Consumer Choice between Recommendation Algorithms - Experimental Evidence

Felix Schleef*

July 13, 2026

[Link to Latest Version]

Abstract

Regulators are increasingly concerned with the power of large online platforms to bias consumer recommendations. In light of these concerns, I study experimentally whether giving consumers the choice between several recommendation algorithms can improve consumer welfare. I develop a novel experimental setting in which I first elicit the subjects' preferences and then expose them to choice tasks intermediated by personalized recommendation algorithms. The experiment achieves the desired variation - when faced with an algorithm that optimizes for the users utility alone, the users choose alternatives that have higher expected value and that are higher ranked, compared to when they face a biased algorithm. When faced with a costly choice between two algorithms, subject' willingness-to-pay for better recommendations is dispersed but positive on average. The findings suggest that giving consumers power over recommendation algorithms to curtail potential abuse is not straightforward. However, it may be a viable business for platforms to offer improved recommendation algorithms to consumers for a fee.

*HEC Paris, CREST Paris. E-mail: felix.schleef@gmail.com. I would like to thank Roxana Fernández Machado, Thibaud Vergé, Michelangelo Rossi, Louis Pape, Guillaume Hollard and Yves le Yaouanq. This research was supported by funding from the ANR-21-CE26-0018-01.

1 Introduction

Choices on online platforms are often too numerous for consumers to compare all relevant products (Natan, 2024). When consumers search products, platforms therefore employ recommendation systems. Recommendation systems shape which products a consumer sees and in which order they are searched (Ursu, 2018). While these algorithms can be used to help consumers, platforms can also exploit the consumers' reliance on recommendations, for example by boosting the ranking of products the platform sells itself (self-preferencing) (Peitz, 2023, Lam, 2021).

Regulators in the recent past have become increasingly concerned with self-preferencing. So far, the answer from European regulators has been to ban self-preferencing for large platforms through the Digital Markets Act (*Digital Markets Act*, 2022). However, this regulation may be difficult to enforce, difficult to comply with[1] and leaves other consumer-welfare related concerns, such as price discrimination through recommendations, unaddressed.

In light of these concerns, I study an alternative proposal for regulation, which would enable the consumer to choose between different recommendation systems. In recent times most prominently discussed by Fukuyama and coauthors as "middleware" (Fukuyama, Richman, Goel, Katz, Melamed and Schaake, 2020), the concept of enabling consumers to have more control over their interaction with online services is not new[2]. In the context of recommendation systems, Resnick and Varian (1997) envisioned a future, where e-commerce stores and recommendation systems would be two separate entities, owned and operated by two different firms. The e-commerce stores would be competing intensely with each other, and the consumer could use the recommendation system that they perceive to best fit their needs. Nowadays, e-commerce platforms such as Amazon or content platforms like Spotify

[1] Given that the past purchase data on large platforms was generated in a setting with potentially biased (or self-preferencing) algorithms, it may be difficult to generate truly unbiased rankings, even when using an unbiased algorithm. How to de-bias data for recommendation applications is an active area of research (Chen, Dong, Qiu, He, Xin, Chen, Lin and Yang, 2021, Lin, Liu, Pan and Ming, 2021).

[2] This paper also relates to recent contributions of Bergemann, Cremer, Dinielli, Groh, Heidhues, Schafer, Schnitzer, Morton, Seim and Sullivan (2023), Esber, Kominers, Kornfeld, Belsky, Chitnis, Chmielinski, Deighton, Eliot, Rossini, Searls, Shah and Williams (2024) and Posner and Weyl (2018) that consider ways for consumers to better govern and make use of their personal data.

and Netflix come bundled with their own recommendation system, and consumers do not have an alternative recommendation system to choose from, unless they leave the platform. This leads to the question whether it would be a promising policy idea to request large platforms to let consumers choose from multiple recommendation systems. Apart from how the incentives for platforms might change, for such a policy to be effective it is necessary that consumers are able to judge different recommendation systems well.

I use a laboratory experiment to study how subjects use and assess different recommendation algorithms. I first elicit the subjects risk aversion, which I then use to create and order lists of three-outcome lotteries with two different algorithms: one that selects and orders lotteries based on the subjects expected utility (expected utility algorithm), and one that weighs the expected utility of the subject and the objective of the platform (platform algorithm). I then let subjects make choices from lists of lotteries that have been selected and ordered by these two algorithms. After 10 rounds of the choice task with each algorithm, I elicit their willingness to pay for either algorithm - the subject gets to change the probability with which they are going to face either the expected utility or the platform algorithm for a last set of 10 choice tasks.

While lotteries are not products, the choice consumers face online is often lottery-like. Once a consumer has settled on a product's attributes, they are still confronted with a distribution of user ratings. Choosing between several products with different ratings distributions is effectively choosing between safer and riskier lotteries over one's own future experience. I further explore the links between the experimental design and the platform consumer choice problem in section 3.3 once the recommendation algorithms are defined.

I document three key findings. First, I find that in both algorithmic treatments, subjects are more likely to choose items that are towards the top of the list. This is consistent with our empirical knowledge of consumer behavior on online platforms, where defaults and particularly salient options are more likely to be chosen. Second, the platform algorithm leads subjects to choose lotteries with lower expected values, and lotteries that are further down the list of ordered lotteries. The experiment therefore successfully creates a setting

where the platform algorithm induces worse outcomes for the subjects, where subjects adapt their behavior to the different algorithms, and where a hypothetical platform would gain from the adoption of such an algorithm. Third, subjects have a positive average willingness-to-pay for the expected utility algorithm, with substantial heterogeneity. This willingness-to-pay is, however, not very responsive to differences in the gain the algorithm offers a given subject. Willingness to pay thus reflects factors other than the algorithm's value to the consumer.

The contribution of this project is twofold. First, I develop a parsimonious experimental paradigm for studying *personalized* recommendation algorithms: existing field experiments cannot observe algorithms' objectives, while existing lab experiments typically study non-personalized rankings. By eliciting risk preferences and using them to construct personalized rankings in real time, my design gives the researcher control over both the algorithm's objective and knowledge of the consumer's preferences, enabling clean measurement of how algorithmic (mis)alignment affects consumer outcomes. The framework accommodates varying degrees of congruence or conflict between platform and consumer interests de Cornière and Taylor (2019) and can be extended to information disclosure, advertising, or framing interventions. Second, I apply it to ask whether consumers can choose effectively between algorithms that differ in alignment with their interests. Although average willingness to pay for the better algorithm is positive, it is largely unrelated to the algorithm's actual value to the consumer, who does not reliably identify or price the alignment of the recommender they face. Simply offering consumers a choice between algorithms may therefore be insufficient to protect consumer welfare.

1.1 Literature

I contribute to a growing empirical literature on rankings and platform recommendation systems (Donelly, Kanodia and Morozov, 2023; Zhang, Ferreira, Matos and Belo, 2021; Farronato, Fradkin and MacKay, 2023; Reimers and Waldfogel, 2023; Lee and Musolff, 2021; Greminger, 2022; Greminger, 2023; Kaye, 2024; Ghose, Ipeirotis and Li, 2014; De los

Santos and Koulayev, 2017; Derakhshan, Golrezaei, Manshadi and Mirrokni, 2022). Ursu (2018) studies the effect of rankings on consumer choices using experimental data from the online travel platform Expedia. A subset of consumers looking for hotels were shown a random ranking, while the others were shown a ranking by "relevance" according to Expedia. The subjects were not aware that their rankings were created differently. She finds that rankings affect consumer search, but do not affect purchase decisions after controlling for search. Donelly et al. (2023) investigate the effects of recommendation systems using an experiment on the furniture shopping platform Wayfair. The experiment randomly assigned shoppers to see either personalized or non-personalized rankings of products. They find that personalized rankings induce more active consumer search and increase purchase probabilities. Importantly however, the experiment was conducted unbeknownst to the user - the users may have expected their product search results to be personalized based on their experience on similar E-commerce websites, or they might have changed their search behavior if they had known that their rankings are not personalized. Zhang et al. (2021) run an experiment on a Video-on-demand system. They randomize both the slot (i.e. the ranking of the movie, how saliently it is presented) and the price of the movie. They find that movies that are prominently displayed have less price elastic demand. Similarly to the previous studies, subjects are not aware that the algorithm has changed for them as compared to when they used the same service before the researchers intervention. Farronato et al. (2023) document that products of Amazon brands are more prominently displayed than other products. They also show that steering can be potentially very effective, since in 72.1% of searches consumers stay on the first page of search results.

This paper also relates to the emergent experimental literature on platform recommendation systems. Boldrini and Clavorà Braulin (2025) conducts a framed experiment related to the Amazon "Buybox" to study the effect of information provision about the algorithm on consumer choices. Düll, Karle, Martin and Schumacher (2025) study the performance of search cost estimation methods when subjects have priors about the returns to search (which could be induced by recommendation algorithms). Most closely related to my work is the paper by Fong, Natan and Pantle (2024), where they conduct an experiment in or-

der to disentangle position specific search costs from beliefs about the recommendation (or ranking) algorithm. They find that both mechanisms are important, and that not accounting for beliefs leads to overestimating position-specific search costs. Lastly, they also show that consumers learn about algorithms in experimental settings where the recommendation algorithm changes, and they are neither informed about the change nor the algorithms. Their findings are important for the simulation of counterfactual recommendation procedures, as overestimated search costs lead to overestimating gains from potentially better algorithms. Instead, I elicit preferences for different algorithms directly.

Regarding the literature on recommendation algorithms beyond their use as ranking tools on platforms, I relate to Chen, Wu and Zhong (2023), from whom I borrow the three outcome lottery design. Chen et al. (2023) study how subjects make binary choices between three outcome lotteries when they can use recommendation systems. They find, that subjects tend to follow the recommendations, and that they make better and faster decisions when they have access to recommendations. They also document that subjects are willing to pay a fee to have access to the recommendations. To generate recommendations, they use algorithms that are employed in real-life recommendation systems, such as collaborative filtering. In contrast to their research, I do not consider technical intricacies of different recommendation algorithms. I start from the view that platforms use accurate utility estimates for their recommendations, and then investigate how consumers choose from ordered lists, rather than how they make a binary choice. Their finding that consumers are willing to pay for recommendations rather than not having recommendations is an important result, and and I will be able to contrast this with my findings on whether subjects are willing to pay for a recommendation system that is aligned with their preferences rather than one that is not aligned. Caplin, Dean and Martin (2011) study how complexity affects choices from lists. They show that many choices from lists are consistent with sequential search, which is corroborates the steering power of rankings. While I will not vary the complexity of choices in my experiment, the experimental setting could be easily amended to vary complexity by increasing the number of lottery choices. Kaye (2024) studies outcomes of firms and consumers with recommendation algorithms where firms price endogenously. He

finds that personalization improves the matches of consumer to products, but it implies that firms are matched with consumers that have more inelastic demand for their offering. Facing this new demand, firms find it optimal to increase their price, negating the welfare benefits that consumers derive from the improved matches. Importantly for my experiment, he emphasizes the role of consumer expectations about recommendation algorithms. In experiments on platforms like Expedia, consumers may expect the items to be ordered by a recommendation system, and may infer something about unobserved product attributes from the items position in a list, even if the recommendation system has been deactivated for the user as part of an experiment. I therefore emphasize to subjects in the experiment, that the items on the lists that they are able to choose from are selected and ordered by different algorithms.

I also relate to an active theoretical literature on biased intermediation on platforms (Hagiu and Jullien, 2011; Hagiu and Jullien, 2014; Armstrong, Vickers and Zhou, 2009; Armstrong and Zhou, 2011; Reimers and Waldfogel, 2023; de Cornière and Taylor (2019); Bergemann and Bonatti (2023)). Theoretical contributions highlight consumer heterogeneity and the role of information in platform intermediation. de Cornière and Taylor (2019) outline a model where recommendations by the intermediary can either harm or benefit consumers, depending on whether the intermediaries' and the consumers' interests are conflicting or congruent. Heidhues, Köster and Kőszegi (2023) focus on how recommendations can affect consumers when they make mistakes. Mistakes in this model fall into two categories - either the consumer purchases something that they should not have purchased, or they do not purchase something that they should have purchased. However, this model abstracts from consumer search. Bourreau and Gaudin (2022) study biased intermediation on streaming platforms. In this setting, the user does not pay per song or movie they stream, but the platform bears a marginal cost for each stream. The platform chooses which items to recommend, and the subscription price. They show that the platform optimally steers consumers to items with lower marginal costs. In the model of Bergemann and Bonatti (2023), the platform has an informational advantage over the user - the platform knows the user-item match specific utility exactly, while the user is uncertain. In this setting, the

platform matches products and users with targeted ads, and monetizes by charging sellers for the advertising. Reimers and Waldfogel (2023) develop a discrete choice framework that sheds light on how biased intermediation can be studied without data from inside the platform. They trace out the a welfare frontier, where total welfare is divided by between buyers and sellers on the platform, and label departures from this welfare frontier as biased intermediation. Using simulations, they illustrate that regressing rank on observable product characteristics as well as a dummy that equals to 1 if a product is sold by the platform (or "platform-preferred") does not produce reliable evidence for or against biased intermediation.

While I do not study biased intermediation theoretically, my experiment relies on these contributions in important ways. Firstly, as in de Cornière and Taylor (2019), I study a setting where alignment of interest between the user and the platform vary from user to user. In the experiment, users that are risk neutral or close to risk neutral have completely conflicting interests with the platform, while the interests of either very risk averse or risk seeking consumers are more aligned. I will be able to study if consumers are more or less likely to "make mistakes" or choose options the lead to a lower expected utility based on certain characteristics, as in Heidhues et al. (2023), and even though I abstract from "mistake-based" steering, the experimental setting could be extended to include such alternative steering algorithms. As in Bourreau and Gaudin (2022) I study a setting where user choose items at no cost. The choice between different recommendation systems, where the recommended items are free for the user is similar to platform competition between streaming services. For the design of the algorithms in the experiment, I rely extensively on Reimers and Waldfogel (2023). In my conceptual framework, I use their simple logit formulation of rank-dependent expected utility, and I rely on their results on consumer- and seller optimal algorithms for the design of my experiment. Lastly, mirroring Bergemann and Bonatti (2023), I assume that the platform knows the users preferences, and can therefore compute the (rank-independent) utility of each choice for each user.

The paper proceeds as follows. In Section 2, I describe a conceptual framework motivating the different recommendation algorithms used in the experiment. Section 3 describes the

design of the experiment. In section 4, I show the results. Section 5 concludes.

2 Conceptual Framework

Just as consumers on e-commerce platforms, subjects in my experiment will be faced with lists of items. While in the platform case, these items are likely to be products (or songs or movies), in the case of my experiment, they are three-outcome lotteries. The outcomes of the lotteries are $x_1 = \$0$, $x_2 = \$5$ and $x_3 = \$10$, and I denote the corresponding probabilities as p_1, p_2 and p_3. I follow Reimers and Waldfogel (2023) logit approach in modeling the subject's utility. The subject chooses among L lotteries based on the lotteries characteristics, and its ranking in the list r_j. I model the subject's utility of choosing a lottery j ranked at r_j as:

$$u_{ij} = \delta_{ij} + \gamma_i r_j + \xi_j + \epsilon_{ij} \tag{1}$$

where γ_i denotes the user-specific effect of rank on utility, ξ_j is are unobserved lottery attributes (for example, lotteries where the probability of one outcome is equal to 0 might be more easy to understand for subjects, leading to them being chosen more frequently), and ϵ_{ij} is an extreme-value error. Lastly, δ_{ij} denotes the rank-independent component of utility, which I choose to model using a constant relative risk aversion (CRRA) utility function:

$$\delta_{ij} = \begin{cases} \sum_{k=1}^{3} p(c_k) \frac{c_k^{1-\omega_i}}{1-\omega_i} & \text{if} \quad \omega \neq 1 \\ \sum_{k=1}^{3} p(c_k) ln(c_k) & \text{if} \quad \omega = 1 \end{cases}$$

The variable k indexes the three different outcomes (c_k) of lottery j, and ω_i is the individual-specific risk aversion. Given that there is no outside option in the experimental setting, the choice probability for a lottery j ranked at r_j can be computed as:

$$P_{ij} = \frac{e^{v_{ij}}}{\sum_{j \in J} e^{vij}}$$

with $v_{ij} = \delta_{ij} + \gamma_i r_j + \xi_j$. The ranking of a lottery affects the subjects utility of choosing a lottery, and therefore also the probability that the lottery will be chosen.

