Bao, Yongping et al.

Working Paper

Similarity and Consistency in Algorithm-Guided Exploration

CESifo Working Paper, No. 10188

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Bao, Yongping et al. (2022) : Similarity and Consistency in Algorithm-Guided Exploration, CESifo Working Paper, No. 10188, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/271832

CESifo WORKING PAPERS

10188
2022

December 2022

Similarity and Consistency in Algorithm-Guided Exploration

Yongping Bao, Ludwig Danwitz, Fabian Dvorak, Sebastian Fehrler, Lars Hornuf, Hsuan Yu Lin, Bettina von Helversen

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10188*

Similarity and Consistency in Algorithm-Guided Exploration

Abstract

Algorithm-based decision support systems play an increasingly important role in decisions involving exploration tasks, such as product searches, portfolio choices, and human resource procurement. These tasks often involve a trade-off between exploration and exploitation, which can be highly dependent on individual preferences. In an online experiment, we study whether the willingness of participants to follow the advice of a reinforcement learning algorithm depends on the fit between their own exploration preferences and the algorithm's advice. We vary the weight that the algorithm places on exploration rather than exploitation, and model the participants' decision-making processes using a learning model comparable to the algorithm's. This allows us to measure the degree to which one's willingness to accept the algorithm's advice depends on the weight it places on exploration and on the similarity between the exploration tendencies of the algorithm and the participant. We find that the algorithm's advice affects and improves participants' choices in all treatments. However, the degree to which participants are willing to follow the advice depends heavily on the algorithm's exploration tendency. Participants are more likely to follow an algorithm that is more exploitative than they are, possibly interpreting the algorithm's relative consistency over time as a signal of expertise. Similarity between human choices and the algorithm's recommendations does not increase humans' willingness to follow the recommendations. Hence, our results suggest that the consistency of an algorithm's recommendations over time is key to inducing people to follow algorithmic advice in exploration tasks.

JEL-Codes: C910, D830.

Keywords: algorithms, decision support systems, recommender systems, advice-taking, multi-armed bandit, search, exploration-exploitation, cognitive modeling.

*Yongping Bao**
University of Bremen / Germany
yongping.bao@uni-bremen.de

Ludwig Danwitz
University of Bremen / Germany
danwitz@uni-bremen.de

Fabian Dvorak
University of Konstanz / Germany
fabian.dvorak@uni-konstanz.de

Sebastian Fehrler
University of Bremen / Germany
sebastian.fehrler@uni-bremen.de

Lars Hornuf
University of Bremen / Germany
hornuf@uni-bremen.de

Hsuan Yu Lin
University of Bremen / Germany
hslin@uni-bremen.de

Bettina von Helversen
University of Bremen / Germany
b.helversen@uni-bremen.de

*corresponding author

December 21, 2022

1 Introduction

Algorithm-based decision support systems are becoming increasingly influential in the decision-making practices of individuals and organizations, including product searches, portfolio choices, and human resource procurement (e.g., Adomavicius and Tuzhilin, 2005b; Panniello et al., 2016; Tauchert and Mesbah, 2019; Zhou et al., 2021; Wang et al., 2022; Uotila et al., 2009). They also affect search and rescue missions and production settings, as well as geological and space exploration, in which teams of humans and algorithm-driven robots jointly perform search tasks (e.g., Orth et al., 2021; Schoonderwoerd et al., 2022; Pouya and Madni, 2021). Such decision problems are typically characterized by sequential choices and learning over time. The key trade-off is between searching longer for the best option, also referred to as *exploration*, or sticking to the option with the highest expected return at an earlier point in time, or *exploitation* (e.g., Addicott et al., 2017). People often struggle with such exploration–exploitation trade-offs due to the complexity of the tasks and constraints on cognitive resources (e.g., Gershman, 2020; Laureiro-Martinez et al., 2019). In contrast, it is widely recognized that reinforcement learning algorithms can perform very well in exploration–exploitation settings (Sutton and Barto, 2018), such as in the canonical multi-armed bandit problem (Thompson, 1933), making these tasks a compelling target for developing decision support systems (e.g., Zhou et al., 2021).

The usefulness and impact of advice from such algorithms crucially depends on the willingness of humans to follow it, and the literature on human–algorithm interactions provides little guidance in this regard. Previous research has primarily focused on one-shot decisions and the results have been fairly inconsistent. While in some judgment tasks, such as shopping online, humans happily follow an algorithm's advice (e.g., Zhou et al., 2021), in others they exhibit considerable reluctance to do so (Logg et al., 2019; Dietvorst et al., 2015). Indeed, recent reviews suggest that whether and when people are willing to take advice is complex, depending on the characteristics of the individual, the algorithm, and the task (Kawaguchi, 2021; Mahmud et al., 2022).

In this study, we investigate whether people accept and benefit from the advice of an algorithm in an exploration–exploitation task, and which characteristics of the algorithm and the human decision-maker influence the willingness to accept the algorithm's recommendation. In these tasks, the human decision-maker and the algorithm interact repeatedly, providing the opportunity to learn about the task but also allowing the human to observe the algorithm's decision behavior and quality of advice. The characteristics of the algorithm and the human decision-maker are therefore likely to jointly determine whether the advice is accepted and how effectively the task is solved. Given that the ability to balance exploration and exploitation is crucial to success in these tasks, we focus on whether the algorithm's tendency to explore and the similarity in tendencies between the algorithm and the human decision-maker will affect the latter's inclination to follow the algorithm's recommendations.

