithm aversion, i.e. not choosing the algorithm, and the number of algorithm choices 5 to 9 as weakly leaning towards algorithm appreciation, i.e. choosing the algorithm (see Figure A3 in the Appendix). The shares of participants who chose the algorithm obtained in this way reveal a slight tendency towards algorithm appreciation in both conditions (62.26% of the participants in Human and 65.09% in Human-Algorithm). However, the difference between the two conditions is not statistically significant, $t(210) = -0.43$, $p = 0.67$, based on a t -test. This exercise shows that our suggested fine-grained method is able to elicit differences in algorithm preferences, where the commonly used binary measures might fail.

4. Conclusion

In this paper, we introduce a simple method for elicitation of human preferences for algorithms. By varying the size of payment reduction when the algorithm decision becomes payoff relevant in several choice problems and counting the number of choice problems in which the algorithm is selected, we obtain a measure of the degree of algorithm aversion (resp. appreciation). Our results reveal that binary measures of algorithm aversion typically used in experimental studies may be insufficient given the heterogeneity of people and the complexity of the underlying decision problems. The procedure we propose, captures this heterogeneity. It would also allow using the degree of algorithm aversion/appreciation as an explanatory variable in other applications, for example, when studying the use of algorithms in hiring decisions of managers (Dargnies et al., 2022) or the reliance on algorithm advice and performance in price estimation tasks (Greiner et al., 2022).

Forecasts are an important ingredient of several business decisions ranging from financial reporting to budgeting and production. As algorithm forecasts are proven to be superior to human forecasts in many economic applications (e.g., stock market forecasting), we apply the suggested method to measure the degree of algorithm aversion in a forecasting task in two environments that differ with respect to the information about the algorithm performance. Our findings suggest that an important factor in acknowledging the superior performance of algorithms is the improvement of the feedback on the algorithm performance. Providing more information about what algorithms do seems to help reduce human reluctance to rely on algorithm advice. Thus, identifying the appropriate channels in explaining how and why an algorithm is used and how it works is the natural next step in improving our understanding of the nature of human preferences for algorithms.

References

- Arkes, H.R., Dawes, R.M., Christensen, C., 1986. Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes* 37, 93–110.
- 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, 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, 2250–2258.
- Burton, J.W., Stein, M.K., Jensen, T.B., 2015. A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making* 33, 220–239.
- Chugunova, M., Sele, D., 2020. We and it: An interdisciplinary review of the experimental evidence on human-machine interaction. *Center for Law & Economics Working Paper Series* 12/2020 .
- Dargnies, M.P., Hakimov, R., Kübler, D., 2022. Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence. *CRC Discussion Paper* 334.
- Dietvorst, B.J., Bharti, S., 2020. People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science* 31, 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, 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, 1155–1170.
- Filiz, I., Judek, J.R., Lorenz, M., Spiwoks, M., 2021. Reducing algorithm aversion through experience. *Journal of Behavioral and Experimental Finance* 31.
- Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171–178. doi:10.1007/s10683-006-9159-4.
- Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association* 1, 114–125. doi:10.1007/s40881-015-0004-4.
- Greiner, B., Grünwald, P., Lindner, T., Lintner, G., Wiernsperger, M., 2022. Incentives, framing, and trust in algorithmic advice: An experimental study. *Working paper* .
- Johar, V., 2022. Artificial intelligence in hiring: A tool for recruiters. *Forbes Business Council* .
- Jung, M., Seiter, M., 2021. Towards a better understanding on mitigating algorithm aversion in forecasting: an experimental study. *Journal of Management Control* 32, 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.

- Kimbrough, S. O., W.D.J.Z.F., 2002. Computers play the beer game: can artificial agents manage supply chains? *Decision support systems* 33, 323–333.
- Kocher, M.G., Lahni, A.M., Trautmann, S.T., 2018. Ambiguity aversion is not universal. *European Economic Review* 101, 268–283.
- 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.
- Polli, F., 2019. Using AI to eliminate bias from hiring. *Harvard Business Review* .