2.1 Ranking algorithms

In the experiment, subjects face two different algorithms. The first algorithm, which I will call the expected utility algorithm, is fully aligned with the interests of the subjects and orders lotteries. The other algorithm, which I will call the platform algorithm, weighs consumer utility and the expected cost of providing a lottery to a subject.

Expected Utility Algorithm:
The expected utility algorithm ranks lotteries by the following index:

$$I_{ij} = \underbrace{\sum_{k=1}^{3} p(c_k) \frac{c_k^{1-\omega_i}}{1-\omega_i}}_{\delta_{ij}} \tag{2}$$

Reimers and Waldfogel (2023) show that this ordering by *rank-independent* expected utility is optimal for consumers.

Platform Algorithm:
The platform algorithm ranks lotteries by the following index:

$$I'_{ij} = \kappa_1 S_u(\delta_{ij}) - \kappa_2 S_\pi \left(\sum_{k=1}^{3} p(c_k) c_k \right) \tag{3}$$

where κ_1, κ_2 are the weights on the consumer and the "platform" objective respectively (in the experiment, $\kappa_1 = \kappa_2 = 1$), where $S_\pi(.)$ and $S_u(.)$ are min-max scalers: $S_u(x) = x \frac{1-\omega_i}{10^{1-\omega_i}}$ and $S_\pi(x) = \frac{x}{10}$. The minimums and maximums correspond to the profits and utilities in the case where the subject receives the degenerate lottery $(p_1 = 1, p_2 = 0, p_3 = 0)$ and $(p_1 = 0, p_2 = 0, p_3 = 1)$ respectively. Intuitively, these scaling functions are required,

because otherwise the value that is associated with the utility of the subject could be much higher or lower than the number associated with the platforms profit, since it is scaled by their risk aversion. However, there is no reason why a platform should disadvantage users with a given set of preferences more than others, therefore the values are appropriately rescaled.

To see how platform algorithm orders lotteries, note that the consumer's expected utility equals the utility of their certainty equivalent, $\delta_{ij} = u(CE_{ij})$. Substituting into the scalers (and setting $\kappa_1 = \kappa_2 = 1$ as in the experiment), the index collapses to

$$I'_{ij} \;=\; \left(\frac{CE_{ij}}{10}\right)^{1-\omega_i} \;-\; \frac{EV_{ij}}{10}. \tag{4}$$

The first term is what lottery j is *worth* to consumer i, the second term is what the lottery *costs* the platform in expectation. The algorithm therefore ranks lotteries by the difference between subjective value and expected cost. The lotteries that are ranked highly are lotteries that the consumer values highly but that are also cheap for the platform to provide.

This difference is larger when the subject has more extreme risk preferences. A risk-averse subject may derive the same utility from two lotteries - a risky lottery with relatively high expected value, and a safe lottery with lower expected value. The platform however is risk-neutral - even though the consumer is indifferent between the two lotteries, the lottery with the lower expected value is cheaper for the platform to provide, and accordingly it is the lottery that will be ranked higher by the platform algorithm. For a risk-seeking subject the logic is the same - the platform can offer a higher risk, lower expected value lottery over a safer, higher expected value alternative.

Risk-neutral consumers pose a problem for the platform algorithm - since they value a lottery only by its expected payout, their objective is exactly opposed to the platform's objective. If $\omega \to 1$, $CE_{ij} \to EV_{ij}$, which means that and $I'_{ij} \to 0$ for *every* lottery. For the experiment, risk-neutral subjects were randomized to $\omega_i = \pm 0.01$ to be able to generate an ordering that is different from random in the platform algorithm treatment rounds.

3 Experimental Design

The experiment features three kinds of tasks: a risk preference elicitation task, a choice task, and a willingness-to-pay elicitation task.

The *risk preference elicitation task* follows Johnson, Baillon, Bleichrodt, Li, van Dolder and Wakker (2021). As shown in figure A11, subjects are presented with the prospect of gaining either the outcome of a lottery that pays \$10 or \$0 with equal probability or receiving a sure but unknown monetary amount X, randomly drawn with uniform probability between \$0 and \$10. Subjects are asked to declare a threshold such that they would prefer receiving X if it is above the threshold, and they would prefer receiving the lottery if X is below the threshold. The threshold of a risk neutral is \$5, whereas the thresholds of risk averse (risk seeking) individuals are lower (higher) than \$5. Their choice in this task is used to compute their coefficient of risk aversion ω, assuming CRRA preferences.

The central task of this experiment is a *choice task*, which subjects complete 30 times. An example of the choice task is shown in figure 1. It is designed to resemble choices from ordered lists on online platforms. For each round of this task, subjects are asked to choose from lists of three-outcome lotteries. The experiment uses 30 sets of 20 lotteries each with with payoffs $(\pi_1 = 10\$, \pi_2 = 5\$, \pi_3 = 0\$)$ probabilities $(p_1 \leq 0.5, \quad p_2 \leq 1, \quad p_3 \leq 0.5)$. In each round, the lotteries have been ordered by either the platform or the expected utility algorithm as described in section 2. The top 10 lotteries according to the algorithms ranking criterion are displayed to the subject. Subjects are not informed about any properties of the algorithms, but they are aware of the different algorithms that select and order the lotteries. Throughout the choice task, subjects see either "Algorithm 1" or "Algorithm 2" at the top of the page, colored green or blue (the color order is randomized across individuals). They can thus learn through there repeated interactions whether algorithm 1 or algorithm 2 creates better recommendations for them.

After the first 20 rounds of the choice task, I elicit subjects willingness-to-pay for one algorithm over the other. For this step, I give subjects an endowment of \$1. Using a slider

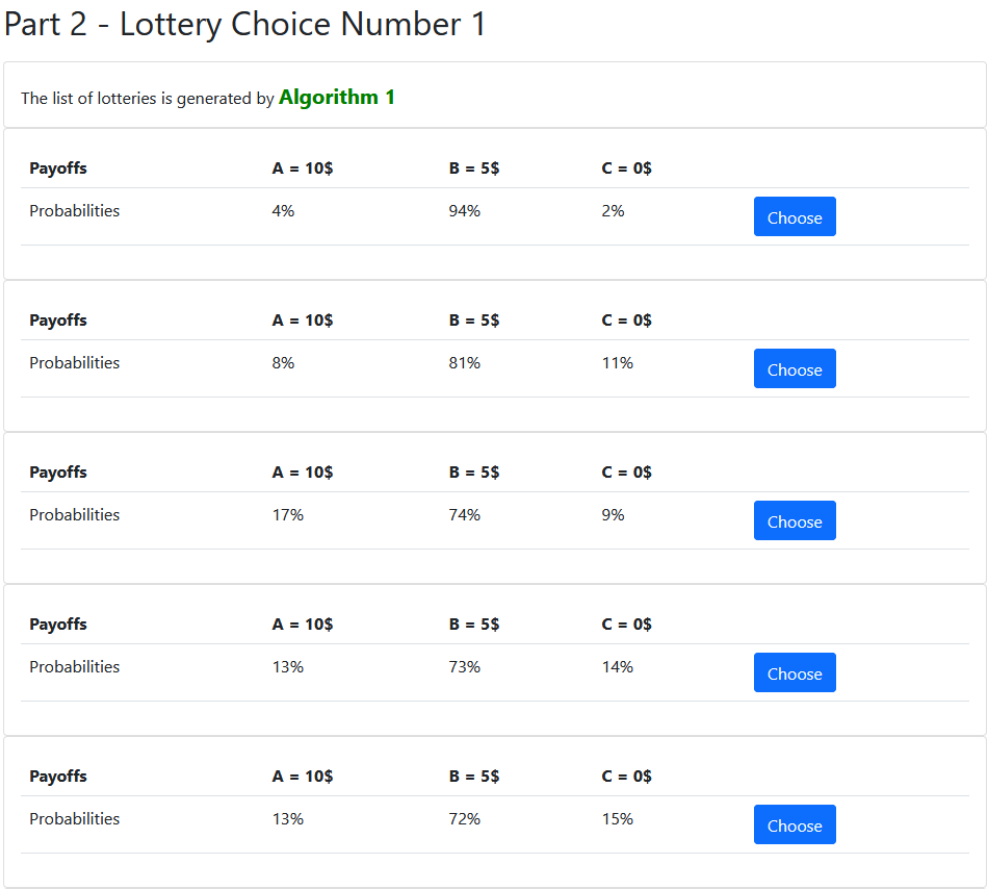


Figure 1: A screenshot of round 1 of the choice task for a risk-averse subject. The lotteries are selected and ordered by the platform algorithm which ranks safe but low expected value lotteries high. The subject sees that the lotteries in this round of the choice task are selected and ordered by algorithm 1.

with no default, subjects can change the probability that they will face either algorithm 1 or algorithm 2 for the last set of 10 choice rounds. Choosing equal probability between the two algorithms (50% - 50%) is costless. Changing the probabilities by 1%, i.e. to 49% - 51% costs \$0.02. The cost increases linearly, so that the cost of setting the probability to 1 to have either algorithm for the last 10 choice rounds is equal to the full endowment of \$1.

The cost schedule has two implications. First, both the cost and the expected gain from having the better algorithm are linear. Therefore, a subject should choose to pay the full endowment if the benefit of the better algorithm exceeds the cost. Given that it is costless to select 50-50, it is only worthwhile to pay for an algorithm if their value differs by more than \$2. Second, the lottery menus are fixed in advance, so we can evaluate the *monetary* difference in the lotteries recommended by the two algorithms ex-ante. As graph 2 shows, for every risk preference, the value difference between the two algorithms' top recommendations falls short of \$2. The payoff-maximizing report is thus the costless 50–50 position for every subject. Subjects can still rationally deviate from the 50-50 benchmark, most importantly if they identify differences in time spent and therefore opportunity cost when choosing with the expected utility rather than the platform algorithm. Figure 2 plots a hypothetical benefit curve if subjects save six seconds per round (one minute for the block of 10 choice rounds) when choosing with the expected utility rather than the platform algorithm (evaluated in \$ terms using the realized average payoff of the experiment). In addition, subjects may also find searching the ranked alternatives more cognitively costly when faced with the platform rather than the expected utility algorithm.

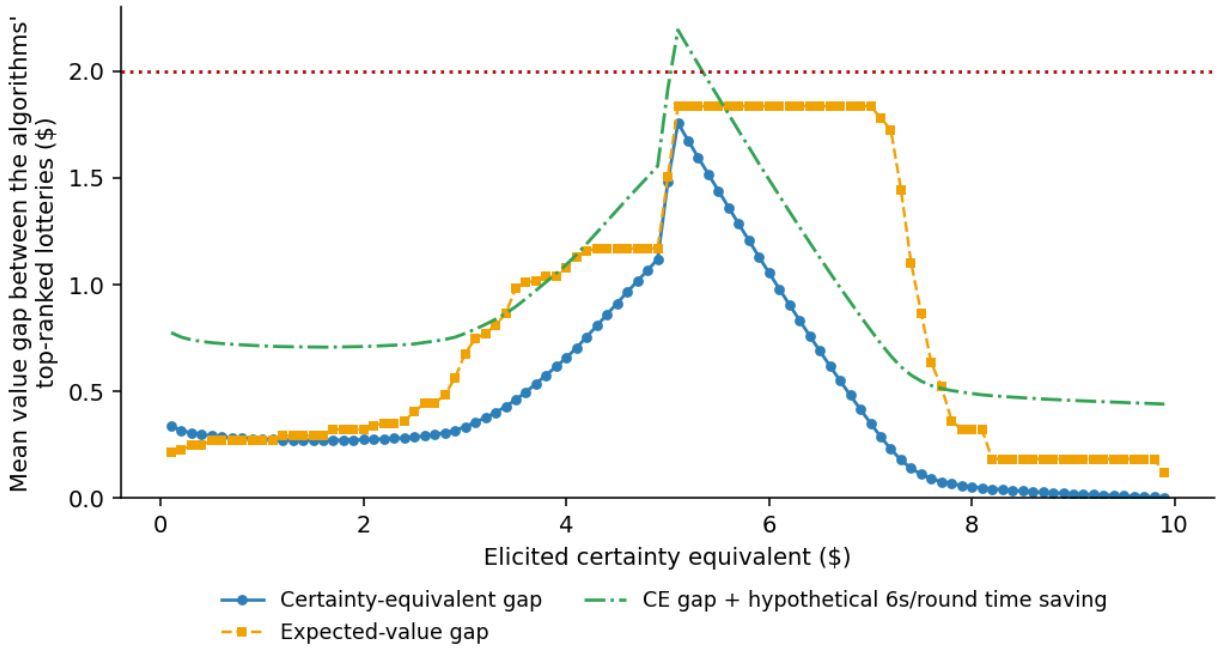


Figure 2: Ex-ante average value difference in the top recommendation of the expected utility and the platform algorithm for each possible choice of certainty equivalent (CE). CE=5 signifies risk neutrality, CE<5 risk-averse, and CE>5 risk-seeking preferences.

3.1 Incentives

Subjects receive a fixed 4$ payout for completing the experiment. In addition, subjects receive a bonus that depends on their choices in the experiment.

For the bonus payment, I divide the experiment into three parts, of which one is randomly selected with equal probability. The three parts are risk preference elicitation ($t = 2$), the first 20 rounds of the choice task ($t = 3$, $t = 4$), and the willingness-to-pay elicitation task including the final set of choice rounds ($t = 5$, $t = 6$), as outlined in table 1.

In case the randomly selected part is part 1, subjects get a bonus according to their choice. In case the randomly drawn number X was above their threshold, they receive the X as the bonus, otherwise they receive the outcome of the lottery that pays 10$ with probability 50% (and 0$ with probability 50%).

If the randomly selected part is part 2, I select 1 of the 20 rounds at random and the subject

Timing	Description
t=1	Instructions • Consent form and general Instructions about the experiment - figure A3, A4 • Instructions on the risk elicitation task - figure A5, A6 • Instructions on the lottery choice task - figure A7 • Instructions on the Willingness-to-pay (WTP) elicitation mechanism - figure A8, A9
	Start of the main experiment - figure A10
t=2	Risk preference elicitation task - figure A11
t=3	10 rounds of choice task - figure A13 Choice task with either platform or expected utility algorithm (in case of platform algorithm, expected utility algorithm in the next set of 10 choice tasks and vice-versa; order randomized).
t=4	Buffer Screen informing subjects of a change in the algorithm - figure A14
t=5	10 rounds of choice task - figure A15
t=6	Willingness to pay elicitation - figure A16, A17 • Subjects receive an endowment of 1$ • Subjects can choose the probabilities $(x,\ 1-x)$ that they face the expected utility algorithm or the platform algorithm respectively for the next set of 10 choice tasks. • The cost of changing the probability is equal to $p = \$(x - 0.5) \times 2$
t=7	10 rounds of choice task With either expected utility of platform algorithm based on a random draw with the probabilities (x, 1-x).
t = 8	Survey - figure A18, A19
t = 9	Disclosure of the realization of the bonus payment, end of the experiment - figure A20

Table 1: Timing of the experiment

gets the outcome of the lottery they chose in that round. For example, a subject may have chosen the lottery [P(10$)=0.03, P(5$)=0.9, P(0$)=0.07] in the randomly selected round. The outcome drawn based on these probabilities is the subjects bonus payoff.