For this purpose, we conducted a tightly controlled online experiment using a stationary multi-armed bandit task—a typical task used to investigate exploration–exploitation trade-offs (Speeken-

brink, 2022). Our analysis compares performance when human participants solve the task on their own versus when they receive advice from a state-of-the-art learning algorithm, which differs across treatments in its tendency to explore or exploit. We also use a cognitive modeling approach to estimate latent characteristics of the human decision-maker and compare these with the characteristics of the algorithm.

Our findings provide strong evidence that in a stationary multi-armed bandit task, participants benefit from the advice of a state-of-the-art learning algorithm independent of its exploration tendency. However, participants are more willing to follow an algorithm that is less explorative than they are. Participants that are more explorative than the algorithm benefit more from the algorithm's advice, which demonstrates the need to take the characteristics of both the human advisees and the algorithm into account when designing algorithm-based decision support systems.

The remainder of the paper is organized as follows: In Section 2, we provide a review of the literature, summarize the relevant theory, and explicate our research questions and hypotheses. Section 3 details our experimental method and design, and Section 4 presents the empirical results in line with our pre-registration and provides additional exploratory analyses. Section 5 discusses our findings and Section 6 concludes.

2 Related Literature and Behavioral Predictions

Experimental designs investigating how humans tackle exploration–exploitation trade-offs have frequently employed what are known as multi-armed bandit tasks (Daw et al., 2006; Gershman, 2020; Thompson, 1933; Sutton and Barto, 2018). In these tasks, the decision-maker must choose repeatedly between two or more options that differ in their expected rewards. Each choice results in probabilistic feedback drawn from an underlying reward distribution. In stationary multi-armed bandit tasks, the reward distribution across each of the options (or "bandits") is stable over time (for a recent review, see Speekenbrink, 2022). The decision-maker thus needs to balance exploration of previously unseen or rarely chosen options with exploiting options that have produced comparatively high rewards in the past. Except for a few special cases, it is impossible to compute optimal solutions to the trade-off between exploration and exploitation in the multi-armed bandit task (Schulz and Gershman, 2019). However, reinforcement algorithms have been shown to perform well in these tasks (Sutton and Barto, 2018) and usually outperform humans in experimental settings (Gershman et al., 2015).

One reason why people tend to underperform in exploration tasks is their reliance on random exploration rather than directed exploration. Random exploration refers to a noisy decision process, that is, one that involves choices that are not goal-directed and do not maximize rewards. Exploration is considered directed if it is undertaken to reduce uncertainty about the environment by exploring particularly informative options. Directed exploration thus leads to a preference for options that have only rarely been explored and thus are high in uncertainty (Wilson et al., 2021, 2014). Although

humans have been shown to prefer exploring options with high uncertainty (Speekenbrink, 2022; Wilson et al., 2014; Wiehler et al., 2021; Zajkowski et al., 2017), their exploration behavior is often too random, particularly when cognitive resources such as working memory capacity are limited (Brown et al., 2022; Laureiro-Martinez et al., 2019; Meder et al., 2021; Wu et al., 2022). While algorithms are subject to similar limitations regarding working memory and speed, their constraints are often far less restrictive than those of humans. Hence, algorithm-based decision support systems with directed exploration have the potential to effectively help humans in exploration–exploitation tasks—if they are willing to take the advice of the algorithm.

Previous research on taking advice from algorithms has shown large differences in individuals' willingness to follow the recommendations of an algorithm (e.g., Kawaguchi, 2021; Logg et al., 2019; Mahmud et al., 2022). In general, humans seem to be willing to accept the advice of an algorithm if it is perceived to be of high quality and the expertise of the human is low (e.g., Logg et al., 2019; Saragih and Morrison, 2022; Tauchert and Mesbah, 2019; Van Swol et al., 2018). However, according to what Madhavan and Wiegmann (2007) refer to as the perfect automation scheme humans expect algorithms to work perfectly, unlike other humans, and adherence to an algorithm's recommendations decreases rapidly once the recommendations are perceived as imperfect. This can result in algorithm aversion in settings with stochastic environments, in which it is impossible to give perfect recommendations (Dietvorst et al., 2015; Dietvorst and Bharti, 2020; Prahl and Van Swol, 2017). Accordingly, one might expect algorithm aversion in a probabilistic task such as the multi-armed bandit task. In more recent research, however, participants have reported being likely to adopt an algorithm's advice when the algorithm's accuracy was above average or far greater than human performance (Saragih and Morrison, 2022). In addition, Filiz et al. (2021) have shown that algorithm aversion decreased when participants had the opportunity to evaluate the algorithm based on feedback over repeated decisions. Algorithms in multi-armed bandit tasks usually outperform human decision-makers (Gershman et al., 2015), particularly when participants have relatively little expertise in the task and receive feedback over the course of several trials. Therefore, we expect participants in general to follow the advice of a state-of-the-art learning algorithm. Accordingly, people should make more explorative choices when receiving advice from an explorative algorithm than when receiving advice from an exploitative algorithm. But given that both algorithms perform well in our parametrization, participants should benefit from the advice, independent of the exploration preferences of the algorithm.