5. Appendix

5.1. Additional figures

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

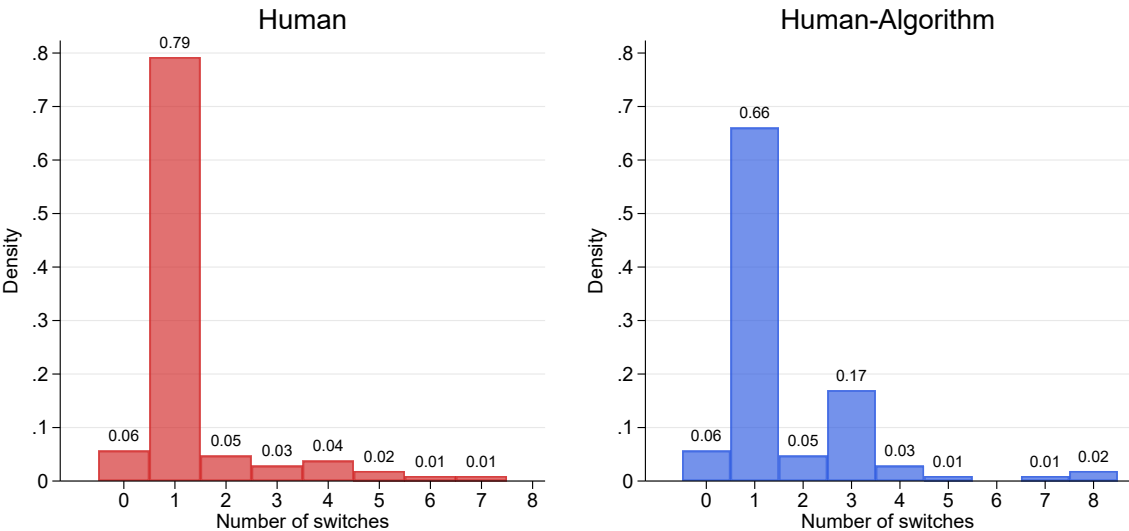


Figure A2: Observed and performance-based predicted number of algorithm choices per condition with fitted values (solid) and 45° line (dashed).

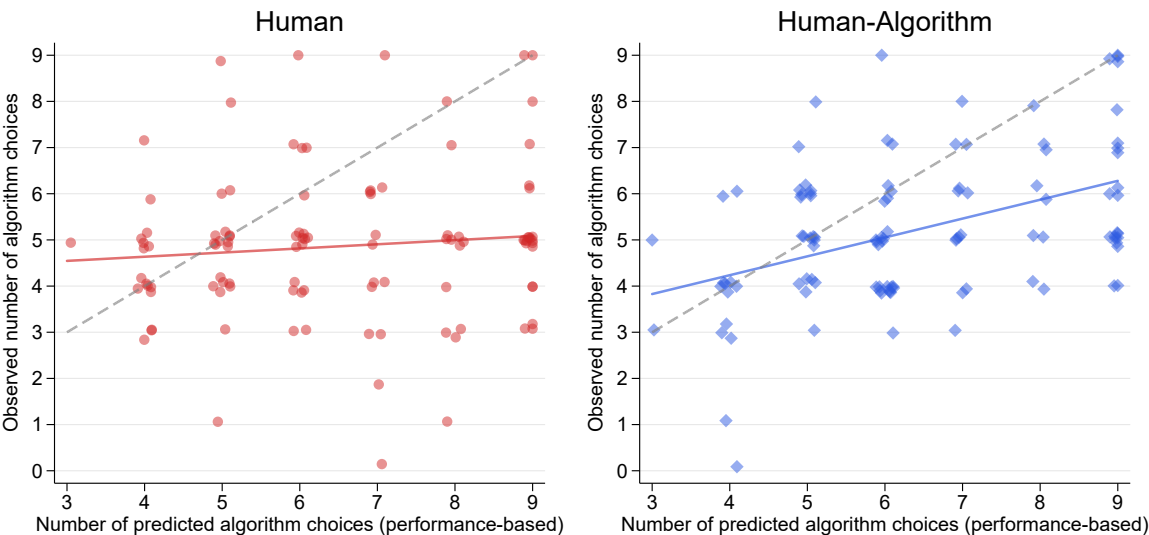
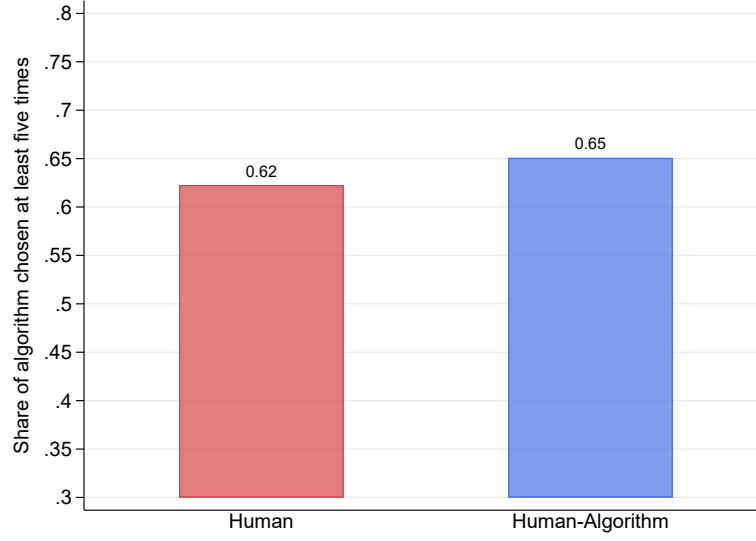


Figure A3: Binary classification: share of participants who chose the algorithm at least five times by condition.



5.2. Additional tables

Table A1: 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
<u>Unincentivized forecasts</u>			
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)	$t(953) = 23.42, p < 0.01$
<u>Incentivized forecast</u>			
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 A2: Belief, confidence and risk attitude questions

After Part 1

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 predict the state's rank almost perfectly? (1= *extremely unlikely*; 9= *extremely likely*)

How likely is it that the algorithm predicts the state's rank almost perfectly? (1= *extremely unlikely*; 9= *extremely likely*)

After Part 2

Are you generally a person who is willing to take risks or do you try to avoid taking risks?