Lastly, if the randomly selected part is part 3, I add the unspent endowment from the willingness-to-pay elicitation to their bonus payoff. Furthermore, I draw one of the last 10 choice rounds at random, and the outcome drawn based on the probabilities of the lottery that they chose in that round is added to the bonus payoff as well.

The goal of randomly selecting one section and one round of the choice task is to avoid paying out the results of multiple lotteries, which could lead to subjects exhibiting higher risk appetite than in other situations since the risk is diversified (Azrieli, Chambers and Healy, 2018). I minimize this problem by randomly selecting one part to reward. As suggested by Plott and Zeiler (2005), subjects will be exposed to training and practice rounds with the elicitation mechanism in an anonymous fashion. The practice rounds will however not be incentivized, as this would lead back to the issue of paying out multiple lotteries.

3.2 Hypotheses

The primitive data set consists of each participant's choice in the risk-preference elicitation task. Their choices in the 30 rounds of the choice task, as well as their elicited willingness-to-pay. For the choice task, the dataset contains the full list of lotteries that the subject was presented with in each round and their respective rank. I further compute the following values: using the participant's choice in the risk elicitation task, I compute their risk aversion assuming a CRRA utility function. Using this computed coefficient of risk aversion, I compute the average expected utility of all lotteries for each subject and section of the experiment ($t = 3$, $t = 4$, $t = 6$), as well as the corresponding average expected value and variance of the lotteries.

The hypotheses were pre-registered in the AEA RCT Registry (AEARCTR-0014017) (see appendix A.4).

The first group of hypothesis concerns behavior and outcomes in the two algorithmic treatments. By construction, the platform algorithm recommends lower expected value lotteries. Therefore, under the platform algorithm, subjects choose lotteries with lower expected value than under the expected utility algorithm, and choose options placed further down the ranked list.

The second group of hypothesis concerns the valuation of the two algorithms.

Firstly, while we show in section 3, that the realized value differences between the two algorithms' recommendations lie below the cost of shifting the algorithm assignment for every elicitable risk preference, the payoff difference is only one aspect of the value that is provided by the expected utility algorithm. Taking into account opportunity and effort costs, I expect average willingness-to-pay for the expected utility algorithm to be positive.

Secondly, the platform algorithm disadvantages close to risk-neutral subjects more than risk-averse (or risk-seeking) subjects. I therefore expect, willingness-to-pay for the expected utility algorithm to be higher for subjects close to risk neutrality than for strongly risk-averse or strongly risk-seeking subjects.

3.3 Discussion

The goal of the experimental design is to map onto a common e-commerce decision. Consumers routinely settle on a product's attributes and a price range. They then choose among options that differ mainly in the distribution of user ratings (see figure 3 for examples of such rating distributions). Distributions of user ratings have a vertical and a horizontal differentiation component - while all consumers prefer a shift of the ratings distribution towards five stars, consumers have different preferences about the shape the ratings distribution at a given average rating. Risk-averse consumers may prefer a product with many four star ratings to a product with some five star and some one star ratings, and risk-seeking consumers may prefer the reverse.

The experimental design mirrors this decision problem. The vertical differentiation component of the lotteries is their expected value, the horizontal component their risk profile. The expected utility algorithm orders items according to the subjects' expected utility, while the platform algorithm satisfies the horizontal but exploits the vertical differentiation dimension: it fully satisfies the subjects' risk preference at the cost of decreasing the recommended lotteries' expected value.

The experimental setting is deliberately easier for the consumer than the field: the outcome

(a) Rating distribution of with most ratings either 4 or 5 stars

(b) Ratings distributions with a substantial share of 1 star ratings

Figure 3: Two products with the same attributes but different user-rating distributions are similar to a choice between lotteries over future experienced utility. The rating distribution in panel (3a) concentrates on five-star outcomes, which makes it a relatively safe lottery. The rating distribution in panel (3b) has a substantial mass on one-star outcomes, making it a riskier lottery with a higher chance of a poor experience.

probabilities are the only attributes that vary between the choice options, whereas in e-commerce choice settings price and differences in item characteristics would also play a role. If consumers cannot identify or value the better-aligned ranking even in this experimental setting, I expect the difficulty in richer environments to be even greater.

3.4 Data

Participants were recruited on Prolific in four waves during July 2024, with eligibility restricted to U.S.-based workers. Removing subjects that did not finish the experiment, subjects that voided their participation in the experiment, and subjects that failed the embedded instructed-response items (attention checks) during the choice task, 296 participants completed the experiment. Apart from a small technical pilot used only to validate the software, no prior data were collected, and the four waves above constitute the entire data collection.

Of the 296 individuals, 113 have made a choice in the risk elicitation task that is consistent

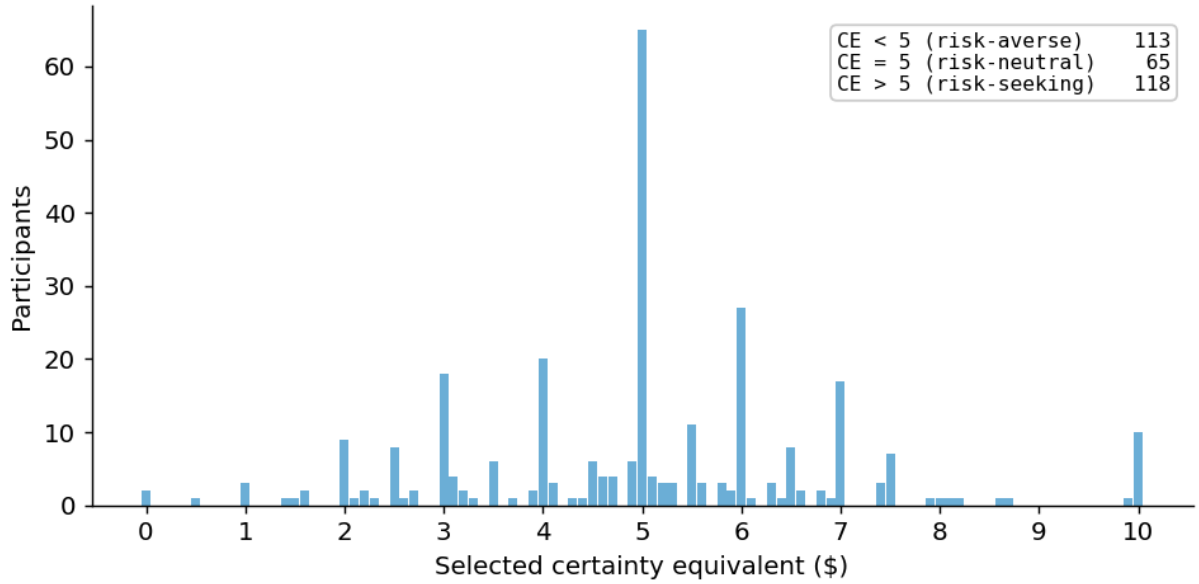


Figure 4: Frequencies of choices in the risk preference elicitation task

variable	n	mean	sd	median	min	max
Completion time (min)	284	22.29	11.54	19.62	2.77	70.87
Bonus payment (USD)	284	5.73	3.35	5.00	0.00	10.00

Table 2: Completion time and bonus payments (final sample).

with risk-aversion, 65 have made a risk-neutral choice, and 118 have made a choice that is consistent with risk-seeking preferences. Twelve subjects selected either 0 or 10 as the fixed threshold that they would prefer over a 50-50 lottery with the payoffs \$0 and \$10. These choices are only consistent with very extreme risk preference profiles and have therefore been excluded from the analysis. This leaves a final analysis sample of N = 284.

Given this sample, the average completion time of subjects that finished the experiment was 22.3 minutes and the average bonus payment was \$5.73 (yielding an average reward of \$9.73 together with the base payment of \$4 for completing the experiment).

4 Results

The analysis uses the 284 participants of the final sample, and, unless noted otherwise, the first twenty rounds of the choice task for each subject. The last ten rounds of the choice task are omitted, since the willingness-to-pay elicitation step would introduce selection bias. Throughout, we use the elicited certainty equivalent relative to risk neutrality, $(CE-5)$, as the measure of risk preference: negative (positive) values indicate risk aversion (seeking).

Section 4.1 examines behavior and outcomes under the two algorithms and section 4.2 examines subjects' choices in the willingness-to-pay elicitation task.

4.1 Behavior and outcomes under the two algorithms

Firstly, figure 5 indicates that with both algorithms, subjects rely on the ranking to inform their choices - the first ranked lottery is the modal choice with either algorithm. However, with the expected utility algorithm, subjects picked the first ranked lottery more than 40% of the time, compared to only $\approx 20\%$ of the time with the platform algorithm. This indicates that subjects adapted their choice behavior to the experimental condition.

In theory, a major benefit of the expected utility algorithm is that subjects can learn to anticipate that their preferred choice is ranked highly and use this to make decisions quicker or with fewer evaluations. Figure 6 shows that this did not happen in the experiment. Subjects spent on average 17.9 seconds when choosing with the expected utility algorithm and 18.1 seconds when choosing with the platform algorithm. Subjects therefore do not seem to use the expected utility algorithm as a tool to make faster decisions. For the optimal willingness-to-pay this implies, that unless the two choice environments create different cognitive costs that do not correlate with decision time, it should be zero for the average subject. It is worth noting, that decision time spikes upon initial exposure to an algorithm and then decreases until it stabilizes between 15 and 20 seconds per round. This effect is more pronounced at the very start, likely because the subjects are still familiarizing themselves with the choice environment. The spike indicates that subjects recognize that

the algorithm has changed and try to adapt accordingly.

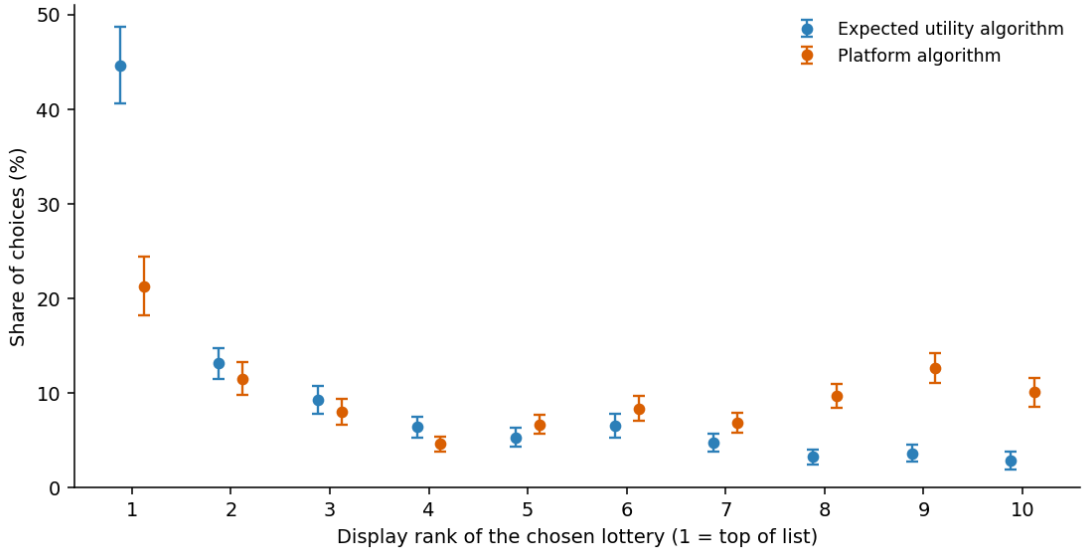


Figure 5: Share of choices with each rank by algorithm

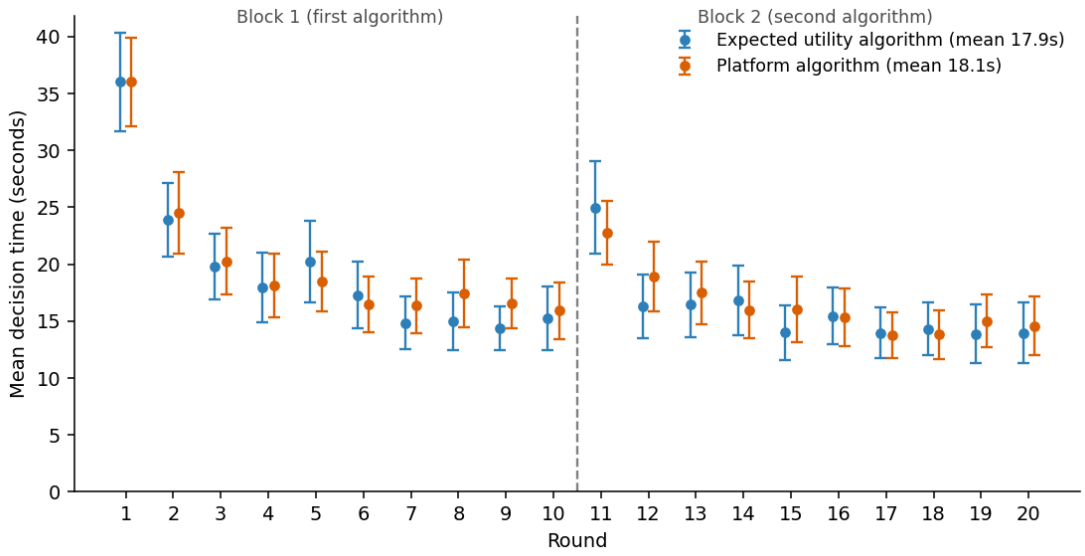


Figure 6: Average time spent per round per algorithm

To examine differences in outcomes under the two treatments, table 3 reports estimates of:

$$y_{ir} = \alpha + \beta_1(CE-5) + \beta_2(CE-5)^2 + \mathbb{1}(A_{ir} = \text{Platform})[\theta_0 + \theta_1(CE-5) + \theta_2(CE-5)^2] + \varepsilon$$

where $A_{ir} \in \{\text{Platform, Expected Utility}\}$ is the algorithm subject i faces in round r, and the outcome y_{ir} is the expected value (columns (1)–(2)) or the certainty equivalent (columns (3)–(4)) of the chosen lottery. Columns (2) and (4) add lottery-set fixed effects and standard errors are clustered at the participant level.

The regression results show that the platform algorithm led to worse outcomes for subjects on both measures. When faced with the platform algorithm, subjects chose lotteries with a \$0.70 lower expected value and a \$0.57 lower certainty equivalent. The risk preference ($(CE-5)$) estimates also reveals that risk seeking subject choose higher expected value lotteries than risk averse subjects in the expected utility treatment rounds, however this difference is erased when they are faced with the platform algorithm. Regarding the results for the regressions with the certainty equivalent of the chosen lotteries as the outcome (columns (3) and (4)), we see that the certainty equivalent decreases by less than the expected value due to the platform algorithm. The platform algorithm was therefore effective in making a tradeoff between reducing the expected value of the chosen lotteries while satisfying the subjects' risk preference. In line with the hypotheses, we also see that interaction Platform $\times (CE-5)^2$ is positive, indicating that subjects with more extreme risk preferences suffer less than risk neutral or close to risk-neutral subjects under the platform algorithm. Accounting for lottery-set fixed effects does not qualitatively change the results.