Even though the algorithm's performance is important, it may not be the only factor influencing how much participants follow it. Most research on acceptance of the advice of algorithms has focused on one-shot decision problems, manipulating the expectations of advice quality by providing the advisee with information regarding the algorithm's past success (Hou and Jung, 2021). In contrast, exploration–exploitation tasks, such as a multiple-armed bandit task, involve a sequence of decisions in which the human decision-maker can repeatedly observe which options the algorithm suggests. The actual reward, however, is based on the participant's choice. On the one hand, this gives

participants the chance to experience the quality of the algorithm's advice for themselves when following it. On the other hand, if participants are skeptical of the algorithm in the beginning and decide not to follow its advice, they may have fewer chances to obtain an accurate impression of the advice quality. Participants' initial perceptions of the algorithm, however, may depend less on the algorithm's ultimate success rate but rather on the characteristics of the algorithm, such as its exploration tendency. Thus our first research question is:

Research Question 1: Do people accept algorithmic advice in multi-armed bandit tasks? And if so, do they benefit from it?

Because multi-armed bandit tasks involve a trade-off between exploration and exploitation, algorithms can be designed to favor exploration or exploitation (Sutton and Barto, 2018) without necessarily affecting the algorithm's performance. This raises the question of whether people might be more likely to follow advice and reap greater benefit from an explorative or an exploitative algorithm.

Algorithms using directed exploration could be attractive as people have been found to show high levels of random exploration in multi-armed bandit tasks but struggle with directed exploration (i.e., strategic information-seeking) when cognitive resources are limited (e.g., Wu et al., 2022; Meder et al., 2021). Using directed exploration, explorative algorithms could help decision-makers to direct their natural tendency to explore options with high uncertainty and thus improve the information gained from exploration. Explorative algorithms are also likely to make suggestions that are less expected by the human decision-maker than exploitative algorithms. Research on recommender systems has identified novel and unexpected advice as a means to increase user satisfaction (e.g., Adamopoulos and Tuzhilin, 2014; Castells et al., 2022).

In contrast, algorithms with a stronger tendency to exploit are by design more consistent over time in their choices. In settings with sequential choices, consistency, as opposed to changing behavior or choices, could be perceived as a signal of expertise (e.g., Falk and Zimmermann, 2017; Fehrler and Hughes, 2018; Soll et al., 2022), which would make exploitative algorithms appear as having a higher ability. Indeed, Ihssen et al. (2016) found that in a reversal learning task, participants were more willing to follow choices that were consistent over time than choices that varied between options. In addition, the variance in the outcomes of choices recommended by exploitative algorithms is likely lower than the variance in outcomes of choices recommended by an explorative algorithm. Accordingly, choices recommended by exploitative algorithms will more likely result in rewards of average magnitude that are less likely to be perceived as a loss. They are also less risky. These aspects could increase the participants' willingness to follow the advice of an exploitative algorithm.

In sum, there are arguments for both preferences—for following a more explorative or a more exploitative algorithm—and our study helps to clarify the relative value of these styles.

Research Question 2: Does the willingness to follow the algorithm's advice depend on the algorithm's exploration tendency?

Lastly, the exploration tendency of the algorithm on its own may matter less than how well it corresponds to the human it provides advice to, as is reflected in the rising importance of personalization in recommender systems (e.g., Adomavicius and Tuzhilin, 2005a; Adomavicius et al., 2008). Research on human decision-making in exploration–exploitation tasks suggests that people exhibit stable tendencies toward either exploration or exploitation in multi-armed bandit tasks (von Helversen et al., 2018; Zettler et al., 2020). Furthermore, research on advice-taking indicates that people are more willing to follow advice of people that they perceive as similar to them, for instance because they share the same gender, come from the same geographical region (Gino et al., 2009), or share similar personality traits (Tauni et al., 2019). In addition, similarity between the advice and one's own judgment in the absence of advice can impact the willingness to accept advice (Minson et al., 2011; Schultze et al., 2015). Advice that differs strongly from the advisee's own independent judgment is often ignored, even when following the advice would enhance the recipient's performance (Ecken and Pibernik, 2016; Yaniv, 2004). Similarly, Capponi et al. (2022) find that clients are especially willing to follow investment advice from a robo-advisor if it adapts to the investor's risk profile, which also results in better investment strategies. These results suggest that people may be more willing to take advice from an algorithm with an explorative tendency similar to their own. Accordingly, we hypothesize that the difference between the exploration tendency of the person and the exploration tendency of the algorithm will influence the person's willingness to follow the advice.

Research Question 3: Does the similarity between the exploration tendency of the algorithm and the human decision-maker influence the latter's willingness to accept the advice?

3 Experimental Design and Method

3.1 Study Overview

To investigate our research questions, we implemented a ten-armed stationary bandit task in an online experiment, in which the participants' goal is to maximize their rewards. Participants' payments depend on the number of points they have earned during the task, either on their own or with the help of advice from a Kalman filter algorithm with exploration bonus (KAE). The Kalman filter algorithm is a state-of-the-art reinforcement learning algorithm (Chakroun et al., 2020; Daw et al., 2006) adjusted to solve stationary bandit problems, and is particularly suited to this research given its strong performance and its ability to model human learning and choice behavior in multi-armed bandit tasks. In particular, the algorithm has been shown to better capture human behavior

than comparable models, such as delta learning models (Chakroun et al., 2020; Speekenbrink and Konstantinidis, 2015). We use the same Kalman filter algorithm to provide advice to participants in the experimental setting and to *ex post* model the human decision-process, which enables us to measure similarity in the exploration behavior based on the algorithm's exploration parameter. For this purpose, the individual explorative tendency of the participants estimated by the model from the experimental data is compared with the explorative tendency of the algorithm.