(0=*Risk-averse*; 10=*Risk-prone*)

Table A3: Ambiguity Aversion Elicitation

Decision item	Ambiguous prospect with unknown probability p	Risky prospects with known probabilities q_i
1	$p = 50\%$	$q_1 = 10\%$
2	$p = 50\%$	$q_2 = 20\%$
3	$p = 50\%$	$q_3 = 30\%$
4	$p = 50\%$	$q_4 = 40\%$
5	$p = 50\%$	$q_5 = 50\%$
6	$p = 50\%$	$q_6 = 60\%$
7	$p = 50\%$	$q_7 = 70\%$
8	$p = 50\%$	$q_8 = 80\%$
9	$p = 50\%$	$q_9 = 90\%$

5.3. 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 with cursive font surrounded by brackets.

Welcome to the experiment and thank you for participating!

General information

The instructions are the same for all participants.

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. This means, that 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 will 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 least 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 eleven rounds.

In the first ten, 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 make a 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 **eleventh**, 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 $7\text{€} - X * |\text{Prognose} - \text{wahrer Rang}|$. 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 Entscheidung:	Auszahlung bei Wahl der Vorhersage des Algorithmus
1.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0\text{€} * \text{Prognose} - \text{wahrer Rang} $
2.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.03\text{€} * \text{Prognose} - \text{wahrer Rang} $
3.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.06\text{€} * \text{Prognose} - \text{wahrer Rang} $
4.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.09\text{€} * \text{Prognose} - \text{wahrer Rang} $
5.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $
6.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.15\text{€} * \text{Prognose} - \text{wahrer Rang} $
7.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.18\text{€} * \text{Prognose} - \text{wahrer Rang} $
8.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.21\text{€} * \text{Prognose} - \text{wahrer Rang} $
9.	$7\text{€} - 0.12\text{€} * \text{Prognose} - \text{wahrer Rang} $	Ihre eigene Vorhersage <input type="radio"/> Vorhersage des Algorithmus <input type="radio"/>	$7\text{€} - 0.24\text{€} * \text{Prognose} - \text{wahrer Rang} $

Bestätigen

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.

[*Human condition:*

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:

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

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.
- $X = 0.12\text{€}$ in all 9 choice problems, if you decide to make the forecast yourself.
- X takes one of the values $\{0\text{€}; 0.03\text{€}; 0.06\text{€}; 0.09\text{€}; 0.12\text{€}; 0.15\text{€}; 0.18\text{€}; 0.21\text{€}; 0.24\text{€}\}$ in the 9 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 have decided to make the forecast yourself, your own forecast will be used for the payout.

Take enough time for your 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 after the end of Part 2. The result of the eleventh 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 3th 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 have decided in line 3 that you make the forecast yourself. In this case, you will be deducted $0.12\text{€} \cdot 5 = 0.60\text{€}$. Your earnings are 6.40€.

Case 2:

You have decided in line 3 that the algorithm makes the forecast. In this case, you will be deducted $0.06\text{€} \cdot 5 = 0.30\text{€}$. 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 have decided in line 7 that you make the forecast yourself. In this case, you will be deducted $0.12\text{€} \cdot 5 = 0.60\text{€}$. Your earnings are 6.40€.

Case 2:

You have decided in line 7 that the algorithm makes the forecast. In this case, you will be deducted $0.18\text{€} \cdot 5 = 0.90\text{€}$. 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 eleventh 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 makes the forecast. The absolute value of $\text{forecast} - \text{true rank}$ is 10.	€
For example, suppose the computer has selected the choice problem on line 5. The person has decided in line 5 that she/he makes the forecast herself/himself. The absolute value of $\text{forecast} - \text{true rank}$ 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 makes the forecast. The absolute value of $\text{forecast} - \text{true rank}$ 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.

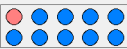
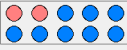







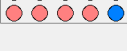
You have to choose one box at a time. Subsequently, one of the nine decision situations is randomly selected, whereby each decision situation is equally likely. Then a ball is randomly drawn from the box you chose in the drawn decision situation. Each of the 10 balls in the box can be drawn with equal probability. You get $2e$ if the color of the randomly drawn ball matches your personal color. If the ball is the other color, you get $0e$. 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		Box B	
1.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 10% Anteil blaue Bälle: 90%
2.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 20% Anteil blaue Bälle: 80%
3.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 30% Anteil blaue Bälle: 70%
4.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 40% Anteil blaue Bälle: 60%
5.	Anteil rote Bälle: unbekannt Anteil blaue Bälle: unbekannt	Box A <input type="radio"/> Box B		Anteil rote Bälle: 50% Anteil blaue Bälle: 50%
6.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 60% Anteil blaue Bälle: 40%
7.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 70% Anteil blaue Bälle: 30%
8.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 80% Anteil blaue Bälle: 20%
9.		Box A <input type="radio"/> Box B		Anteil rote Bälle: 90% Anteil blaue Bälle: 10%

Bestätigen