Taken together, the results of this section establish that the treatment worked as designed. Subjects relied on rankings under both algorithms but adapted their choice behavior: they took the top-ranked lottery twice as often under the expected utility algorithm (44.6% against 21.3%), reached roughly two positions further down the list under the platform algorithm, and received worse outcomes under the platform algorithm — a loss of \$0.70 of expected value and \$0.57 of certainty-equivalent welfare per round. The losses in terms of certainty equivalent are concentrated on subjects near risk neutrality and largely offset by

	Expected value ($)		Certainty equivalent ($)	
	(1)	(2)	(3)	(4)
Platform	-0.702***	-0.711***	-0.565***	-0.577***
	(0.031)	(0.030)	(0.030)	(0.029)
Platform $\times(CE-5)$	-0.177***	-0.175***	-0.072***	-0.072***
	(0.014)	(0.014)	(0.011)	(0.011)
Platform $\times(CE-5)^2$	0.006	0.006	0.053***	0.054***
	(0.006)	(0.006)	(0.007)	(0.007)
$(CE-5)$	0.199***	0.115***	0.533***	0.458***
	(0.017)	(0.021)	(0.012)	(0.021)
$(CE-5)^2$	-0.004	-0.006	0.032***	0.030***
	(0.007)	(0.006)	(0.008)	(0.009)
Constant	6.007***	5.751***	5.979***	5.782***
	(0.035)	(0.039)	(0.032)	(0.041)
Lottery-set FE	No	Yes	No	Yes
Observations	5680	5680	5680	5680
R^2	0.360	0.453	0.737	0.780

DV in columns (1)–(2): expected value of the chosen lottery; in columns (3)–(4): its certainty equivalent under the subject's elicited risk preference. Standard errors (in parentheses) clustered at the participant level; round-level observations, rounds 1–20. Lottery-set fixed effects are dictionary × menu cells.

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Chosen-lottery outcomes: expected value and certainty equivalent

the risk-profile compensation at the preference extremes. Consistent with the hypotheses, consumer outcomes are unambiguously worse under the platform algorithm. Despite these differences, subjects spent on average the same time per choice under both algorithms. Subjects adapted *which* option they took, but do not seem to have adapted *how* they evaluated the list.

4.2 Willingness-to-pay

Before turning to the results of the willingness-to-pay elicitation, recall the experimental procedure. The WTP elicitation screen gives subjects a 1$ endowment and a slider with no default. The middle position of the slider means that the algorithm that the subject faces will be randomly drawn with equal probability. Shifting the slider in either direction costs $0.02 per percentage of probability, such that shifting the slider to receive either the platform or the expected utility algorithm (or "algorithm 1" and "algorithm 2" for the participants) with certainty requires the subject to spend their full endowment (see screenshots A16, A17). In the following analysis, I interpret a positive willingness-to-pay for the platform algorithm as "negative" willingness-to-pay, in the sense that subjects are willing to pay to avoid the expected utility algorithm.

In addition, recall the benchmarks derived for the individually rational willingness-to-pay in Section 3. Given the realized value differences between the two algorithms' recommendations and the $1 cost of increasing the probability of getting the expected utility algorithm by 50% (from the costless 50-50 baseline), the payoff-maximizing willingness to pay is zero for the full range of risk preferences. Figure 6 shows that subjects did not use the expected utility algorithm to decide faster, which means that time savings do not overturn this benchmark. Differences in cognitive effort when choosing with the platform algorithm vs. the expected utility algorithm could still play a role, however if these differences were large we would expect there to be differences in decision speed also.

Observed behavior departs from this benchmark. Average willingness to pay for the expected utility algorithm is positive and significant at $0.14 ($t = 4.88$, $p < 0.001$). Figure 7

shows that willingness to pay is indistinguishable from zero for strongly risk-averse subjects ($0.01 for $CE \leq 3$) and increases for close to risk-neutral and risk-neutral subjects. The bins $(0, 3]; (3, 5); 5; (5, 7]$ are consistent with the hump shape of the value difference graph (figure 2). However, the $(7, 10)$ bin is an outlier with the highest average willingness-to-pay, close to $0.4. This is at odds with the hump shape of the value difference graph. While this is the smallest bin in terms of the number of subjects, and the noisiest estimate, this may be consistent with a setting where subjects use heuristics to make their choices - table 3 shows that the expected value of chosen lotteries decreases more for risk-seeking subjects when they are faced with the platform algorithm.

At the individual level, we find however that willingness-to-pay is largely unresponsive to value: subjects' realized gains from the expected utility algorithm, their decision-time differences, and the presentation order of the algorithms jointly explain almost none of the variation in willingness to pay ($F(5, 278) = 1.76$, $p = 0.12$; Appendix Table A2).

Together, these findings suggest that subjects price the *experience* of choosing from the two lists rather than the monetary stakes or time savings. For subjects near risk neutrality, the platform algorithm's list is least helpful and choosing from it may carry a cognitive cost that leaves no trace in decision time. The fact that other measured differences do not seem to explain willingness-to-pay suggests that the subjects' valuations reflect a coarse judgment of the two algorithms rather than a calibrated assessment of their value.

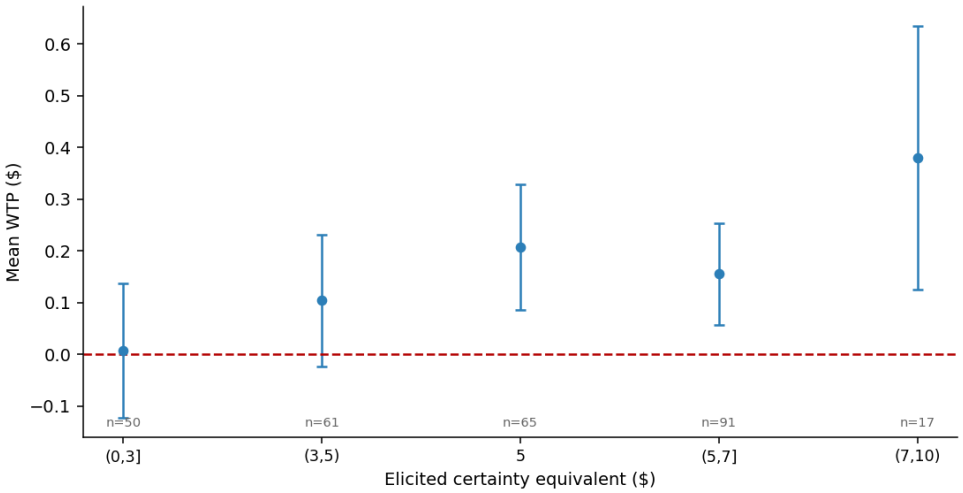


Figure 7: Distribution of willingness-to-pay over *elicited* certainty equivalent in the risk preference elicitation task.

5 Conclusion

In this paper, I develop an online experiment in order to study the effect of recommendation algorithms on consumer choices. The experiment consists of two parts: a first, standard risk elicitation task, and a choice task where the subjects are asked to choose from lists of lotteries, where the lists have been selected and ordered by the different recommendation algorithms. This way of designing the experiment has the important advantage that no prior information about the characteristics of the subjects, or any previous choice data is necessary in order to inform the recommendation algorithms. The recommendation algorithms can be simply constructed using the elicited risk preference. I then use this experimental framework to study the question of whether subjects have a positive willingness to pay for a better recommendation algorithm. However, this experimental framework can in principle be applied to study many other questions related to recommendation systems, including the role of complexity, what happens when information about the algorithm is disclosed, and how subjects search when they are faced with recommendation systems.

Our first set of results concerns how consumers behave under the two recommendation algorithms. Subject's adapted their choice behavior to the algorithms, but suffered from

worse outcomes when faced with the platform algorithm. Against expectation, the time spent per choice is the same under both algorithms, despite the fact that the expected utility algorithm should always rank the best option for the subject first.

Regarding the willingness to pay for the better (expected utility) algorithm, I find that it is positive on average. Risk-neutral subjects that derive a higher benefit from the expected utility algorithm have a higher willingness to pay for it - however, subjects valuations are not responsive to differences in most other measures of benefits that they could derive from the expected utility algorithm. Overall, the results from this experiment suggest, that from a regulatory perspective, giving consumers the option between different recommendation systems may not be ideal. However since willingness-to-pay is positive on average, it could be viable for platform businesses to offer better recommendations for a fee, or to offer the choice of different recommendation systems in order to differentiate themselves from competitors[1]. Future studies could focus on the role of information that subjects have when choosing between algorithms. In this experiment, the only information that helps subjects in their choice comes from their experience with the two algorithms. It would be interesting to understand how choices vary with additional information about the algorithms, and whether subjects are willing to incur costs to learn this information.

[1]The platform Bluesky is essentially offering different recommendation systems by allowing users to subscribe to different "feeds".

References

Armstrong, Mark and Jidong Zhou, "Paying for Prominence," *The Economic Journal*, November 2011, *121* (556), F368–F395.

_ , John Vickers, and Jidong Zhou, "Prominence and consumer search," *The RAND Journal of Economics*, 2009, *40* (2), 209–233.

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy, "Incentives in Experiments: A Theoretical Analysis," *Journal of Political Economy*, August 2018, *126* (4), 1472–1503.

Bergemann, Dirk and Alessandro Bonatti, "Data, Competition, and Digital Platforms," April 2023.

_ , Jacques Cremer, David Dinielli, Carl-Christian Groh, Paul Heidhues, Maximilian Schafer, Monika Schnitzer, Fiona M. Scott Morton, Katja Seim, and Michael Sullivan, "Market Design for Personal Data," *Yale Journal on Regulation*, 2023, *40*, 1056.

Boldrini, Michela and Francesco Clavorà Braulin, "Taming Tech Giants's Algorithms: An analysis of consumer's choices on Amazon.pdf," November 2025.

Bourreau, Marc and Germain Gaudin, "Streaming platform and strategic recommendation bias," *Journal of Economics & Management Strategy*, 2022, *31* (1), 25–47.

Caplin, Andrew, Mark Dean, and Daniel Martin, "Search and Satisficing," *American Economic Review*, December 2011, *101* (7), 2899–2922.

Chen, Jiawei, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang, "AutoDebias: Learning to Debias for Recommendation," in "Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval" SIGIR '21 Association for Computing Machinery New York, NY, USA July 2021, pp. 21–30.

Chen, Yiting, Ziye Wu, and Songfa Zhong, "Risk-Taking Behavior under AI-Based Recommendation," April 2023.

de Cornière, Alexandre and Greg Taylor, "A model of biased intermediation," *The RAND Journal of Economics*, 2019, *50* (4), 854–882.

Derakhshan, Mahsa, Negin Golrezaei, Vahideh Manshadi, and Vahab Mirrokni, "Product Ranking on Online Platforms," *Management Science*, June 2022, *68* (6), 4024–4041.

Digital Markets Act

***Digital Markets Act**, September 2022.*

Donelly, Robert, Ayush Kanodia, and Ilya Morozov, *"Welfare Effects of Personalized Rankings," mimeo, March 2023.*

Düll, Adrian, Heiko Karle, Simon Martin, and Heiner Schumacher, *"Evaluating Search Cost Models: Estimation and Prediction," 2025.*

Esber, Jad, Scott Duke Kominers, Leora Kornfeld, Scott Belsky, Apurva Chitnis, K. Chmielinski, John Deighton, David Eliot, Carolina Almeida Antunes Rossini, Doc Searls, Kinjal Shah, and Reid Williams, *"The Next Phase of the Data Economy: Economic & Technological Perspectives," November 2024.*

Farronato, Chiara, Andrey Fradkin, and Alexander MacKay, *"Self-Preferencing at Amazon: Evidence from Search Rankings," AEA Papers and Proceedings, May 2023, 113, 239–243.*

Fong, Jessica, Olivia R. Natan, and Ranmit Pantle, *"Consumer Inferences from Product Rankings: The Role of Beliefs in Search Behavior," September 2024.*

Fukuyama, Francis, Barak Richman, Ashish Goel, Roberta R Katz, A Douglas Melamed, and Marietje Schaake, *"Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy," 2020.*

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li, *"Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue," Management Science, July 2014, 60 (7), 1632–1654.*

Greminger, Rafael P., *"Optimal Search and Discovery," Management Science, May 2022, 68 (5), 3904–3924.*

_ , *"Heterogeneous Position Effects and the Power of Rankings," December 2023.*

Hagiu, Andrei and Bruno Jullien, *"Why do intermediaries divert search?," The RAND Journal of Economics, 2011, 42 (2), 337–362.*

_ **and** _ , *"Search diversion and platform competition," International Journal of Industrial Organization, March 2014, 33, 48–60.*

Heidhues, Paul, Mats Köster, and Botond Kőszegi, *"Steering Fallible Consumers," The Economic Journal, May 2023, 133 (652), 1430–1465.*

Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie van Dolder, and Peter P. Wakker, *"Prince: An improved method for measuring incentivized preferences," Journal of Risk and Uncertainty, February 2021, 62 (1), 1–28.*

Kaye, Aaron, *"The Personalization Paradox: Welfare Effects of Personalized Recommendations in Two-Sided Digital Markets," January 2024.*

Lam, Tai H., *"Platform search design and market power," Job Market Paper, 2021, Northwestern University.*

Lee, Kwok Hao and Leon Musolff, *"Entry Into Two-Sided Markets Shaped By Platform-Guided Search," September 2021.*

Lin, Zinan, Dugang Liu, Weike Pan, and Zhong Ming, *"Transfer Learning in Collaborative Recommendation for Bias Reduction," in "Proceedings of the 15th ACM Conference on Recommender Systems" RecSys '21 Association for Computing Machinery New York, NY, USA September 2021, pp. 736–740.*

los Santos, Babur De and Sergei Koulayev, *"Optimizing Click-Through in Online Rankings with Endogenous Search Refinement," Marketing Science, July 2017, 36 (4), 542–564.*

Natan, Olivia R., *"Choice Frictions in Large Assortments," October 2024.*

Peitz, Martin, *"How to Apply the Self-Preferencing Prohibition in the DMA," Journal of European Competition Law & Practice, July 2023, 14 (5), 310–315.*

Plott, Charles R. and Kathryn Zeiler, *"The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations," American Economic Review, June 2005, 95 (3), 530–545.*

Posner, Eric A. and E. Glen Weyl, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society, Princeton University Press, May 2018.*

Reimers, Imke and Joel Waldfogel, *"A Framework for Detection, Measurement, and Welfare Analysis of Platform Bias," October 2023. Series: Working Paper Series.*

Resnick, Paul and Hal R Varian, *"Recommender systems," COMMUNICATIONS OF THE ACM, 1997, 40 (3).*

Ursu, Raluca M., *"The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," Marketing Science, August 2018, 37 (4), 530–552.*

Zhang, Xiaochen, Pedro Ferreira, Miguel Godinho de Matos, and Rodrigo Belo, *"Welfare Properties of Profit Maximizing Recommender Systems: Theory and Results from a Randomized Experiment," Management Information Systems Quarterly, March 2021, 45 (1), 1–45.*

A.1 Experimental Design

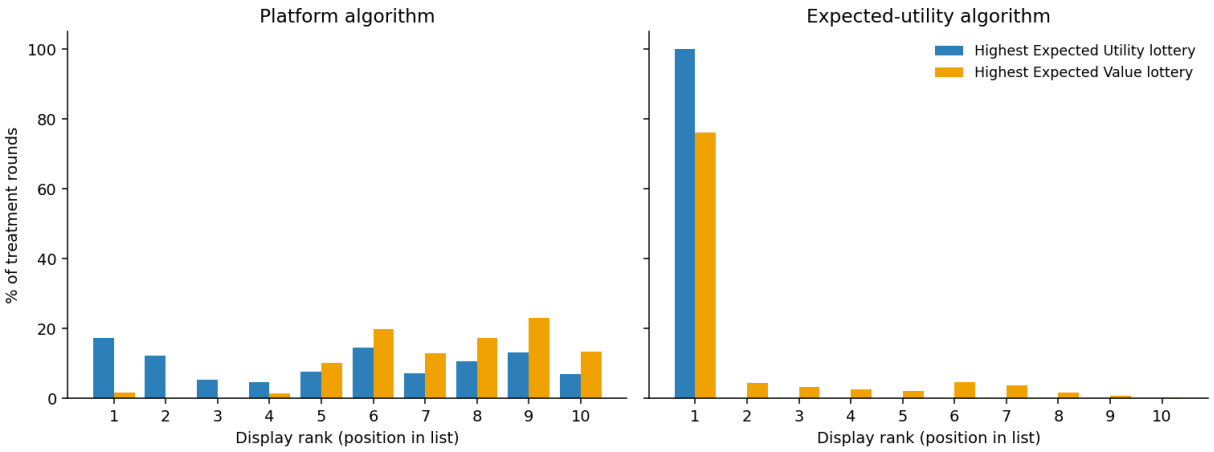


Figure A1: Difference in rank of the highest expected utility/highest expected value lottery for the steering vs. expected utility algorithm