The experiment consists of two phases. In the first phase, all participants perform the multi-armed bandit task without advice, which allows us to estimate the participants' own tendency to explore. In the second phase, we implement four treatment conditions between participants: three treatments in which participants perform the multi-armed bandit task with advice from an algorithm, and a control treatment in which participants perform the task without advice, to ensure that potential increases in performance are due to the advice and not due to learning. In the treatments with advice, the algorithm gives recommendations on which alternative to choose, but leaves it to the human participants to make the final decision. The exploration tendency of the algorithm varies by three levels: explorative, exploitative, and balanced. The algorithm in the explorative treatment has a stronger tendency toward exploration. In the exploitative treatment, the algorithm is more inclined toward exploitation. In the balanced treatment, the algorithm's tendency to explore lies between the explorative and exploitative treatment. Note, that the algorithm provides its outline of recommendations based on the participant's actual choices and the algorithm's tendency to explore. Thus, our experimental design captures a true interaction between a human and an algorithm.[1]

3.2 Participants

Participants were recruited through Prolific Academic. Participation was restricted to UK residents to ensure good knowledge of English and comparable value of the study incentives. Participants had to correctly answer comprehension questions about the task goal, reward scheme, and reward generation procedure in order to participate in the experiment. Participants who failed the comprehension questions were removed from the experiment, and replaced by new participants until the pre-registered number of participants was reached. Six hundred participants finished the study. Among these, thirteen have been be excluded from the data analyses: twelve because their data was not uploaded due to technical issues, and one participant because he indicated that he had failed to read the instruction regarding the algorithm. Participants' payments consisted of a base payment of £2.50 and a performance-dependent bonus payment. The average bonus was £4.14 per participant. Of the participants willing to provide their demographic details, 65.1% are female, 34.2% are male, and 0.7% are non-binary. The average age is 40.8, with a standard deviation of 23.4 years.

[1]The main hypotheses and analyses were preregistered on the Open Science Foundation (OSF) website, `https://osf.io/e3hqa`.

3.3 Bandit Task and Experimental Procedure

In the experiment, participants perform a total of eight instances of a stationary multi-armed bandit task with 10 bandits: one training task and three experimental tasks each in the first and the second phase of the experiment. In each task, participants encounter ten boxes (the bandits) and must undertake several trials. In each trial, participants may choose one box by clicking on it with a mouse. After selecting a box, its contents are revealed and the participant receives feedback about the number of points collected from the box. The goal is to collect as many points as possible, and participants have been told that their payout depends on the average number of points they accumulate over the course of the experiment. The boxes differ in their underlying reward distributions, whose means have been randomly sampled from a normal distribution with a mean of 50 points and a standard deviation of 10. Each time a box is selected, a random number is drawn from a normal distribution with a mean of the average points of the box and again a standard deviation of 10. The drawn number is then rounded to the closest integer. For a visualization of an example reward distribution, see Figure 1, Panel A. A training task consists of 20 trials and a experimental task consists of 70 trials. Within a task, the average points that a box will yield stays the same, that is, within an experimental task participants have 70 choices to learn which boxes result, on average, in high payoffs and which result in low payoffs. For each new task, the average points for each box are redrawn. During a task, participants are informed about the current number of trials and the maximum number of trials in the task. After each task, they are informed about the average number of points they have acquired during the task.

Participants begin the experiment with a declaration of consent informing them about the processing of their data and the university conducting the experiment (University of Bremen). After providing consent, participants receive detailed instructions explaining the task, the experimental procedure, and the reward scheme.[2] In the first phase of the experiment, participants perform one training task followed by three experimental tasks without advice from an algorithm. In the second phase of the experiment, participants again perform one training task followed by three experimental tasks. Depending on the treatment participants are in, they either perform the training and experimental tasks on their own (control) or, at the beginning of the phase, read additional instructions regarding the algorithm's advice, before receiving these recommendations throughout the training and experimental tasks. The advice depends on the treatment and is provided by either an explorative, balanced or exploitative algorithm. The recommendation of the algorithm is indicated by a yellow square surrounding the suggested box. Participants are informed that the algorithm has received the same information as the participants, that is, the algorithm is aware of the distribution of the average bandit points and the standard deviation of each draw from a bandit. The algorithm is also informed of the number of points from the bandits picked by the participants. The recommendations from the algorithm are made based on this information.

[2]See Appendix H for the consent form and the instructions.

After the participants finish the second phase, they are informed about the average points they have acquired over the course of the experiment and the amount of the bonus payment they will receive. Participants are then asked several questions regarding their subjective perceptions of the experiment. Specifically, they are asked whether there are any reasons to disregard their responses, and to rate the effort they put into the experiments and their tendency to rely on exploitation and exploration throughout the experiment. If the participants belong to the treatments with algorithm recommendations, they are also asked about the usefulness of the recommendations, how much they relied on the algorithm, and how much they followed the algorithm, each using a slider that records their responses between 0 (not at all) and 1 (very much). All the participants are given the opportunity to leave comments regarding the experiment before ending the experiment.[3]

3.4 Algorithm Recommendations

To generate the recommendations, we use a Kalman filter algorithm with exploration bonus (KAE) (Chakroun et al., 2020; Daw et al., 2006). In each trial $t \in \{1, \ldots, 70\}$, the algorithm uses Bayesian learning to generate posterior expectations about the mean and standard error of the mean of the points generated by each bandit. Let $\hat{\mu}_{k,t}$ denote the posterior expectation of the mean, and $\hat{\sigma}_{k,t}$ the posterior expectation of the standard error of the mean of bandit k in trial t. In the first trial, the expected means and standard errors correspond to the prior expectation, i.e., $\hat{\mu}_{k,1} = 50$ and $\hat{\sigma}_{k,1} = 10$ for all bandits. In all subsequent trials, the algorithm updates the expected mean and standard error of the mean of the bandit selected in the current trial based on the Kalman filter (Daw et al., 2006). The Kalman filter uses the following three recursive equations to update the posterior expectations of the selected bandit s conditional on the reward R_t obtained by this bandit in trial t:

$$\begin{aligned} \hat{\mu}_{s,t} &= \hat{\mu}_{s,t-1} + K_t(R_t - \hat{\mu}_{s,t-1}) \\ \hat{\sigma}^2_{s,t} &= (1 - K_t)\hat{\sigma}^2_{s,t-1} \\ K_t &= \frac{\hat{\sigma}^2_{s,t-1}}{\hat{\sigma}^2_{s,t-1} + \sigma^2_{s,0}} \end{aligned} \tag{1}$$

The quantity K_t is called the Kalman learning rate, which is updated based on the expected variance of the selected option. As the expectation for the standard error of the mean equals the known standard deviation of the average rewards before the bandit is sampled, and subsequently decreases, the learning rate starts at one-half and decreases over trials. This implies that the expectation of the standard error of the mean decreases every time the bandit is sampled, which is a specific feature of stationary bandits.

The algorithm recommendation in trial t depends on the posterior expectations $\hat{\mu}_{k,t}$ and $\hat{\sigma}_{k,t}$, and a parameter ϕ that reflects the algorithm's tendency toward exploration. The recommendation

[3]The experiment was programmed in PsychoPy (Peirce et al., 2019) and exported to Pavlovia.

for trial t is the bandit with the largest value of the attraction score $q_{k,t} = \mu_{k,t} + \phi\sigma_{k,t}$. If there are multiple bandits sharing the largest attraction score, the bandit with the larger index s will be recommended.

The parameter ϕ controls the exploration tendency of the algorithm by scaling the effect of the estimated standard error of the mean on a bandit's attraction score. If ϕ is large, bandits with a large standard error of the mean (and which are less frequently chosen) have large attraction scores. If ϕ is small, uncertainty does not attract, and the algorithm will mostly recommend bandits based on the posterior expectations about their mean reward.

We use three different values of the exploration parameter ϕ across the three treatments with algorithm recommendations. In the exploitative treatment $\phi = 0.4$. The balanced treatment uses $\phi = 1.4$, and the explorative treatment $\phi = 3.3$.

The three values of ϕ have been selected based on their performance in 10,000 simulated tasks of our experiment. The values used in the exploitative and explorative treatments both generate the same average reward of approximately 61 points across the 10,000 simulated tasks. These values have been selected because they generate an average reward comparable to Bayesian learning in combination with Thompson sampling, which is known to perform well in multi-armed bandit tasks (Sutton and Barto, 2018). The value of $\phi = 1.4$ that we use in the balanced treatment produces an average reward of 62 points across the 10,000 simulation runs, which suggests an optimal balance between exploration and exploitation. Panels B–D of Figure 1 illustrate how the advice by the algorithms in the advice treatments differs by presenting an exemplary sequence of choices when the algorithms perform the task without the human participant.

4 Results

First, we check whether the different algorithms' exploration tendencies in our experimental treatments influence participants' behavior. Following the preregistration, we test whether the algorithm's recommendations influence participants' exploration behavior, the rewards they earned, and how frequently their choices match the recommendation.[4] Second, we report the cognitive modeling results and how participants' exploration tendencies affected their willingness to follow the algorithm's advice, as well as participants' subjective perceptions of the algorithm's usefulness.