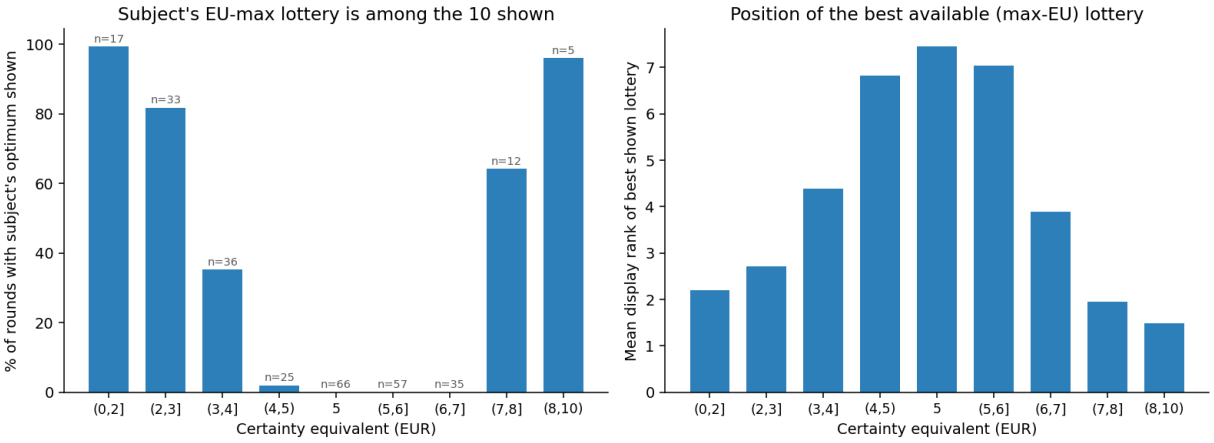


Figure A2: Steering algorithm - availability of best lottery in set (left panel) and average rank of best lottery (right panel) by *elicited* certainty equivalent

variable	mean
N	284.000
Age (mean)	36.310
Female	0.481
Male	0.516
White	0.523
Black	0.249
Mixed	0.121
Asian	0.078
Student	0.216
Employed full-time	0.485
U.S. resident	1.000

Table A1: Demographic summary (final sample; shares of non-missing responses).

A.2 Additional Results

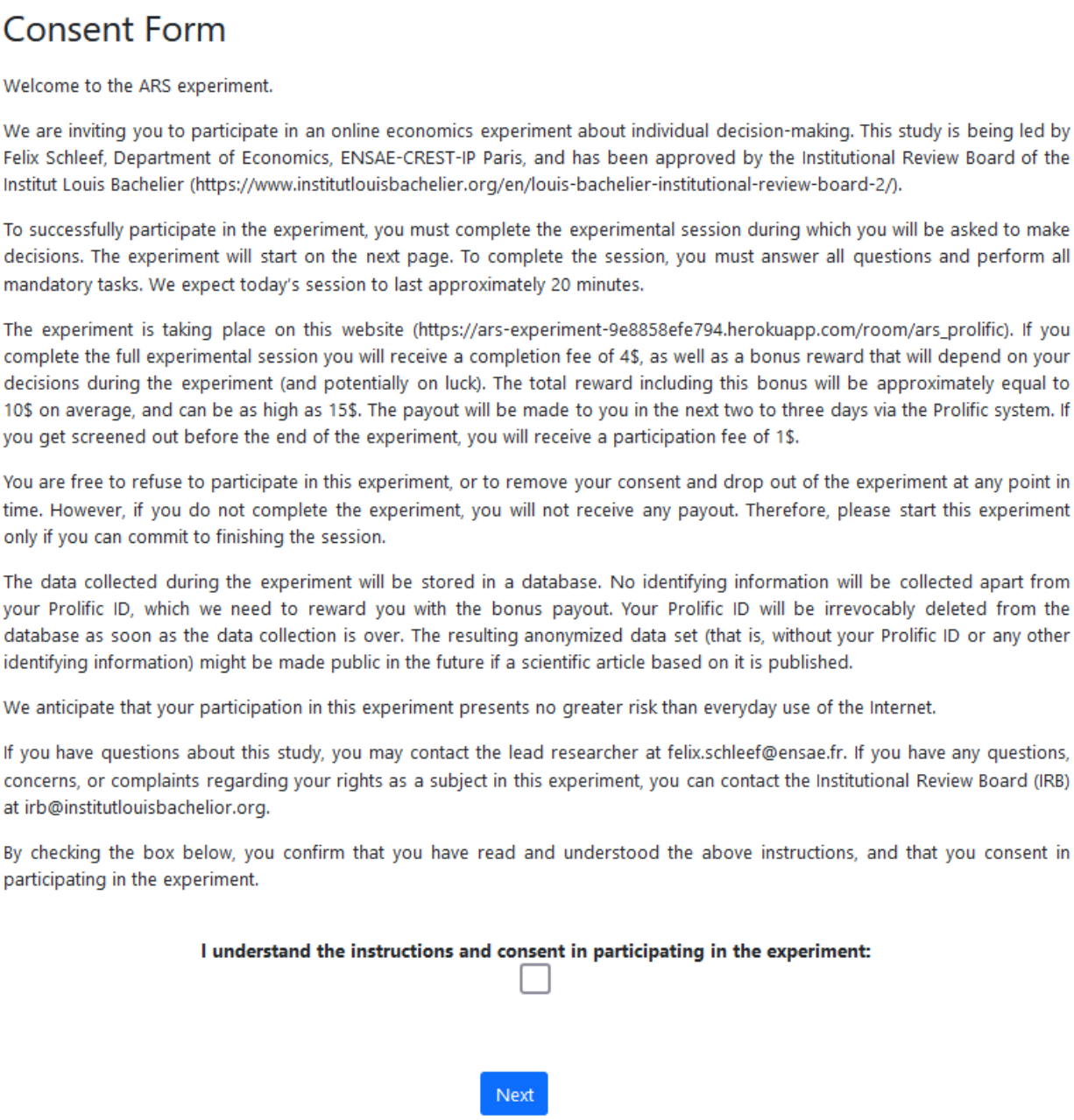
	(1)	(2)	(3)
$(CE-5)$	0.030	0.039*	0.043**
	(0.024)	(0.021)	(0.020)
$(CE-5)^2$	0.004	0.007	0.003
	(0.010)	(0.011)	(0.010)
Δ EV per choice (\$)	0.082		
	(0.073)		
Δ CE per choice (\$)		0.075	
		(0.083)	
Time value of EU algo (\$/block)	-0.020	-0.019	
	(0.024)	(0.024)	
EU shown as Algorithm 1	0.055	0.056	0.072
	(0.063)	(0.063)	(0.059)
Constant	0.054	0.068	0.103**
	(0.067)	(0.066)	(0.049)
Observations	284	284	284
R^2	0.030	0.028	0.024

Δ EV / Δ CE: subject's mean per-choice gain under the EU algorithm. Time value: individual block decision-time saving valued at the realized hourly pay (\$26.2/h).

*** p<0.01, ** p<0.05, * p<0.1

Table A2: Willingness to pay for the expected utility algorithm

A.3 Experiment Screenshots



Consent Form

Welcome to the ARS experiment.

We are inviting you to participate in an online economics experiment about individual decision-making. This study is being led by Felix Schleef, Department of Economics, ENSAE-CREST-IP Paris, and has been approved by the Institutional Review Board of the Institut Louis Bachelier (https://www.institutlouisbachelier.org/en/louis-bachelier-institutional-review-board-2/).

To successfully participate in the experiment, you must complete the experimental session during which you will be asked to make decisions. The experiment will start on the next page. To complete the session, you must answer all questions and perform all mandatory tasks. We expect today's session to last approximately 20 minutes.

The experiment is taking place on this website (https://ars-experiment-9e8858efe794.herokuapp.com/room/ars_prolific). If you complete the full experimental session you will receive a completion fee of 4$, as well as a bonus reward that will depend on your decisions during the experiment (and potentially on luck). The total reward including this bonus will be approximately equal to 10$ on average, and can be as high as 15$. The payout will be made to you in the next two to three days via the Prolific system. If you get screened out before the end of the experiment, you will receive a participation fee of 1$.

You are free to refuse to participate in this experiment, or to remove your consent and drop out of the experiment at any point in time. However, if you do not complete the experiment, you will not receive any payout. Therefore, please start this experiment only if you can commit to finishing the session.

The data collected during the experiment will be stored in a database. No identifying information will be collected apart from your Prolific ID, which we need to reward you with the bonus payout. Your Prolific ID will be irrevocably deleted from the database as soon as the data collection is over. The resulting anonymized data set (that is, without your Prolific ID or any other identifying information) might be made public in the future if a scientific article based on it is published.

We anticipate that your participation in this experiment presents no greater risk than everyday use of the Internet.

If you have questions about this study, you may contact the lead researcher at felix.schleef@ensae.fr. If you have any questions, concerns, or complaints regarding your rights as a subject in this experiment, you can contact the Institutional Review Board (IRB) at irb@institutlouisbachelior.org.

By checking the box below, you confirm that you have read and understood the above instructions, and that you consent in participating in the experiment.

I understand the instructions and consent in participating in the experiment:

Next

Figure A3: Screenshot - Consent Form

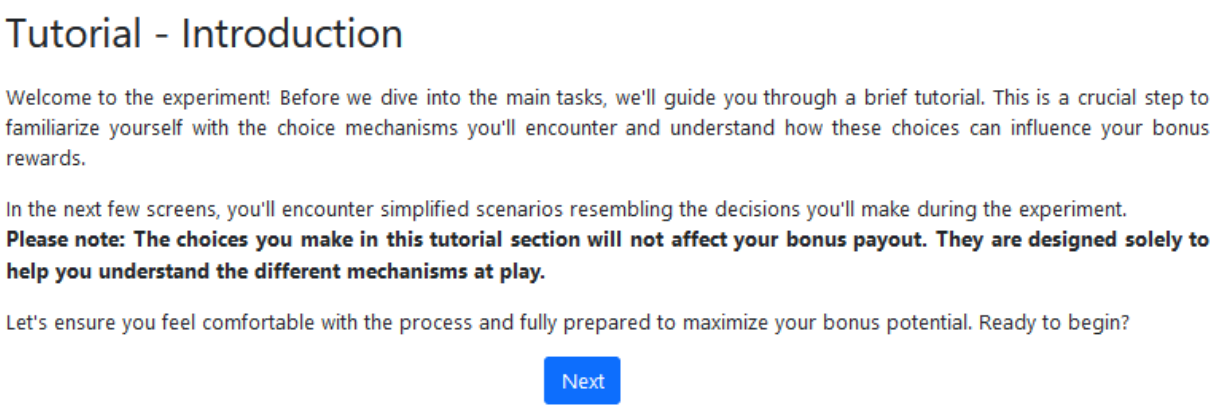


Figure A4: Screenshot - Tutorial Introduction

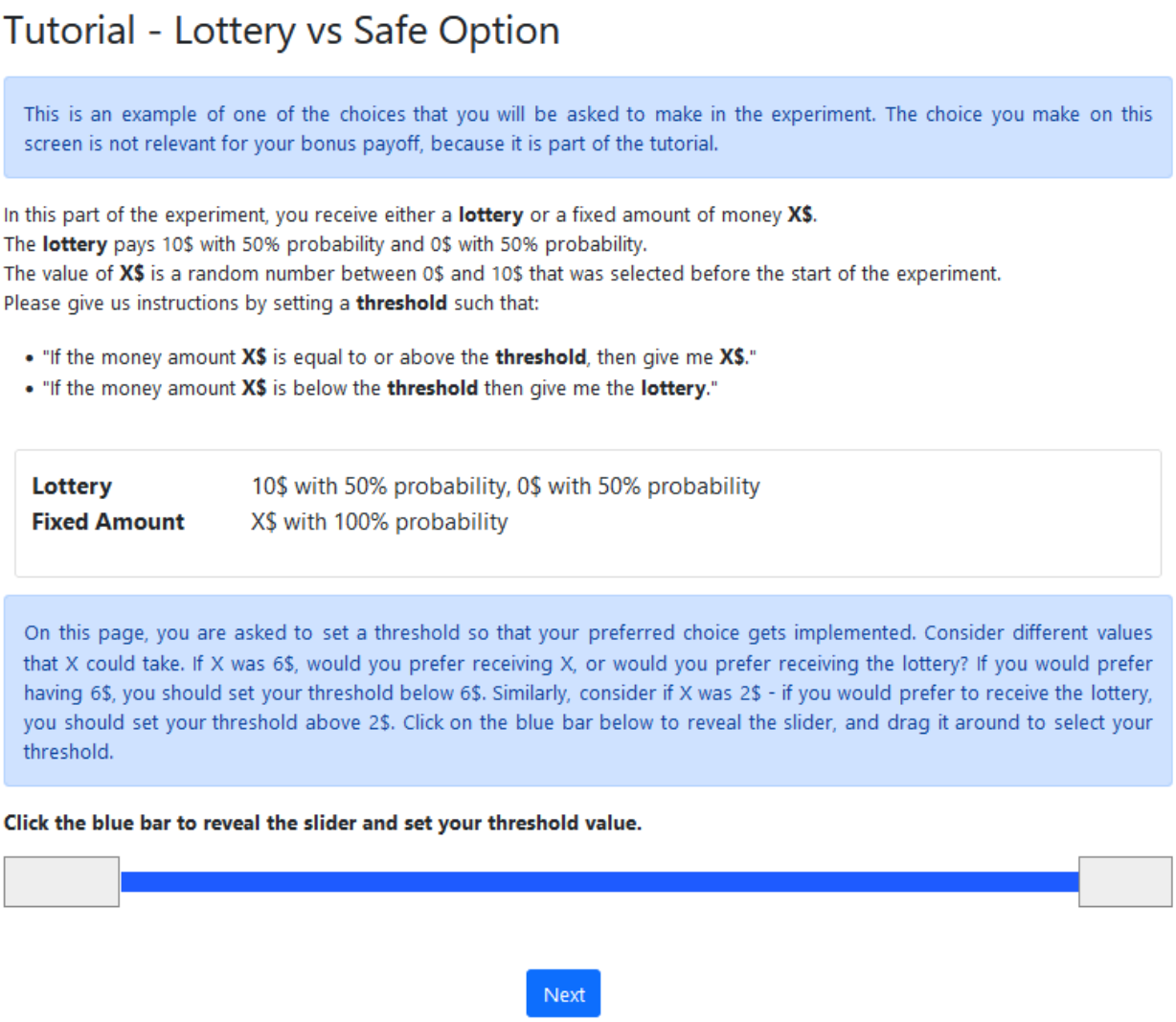


Figure A5: Screenshot - Tutorial - Risk Preference Elicitation

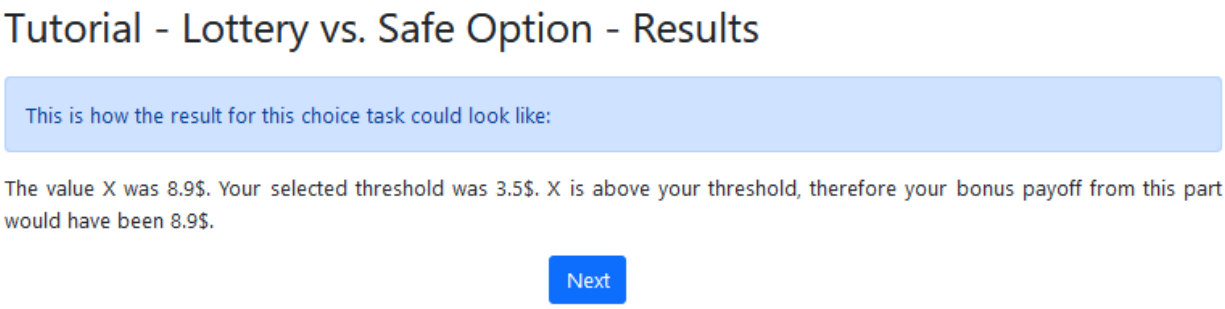


Figure A6: Screenshot - Tutorial - Risk Preference Elicitation Result

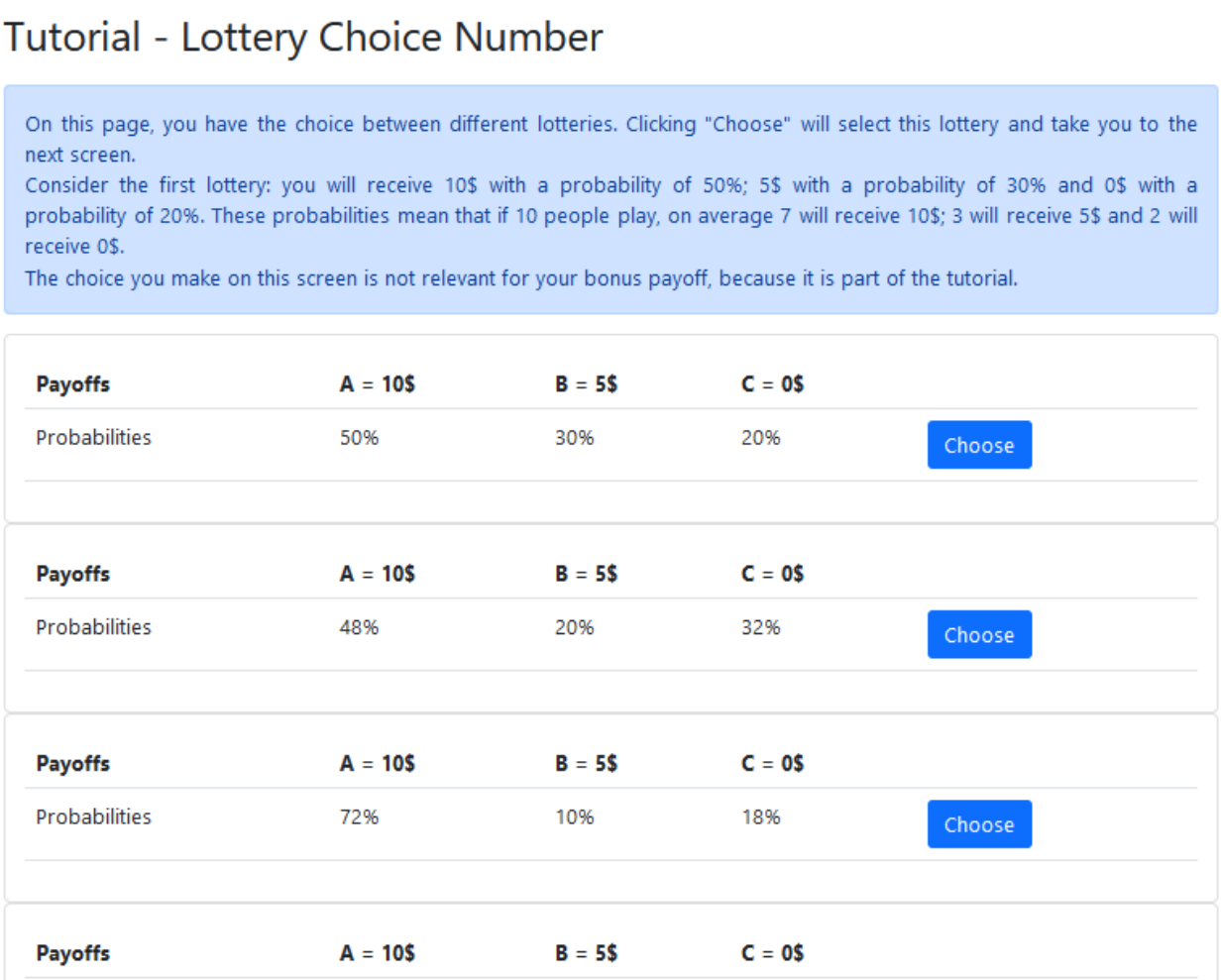


Figure A7: Screenshot - Tutorial - List Choice

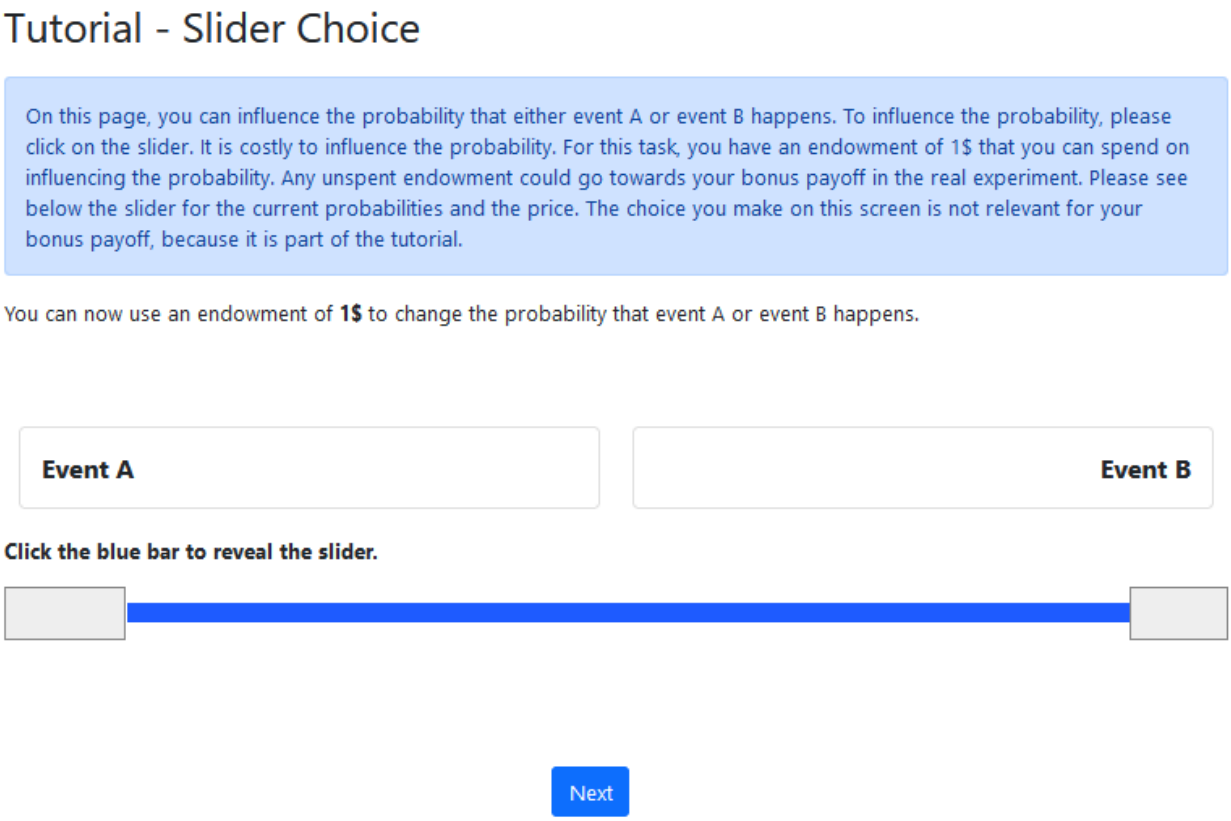


Figure A8: Screenshot - Tutorial - Slider Choice

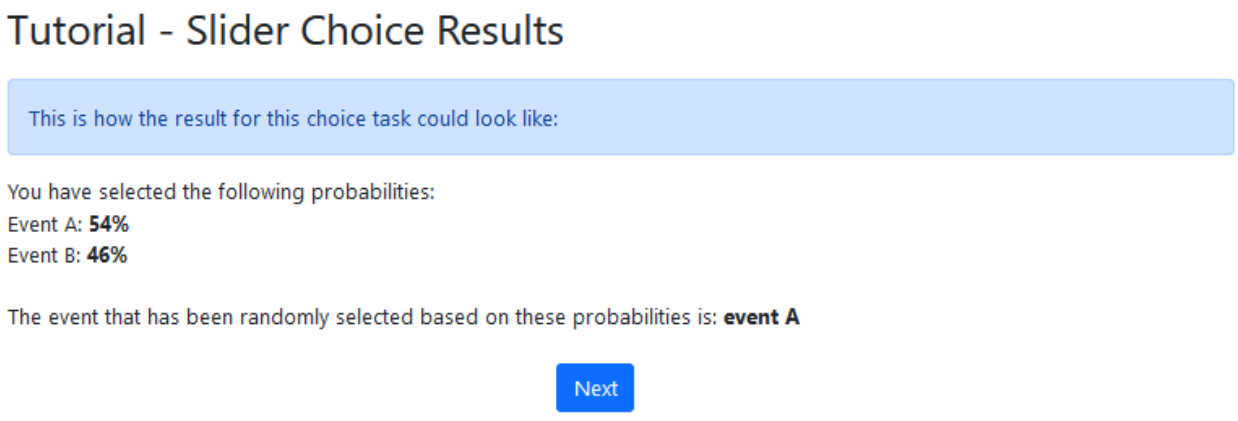


Figure A9: Screenshot - Tutorial - Slider Choice Result

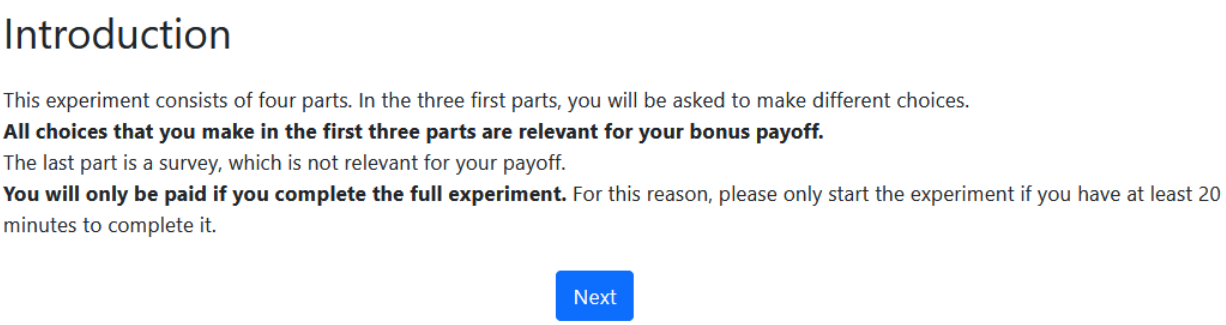


Figure A10: Screenshot - Experiment Introduction

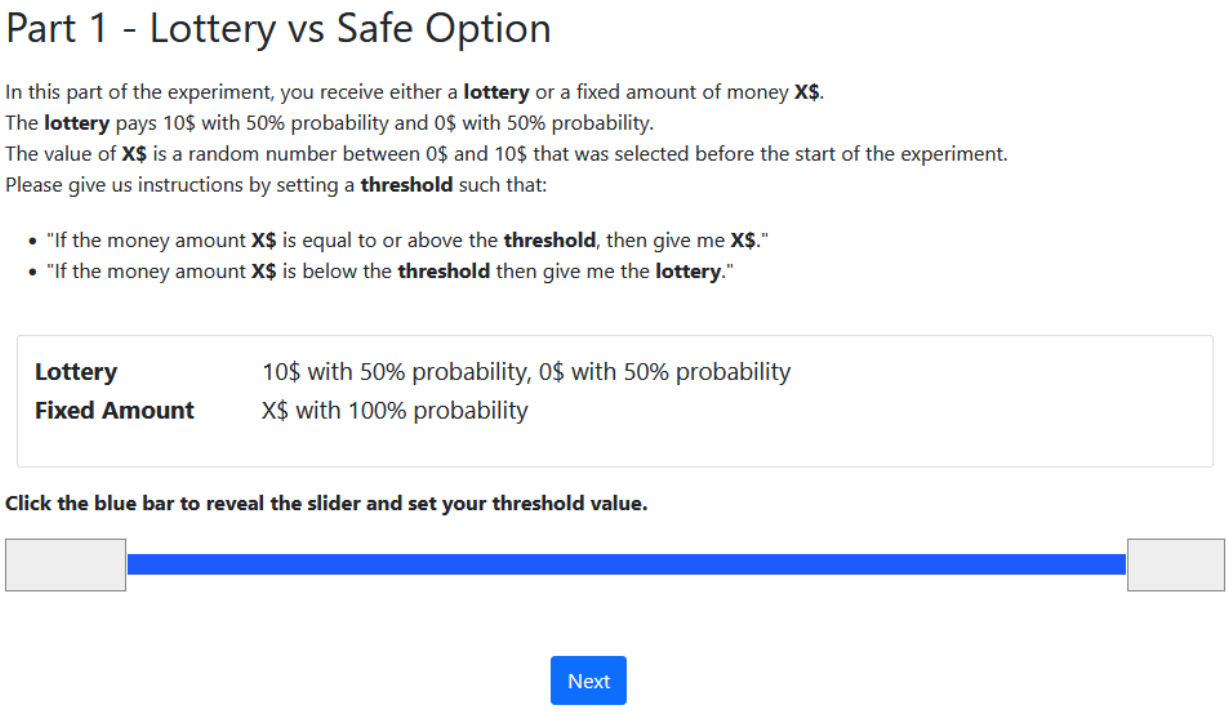


Figure A11: Screenshot - Risk preference elicitation stage