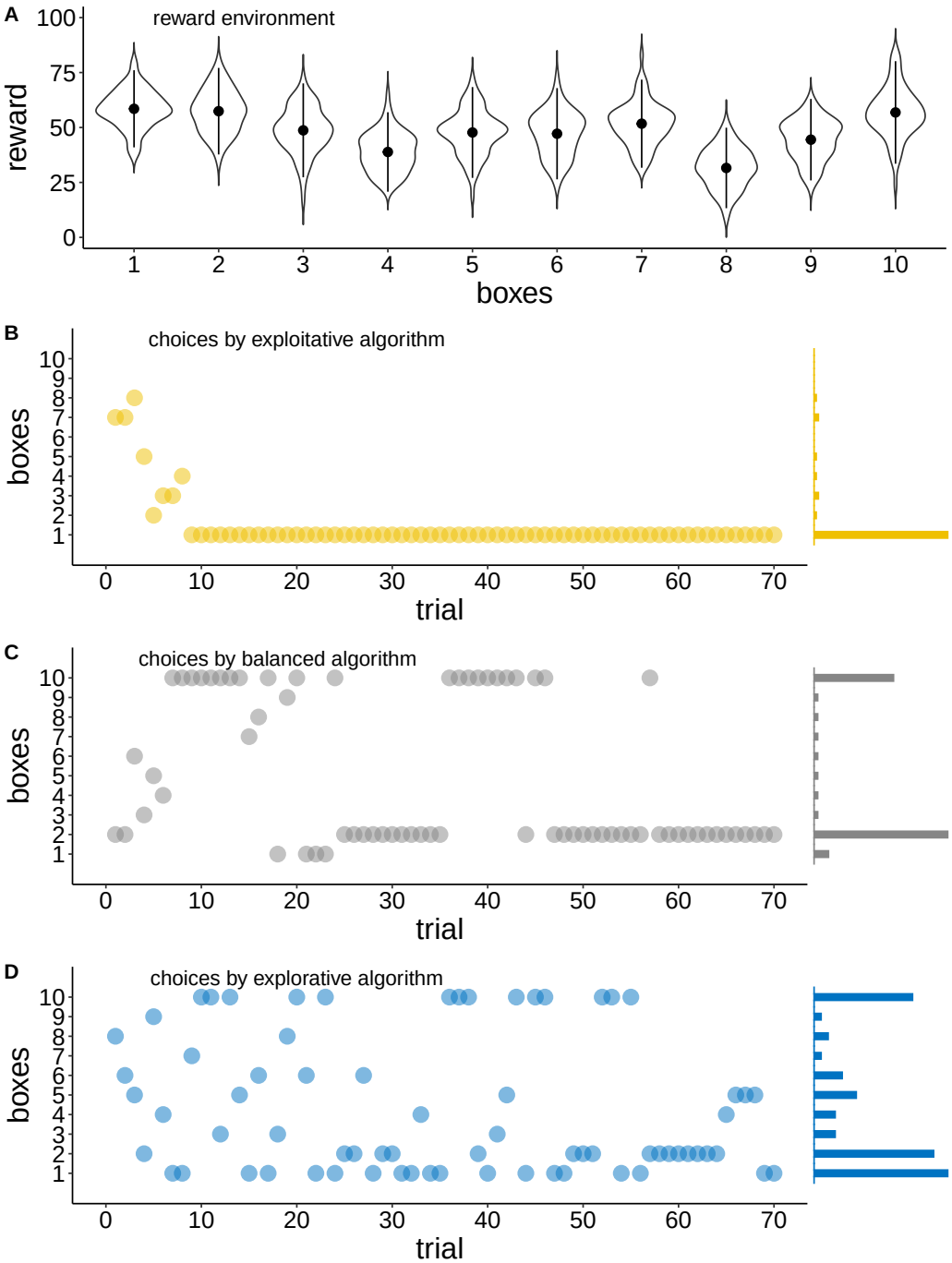
4.1 Behavioral Analyses

4.1.1 Exploration Tendency: Number of Switches.

Participants' exploration tendency is quantified by measuring how frequently they switched between bandits within each of the two experimental phases, averaged across tasks, with a higher number of

[4]The Figures in Appendix F report changes in behavior over time, i.e., across tasks.

Figure 1: Exemplary Algorithm Choices

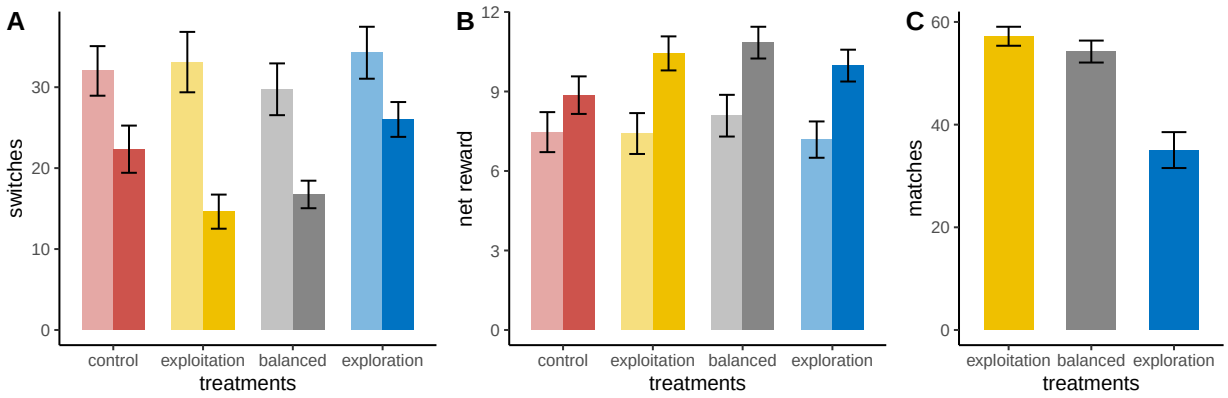


Note: Panel A depicts an exemplary reward environment. When facing this environment alone, i.e., with no human participant involved, the exploitative algorithm produced the choice sequence depicted in Panel B, the balanced algorithm produced the choice sequence depicted in Panel C, and the explorative algorithm produced the choice sequence depicted in Panel D.

switches indicating a higher exploration tendency.

By summarizing the average number of switches per task in the two phases, Panel A of Figure 2 shows that participants in general explored more frequently in the first phase ($M = 32.28$, $SD = 20.47$) than in the second phase ($M = 19.94$, $SD = 14.62$; $F(2, 1172) = 141.3$, $p < .001$, partial $\eta^2 = .11$; see also Table 1).

Figure 2: Summary of Behavioral Measures by Treatment and Phase.



Note: Panel A shows the average number of switches, Panel B the average net rewards minus the expected value of 50, and Panel C the average percentage of matches between participants' choices and the algorithm's suggestions. All measures are averaged across the three tasks in phase 1 (light colors) and phase 2 (dark colors), respectively. Error bars indicate 95% confidence intervals.

To test whether the characteristics of the algorithm influenced participants' exploration behavior, we analyze whether treatment differences affect the number of switches in phase 1 and 2 differently. The ANOVA for phase 1, with no recommendations in all treatments, shows no differences between the four treatments ($F(3, 583) = 1.31$, $p = .271$, partial $\eta^2 = .007$), showing that initial exploration tendencies of the participants were balanced across treatments. In phase 2, both an ANOVA including all four treatments and one focusing only on the three recommendation treatments show a significant effect of treatment ($F(3, 583) = 20.47$, $p < .001$, partial $\eta^2 = .10$ and $F(2, 439) = 35.42$, $p < .001$, partial $\eta^2 = .14$, respectively), indicating that the recommendations changed how much participants explored.

To investigate the differences in exploration behavior between treatments more closely, we conducted post-hoc pairwise comparisons between the treatments using Tukey HSD corrections for multiple tests.[5] The pairwise comparisons show that participants in the exploitation treatment ($M = 14.6$, $SD = 13.0$) explore less than participants in the exploration treatment ($M = 26.0$, $SD = 13.4$; $exploit - explore = -11.39$, $p < .001$), but not compared to participants in the balanced treatment ($M = 16.7$, $SD = 10.6$; $balance - exploit = 2.12$, $p = .559$). Participants in the exploration treatment explored more than in the balanced treatment ($balance - explore = -9.27$,

[5]See Tables E1-E3 in Appendix E for more details. In Appendix G, we show the empirical cumulative distributions of switches in the two phases across treatments.

Table 1: Descriptive Overview of Behavioral Measures by Treatment

	treatments	phase	M	SD	95% lower bound	95% upper bound
number of switches	control	1	32.01	18.82	28.94	35.07
	exploitation	1	33.10	23.00	29.37	36.83
	balanced	1	29.74	19.85	26.54	32.94
	exploration	1	34.26	19.92	31.05	37.47
	control	2	22.34	17.92	19.42	25.25
	exploitation	2	14.62	12.99	12.51	16.73
	balanced	2	16.75	10.55	15.05	18.45
	exploration	2	26.01	13.35	23.86	28.17
rewards	control	1	57.47	4.64	56.71	58.22
	exploitation	1	57.41	4.75	56.64	58.18
	balanced	1	58.09	4.89	57.30	58.87
	exploration	1	57.18	4.26	56.50	57.87
	control	2	58.86	4.35	58.15	59.57
	exploitation	2	60.44	3.96	59.80	61.08
	balanced	2	60.84	3.72	60.24	61.44
	exploration	2	59.98	3.73	59.38	60.58
number of matches	control	-	-	-	-	-
	exploitation	2	57.21	11.40	55.37	59.06
	balanced	2	54.23	13.27	52.09	56.37
	exploration	2	35.04	21.70	31.55	38.54

$p < .001$). In comparison to the control treatment ($M = 22.3$, $SD = 17.9$), participants explored less in both the exploitation and the balanced treatment ($exploit - control = -7.72$, $p < .001$; $balance - control = -5.60$, $p = .004$), but there was no difference in the exploration condition, ($explore - control = 3.68$, $p = .110$).[6]

Result 1: Participants' exploration behavior depends on the exploration tendency of the algorithm. When participants are presented with an exploitative algorithm, they tend to explore less (exploit more) than when they are presented with an explorative algorithm or when exploring on their own.

4.1.2 Rewards.

Next, we turn to the question of whether participants' performance improves when receiving recommendations from the algorithm. Panel B of Figure 2 shows the net rewards (i.e., the average number of points per task minus the expected value of 50) participants achieved during the two phases. Overall, participants' performance increased from phase 1 ($M = 57.54$. $SD = 4.64$) to phase 2 ($M = 60.03$, $SD = 4.00$; $F(1.1172) = 97.37$, $p < .001$, partial $\eta^2 = .08$, see also Table 1).

[6]Using Bonferroni adjustments instead of the Tukey corrections leads to comparable results (details not reported here).

In the first phase, in which participants have not yet received advice from an algorithm, we find no significant differences between the treatment conditions, ($F(3, 583) = 1.02$, $p = .382$, partial $\eta^2 = .005$), as expected. In phase 2, however, net rewards differ significantly depending on the four treatment conditions, ($F(3, 583) = 6.87$, $p < .001$, partial $\eta^2 = .03$). Importantly, this difference seems to stem from lower performance in the control condition, given that an ANOVA focusing on the three recommendation treatments without control does not show a significant difference between these treatments ($F(2, 439) = 1.93$, $p = .147$, partial $\eta^2 = .009$).

Again we conducted follow-up paired comparisons between the treatments using the Tukey HSD correction. In comparison to the control treatment ($M = 58.9$, $SD = 4.35$), subjects earned higher rewards in the exploitation treatment ($M = 60.4$, $SD = 3.96$; $exploit - control = 1.58$, $p = .004$) and the balanced treatment ($M = 60.8$, $SD = 3.72$; $balance - control = 1.98$, $p < .001$), but not in the exploration treatment ($M = 60.0$, $SD = 3.73$; $explore - control = 1.12$, $p = .074$). However, the three treatments with the AI did not differ significantly from each other ($exploit - explore = 0.46$, $p = .749$; $balance - explore = 0.87$, $p = .233$; $balance - exploit = 0.41$, $p = .814$).[7]

Result 2: Participants earn greater rewards when receiving advice from an exploitative or balanced algorithm than when performing the task on their own.

4.1.3 Matches.

Next, we look at how frequently participants' choices coincided with the recommendations from the algorithm, as a rough measure of how willing participants are to follow the algorithm's recommendations. As Panel C of Figure 2 illustrates, the three treatments differ in the average frequency with which participants chose the bandit that was recommended by the algorithm ($F(2, 439) = 82.26$, $p < .001$, partial η^2=.27, see also Table 1).