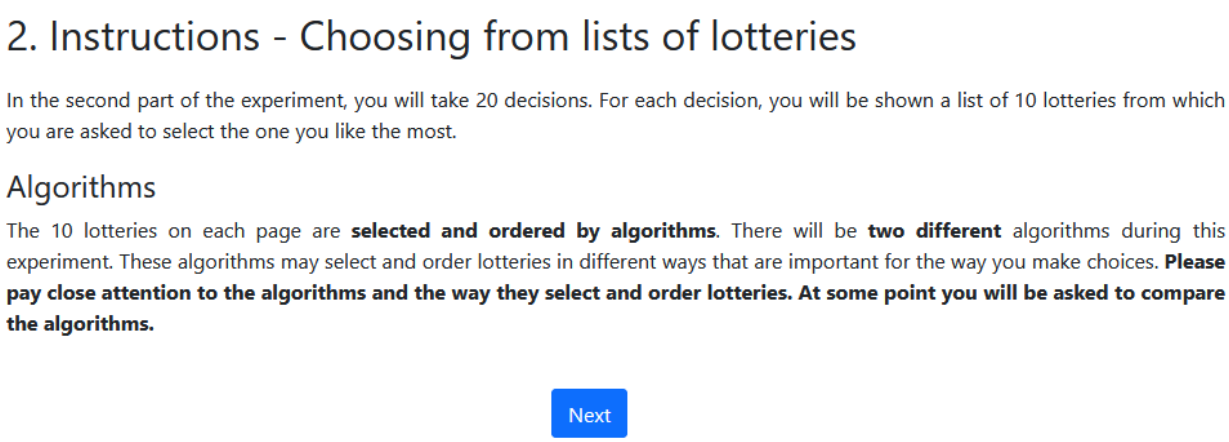


Figure A12: Screenshot - Introduction choice experiment

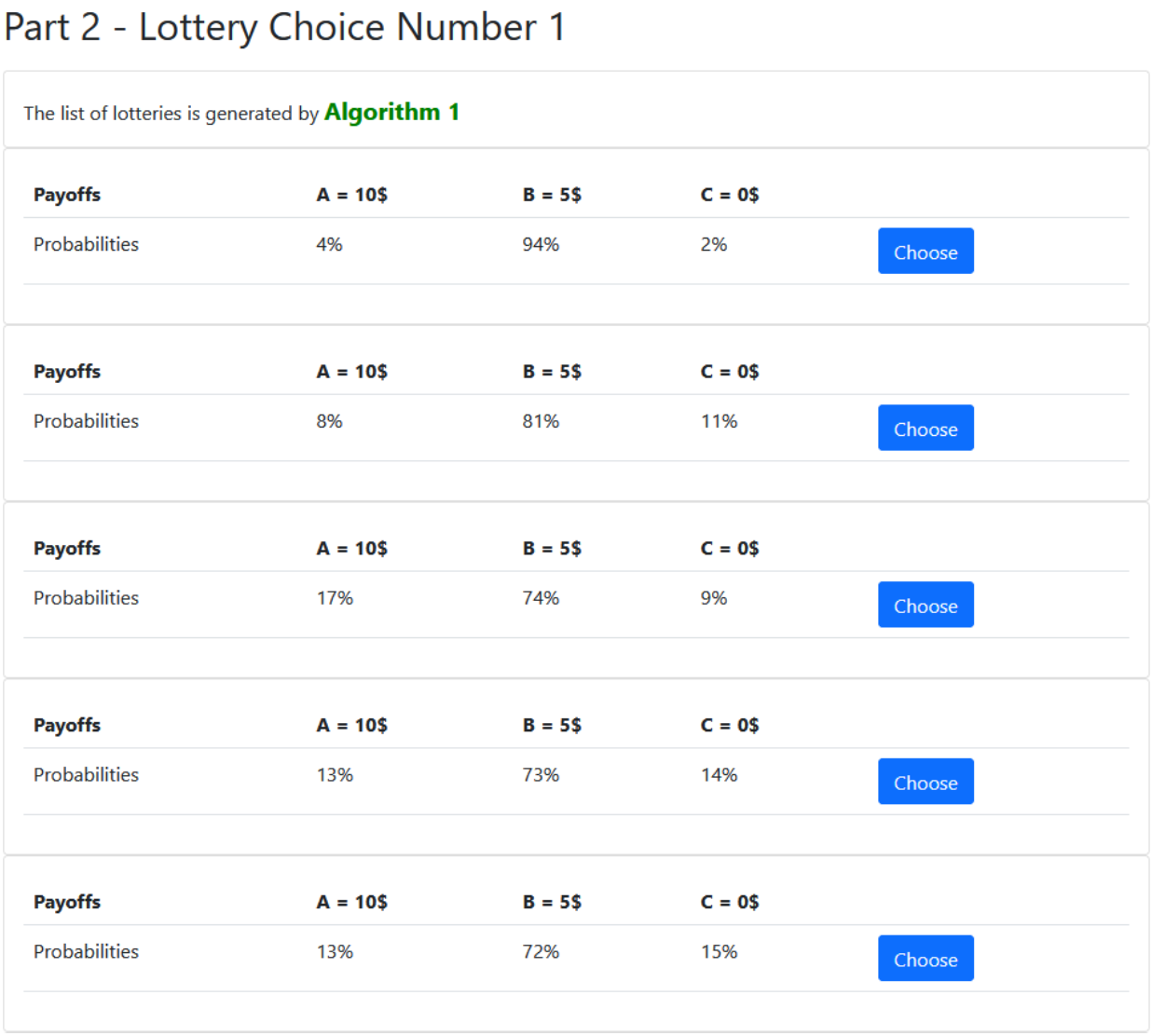


Figure A13: Screenshot - Choice screen where lotteries are selected and ordered by algorithm 1 (Platform algorithm)

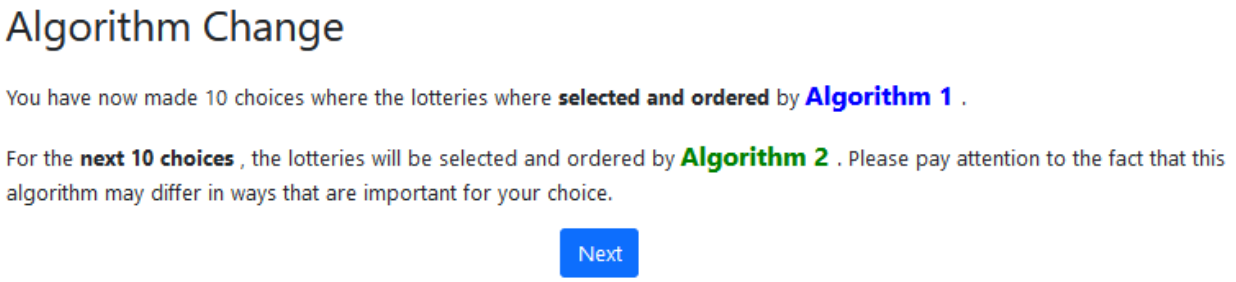


Figure A14: Screenshot - Buffer Screen between rounds 10 and 11

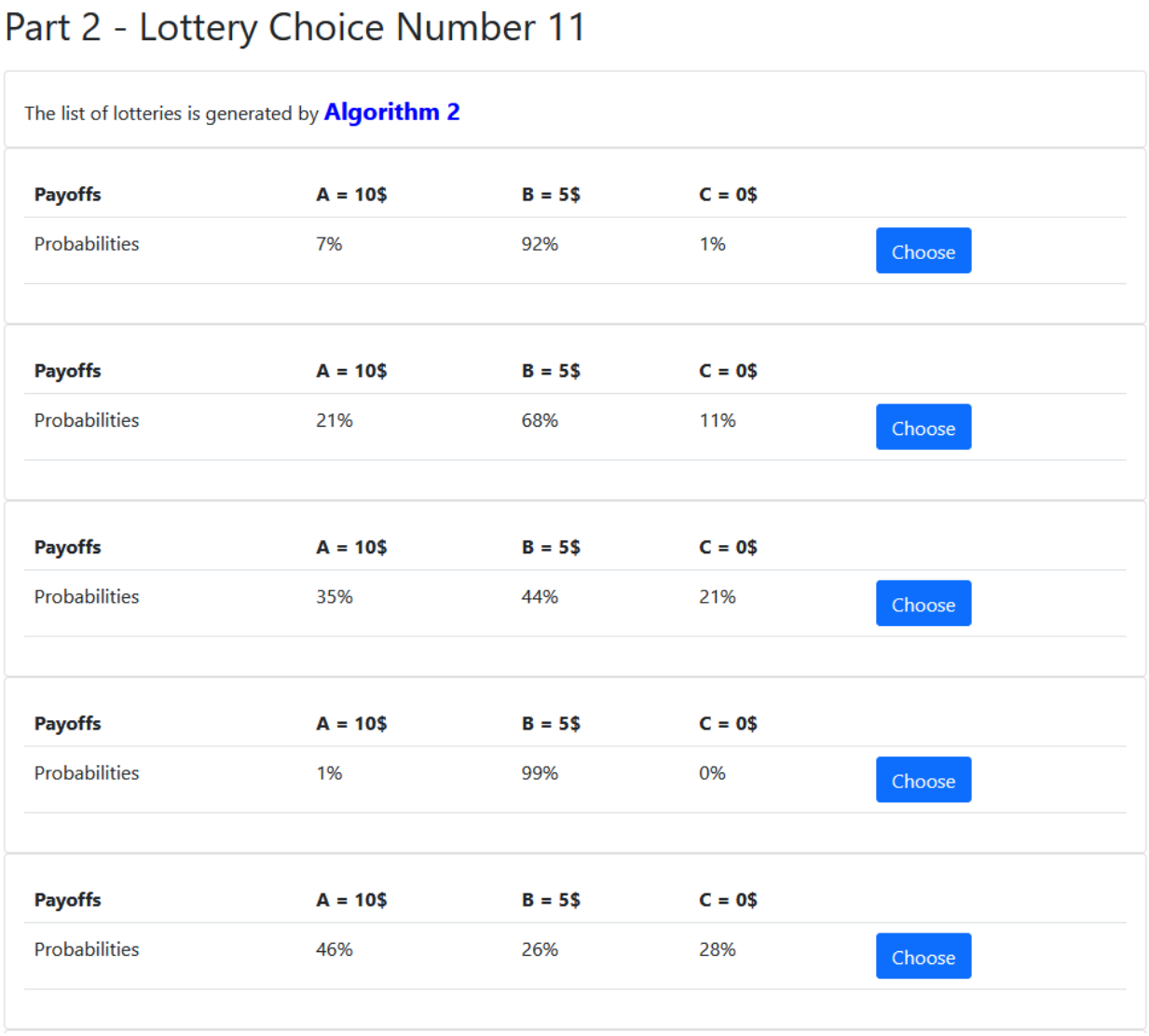


Figure A15: Screenshot - Choice screen where lotteries are selected and ordered by algorithm 2 (Expected utility algorithm)

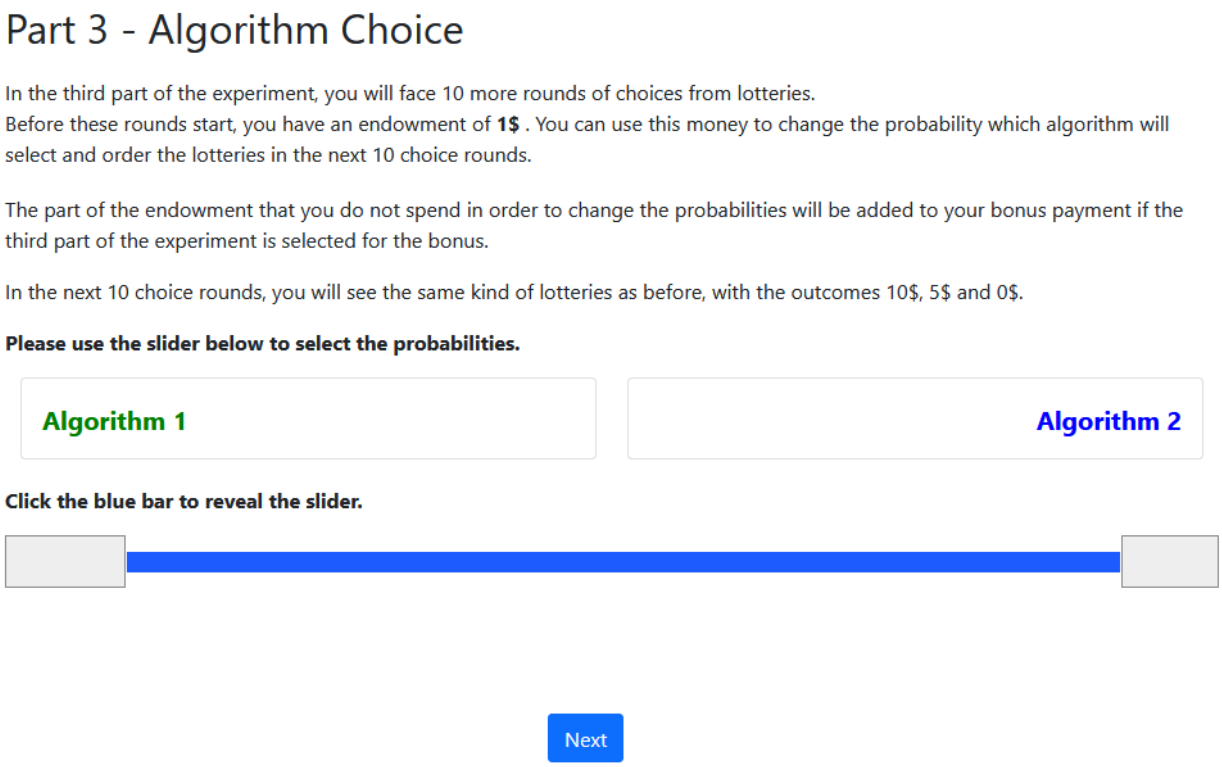


Figure A16: Screenshot - Choice between algorithms

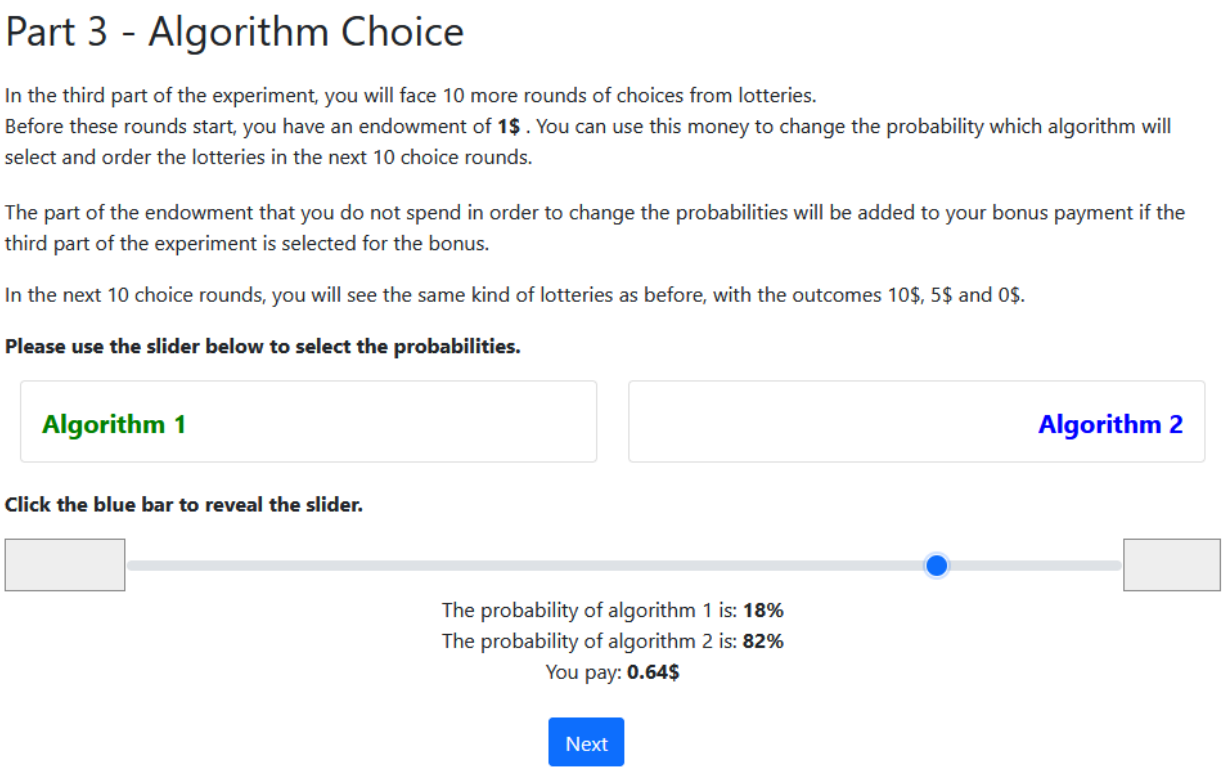


Figure A17: Screenshot - Choice of probability of algorithm 1 or algorithm 2 for the last 10 choice rounds

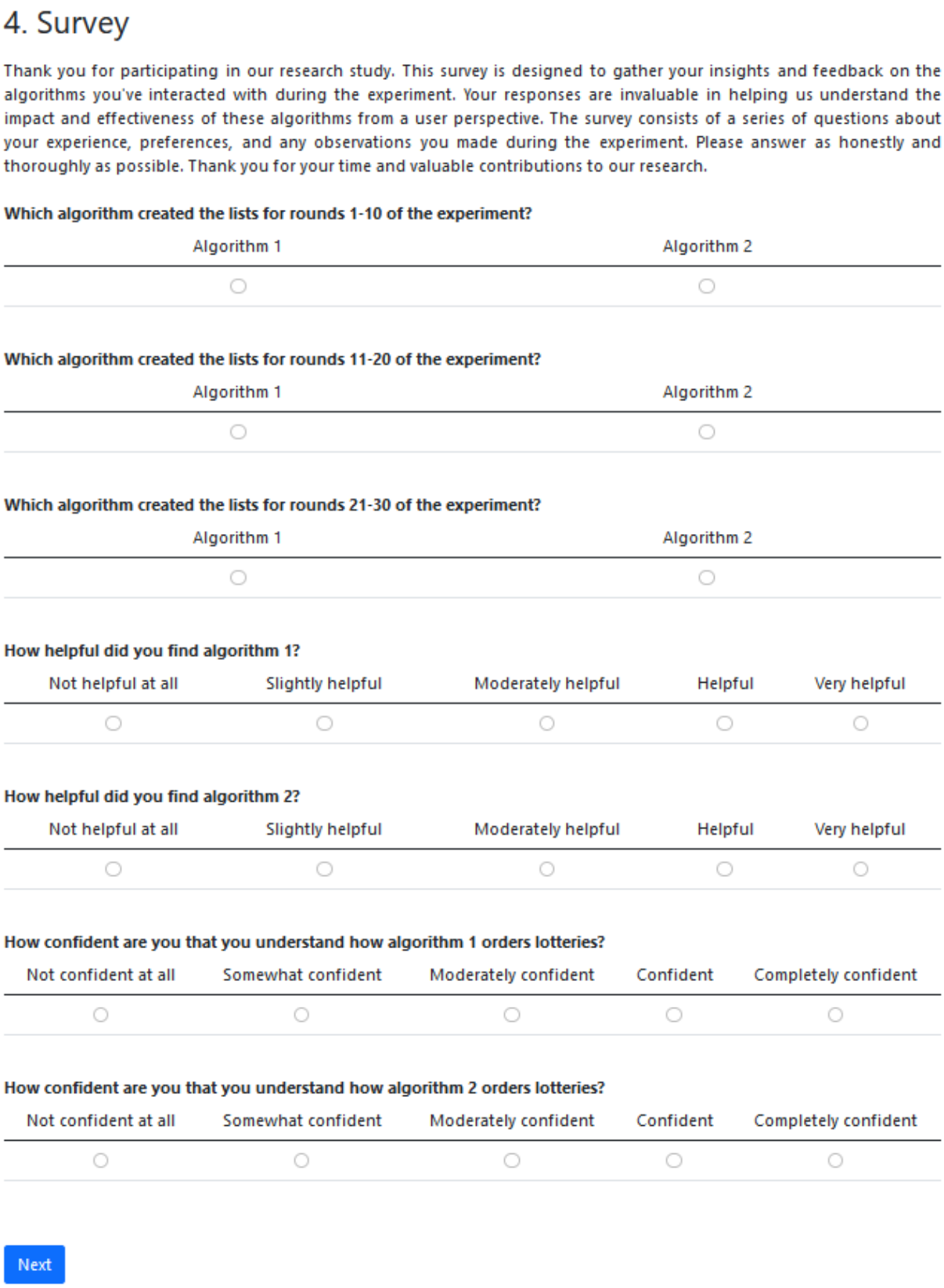


Figure A18: Screenshot - Survey

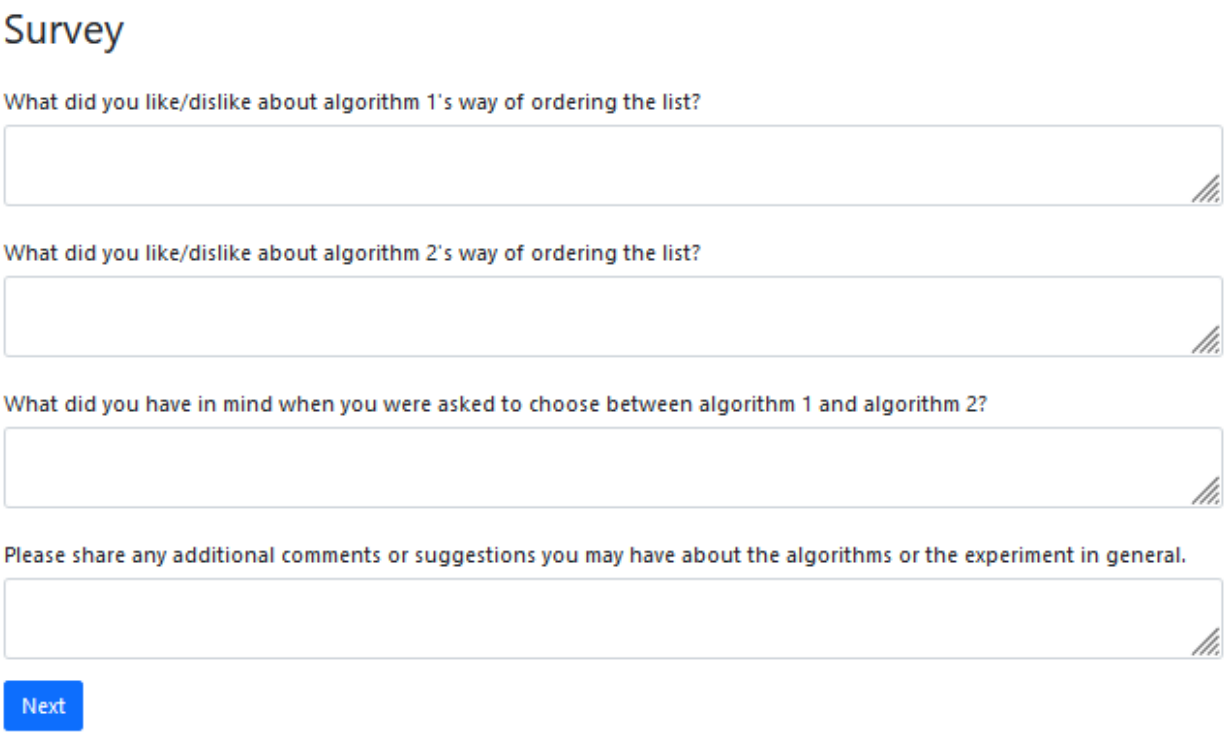


Figure A19: Screenshot - Survey open questions

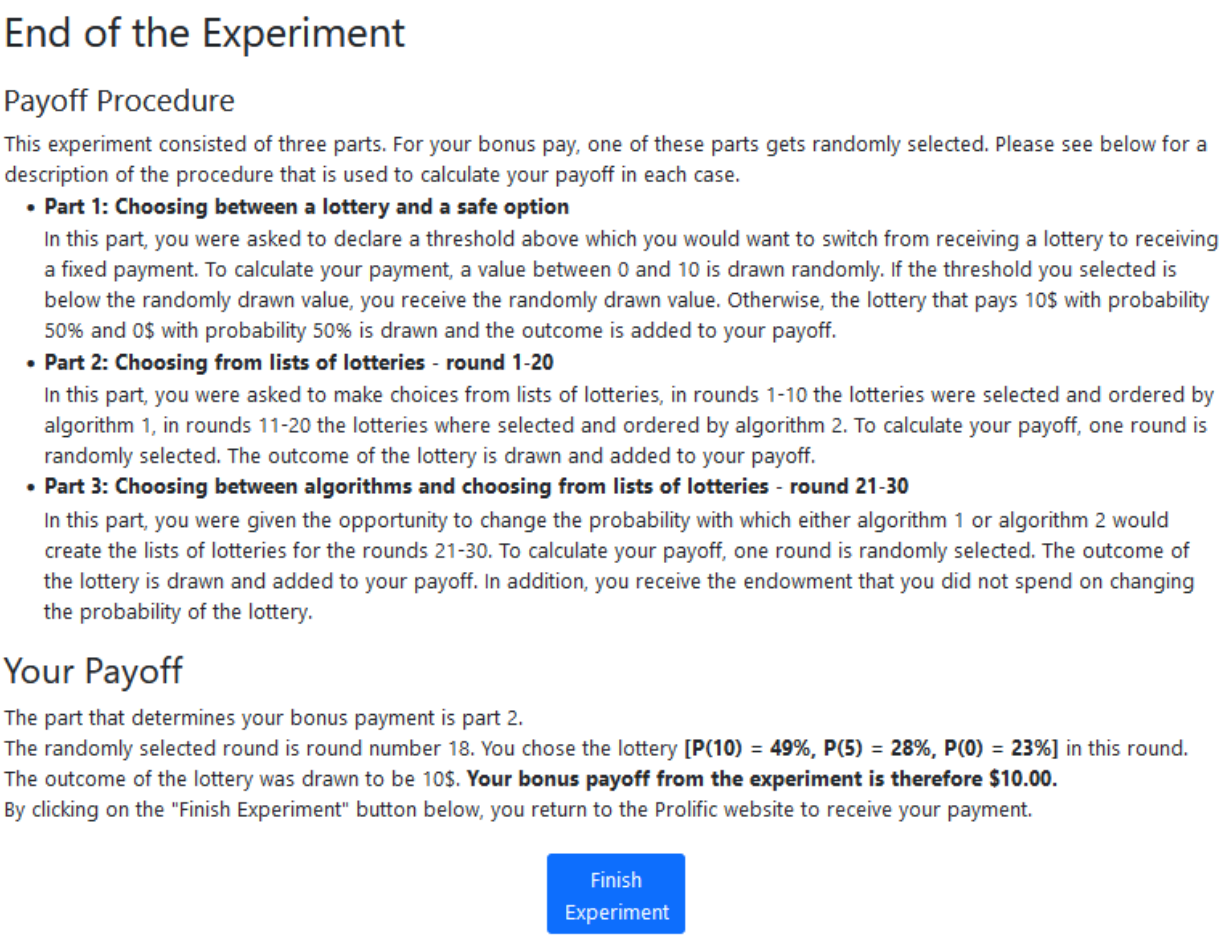


Figure A20: Screenshot - End of the experiment and payment screen

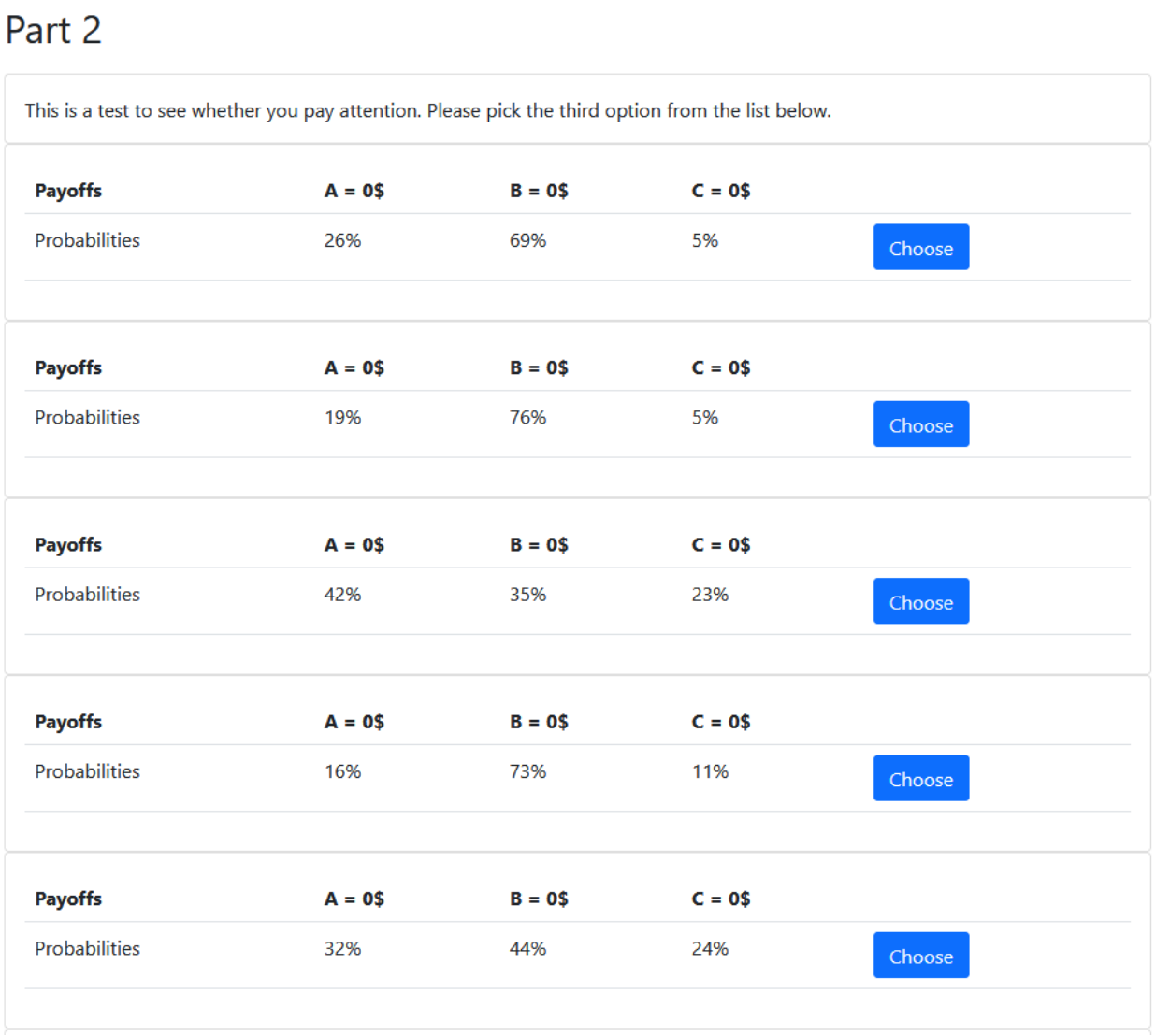


Figure A21: Screenshot - Attention check

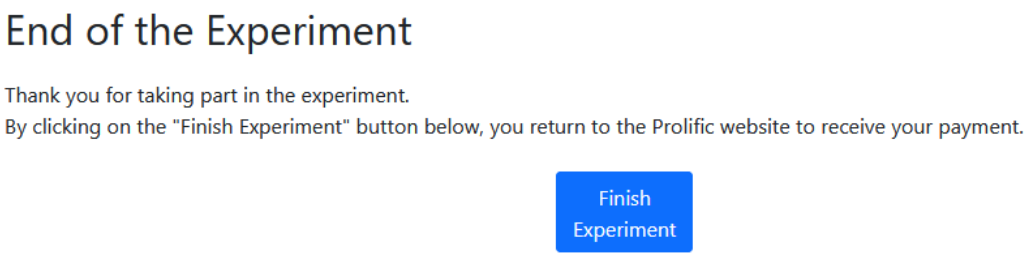


Figure A22: Screenshot - Attention check fail

A.4 Hypotheses

Hypothesis 1: *Subjects choose lotteries with lower expected utility on average when facing the platform algorithm compared to when they face the expected utility algorithm.*

Hypothesis 2: *Subjects choose lotteries with lower expected value on average when facing the platform algorithm compared to when they face the expected utility algorithm.*

Hypothesis 3: *Subjects choose lotteries that are further down the list on average when facing the platform algorithm compared to when they face the expected utility algorithm.*

Hypothesis 4: *Subjects with high search costs are more negatively effected by the platform algorithm than subjects with low search costs.*

Hypothesis 5: *Subjects that are risk neutral or close to risk-neutral are more negatively effected by the platform algorithm than subjects with high risk aversion or high preference for risk.*

Hypothesis 6: *Subjects have a positive willingness-to-pay for the expected utility algorithm.*

Hypothesis 7: *Subjects with high search costs have a higher willingness-to-pay for the expected utility algorithm than subjects with low search costs.*

Hypothesis 8: *Subjects that are risk neutral or close to risk neutral have a higher willingness-to-pay for the expected utility algorithm than subjects with high risk aversion or high preference for risk.*