Specifically, follow-up Tukey tests with HSD corrections show that participants' choices match the algorithm's recommendations less frequently in the exploration treatment than in the other two treatments ($exploit - explore = 22.17$, $p < .001$; $balance - explore = 19.19$, $p < .001$). However, comparing the balanced treatment with the exploitation treatment reveals no significant difference in the frequency of matches between these two treatments ($balance - exploit = -2.98$, $p = .252$). Overall, the results suggest that participants seem to be more willing to follow the algorithm's recommendation when the algorithm is exploitative or balanced than when it is explorative.

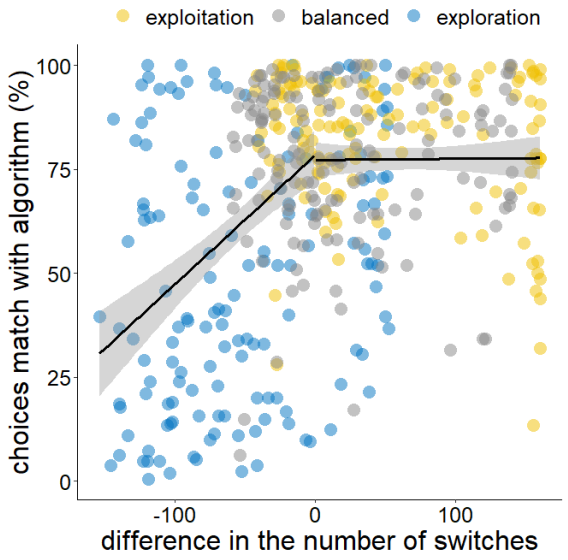
Result 3: Participants' decisions coincide less frequently with the algorithm's recommendations when the algorithm has a more explorative tendency.

[7]Independent sample t-tests with Bonferroni corrections for multiple hypothesis testing confirm the result, showing a significant advantage for the exploitation and the balanced treatment, but not for the exploration treatment over the control treatment (adjusted p-values are $p = .004$, $p < .001$ and $p = .095$, respectively).

To investigate whether the frequency of matches depends on tendency to explore, we regressed the number of times a participant followed the algorithm's recommendations on the difference between the participant's exploration behavior in phase 1 (the number of switches in tasks 1–3) and the algorithm's recommended exploration behavior in phase 2 (the average number of recommended switches by the algorithm in tasks 4–6) and the squared difference using OLS regression controlling for age, gender and treatment conditions.[8] The regression shows that the recommendations of the algorithm are followed more frequently if the algorithm is less explorative than the participant (see Tables B1 and 2 in Appendix B). We also find a smaller but significant effect of the squared difference, which suggests a curve-linear relationship. Moreover, when splitting the data set into participants that are more or less explorative than the algorithm, as shown in Figure 3, we find that the tendency to make the same choice declines when the algorithm is more explorative than the participant but not when the algorithm is more exploitative than the participant (see Table 2 for the corresponding regression results).

Result 4: Participant behavior is more likely to coincide with an algorithm's decisions when the algorithm is more exploitative and therefore more consistent over time than the participant.

Figure 3: Scatter Plot of the Difference in the Tendency toward Exploration of Participants and the Algorithm with Matching



Note: The figure shows two regression lines of percentage of matching on difference in the number of treatments. We look at difference below and above zero. The regression lines use all observations from non-control treatments. The difference in the number of switches is the participant's number of switches in phase 1 minus the algorithm's treatment-wise average recommended number of switches in phase 2.

[8]We also performed a one-way ANOVA with 1,000 iterations as a balance check on age and gender. The p-values are .464 for age and .495 for gender, indicating again that the participants are well randomized into the treatments.

Table 2: Effects of Dissimilarity and Algorithm Consistency on Matching

	without consistency			with consistency		
	Est	SE	p-value	Est	SE	p-value
constant	164.14	5.14	$< .001$	216.07	3.98	$< .001$
I(more explorative)	-11.36	5.86	.053	-2.12	3.84	.582
dissimilarity	-0.86	0.05	$< .001$	-0.17	0.04	$< .001$
I(more explorative) $\times$ dissimilarity	1.01	0.07	$< .001$	0.18	0.06	.001
consistency	-	-	-	-0.75	0.03	$< .001$
N	442			442		
R^2	0.55			0.81		

Note: Linear regressions of number of matches on a dummy variable (*I(more explorative)*) that indicates whether the participant explores more than the algorithm; the *dissimilarity* in the exploration tendency, which is is the absolute difference between the number of switches of participants in phase 1 and the number of recommended switches from the algorithm in phase 2; and the interaction of the dummy and dissimilarity. The second model additionally controls for the *consistency* of the algorithm, i.e., the number of switches recommended by the algorithm in phase 2. Age and gender are used as socio-demographic control variables.

4.2 Cognitive Modeling

A potential issue with behavioral measures of exploration and algorithm following is that these measures do not control for the variation in rewards observed by the participant. For example, in situations in which one bandit stands out due to large rewards, even a participant with a strong preference for exploration will rarely switch. For the same reason, a participant might often follow the recommendations of the algorithm in the second phase of the experiment simply because recommendations and preferences are very similar, which makes it difficult to judge the latent tendency to follow recommendations based on the frequency of following. Thus, to corroborate the behavioral results, we use cognitive modeling to estimate each participant's latent tendency toward exploration as well as the latent tendency to follow the recommendations of the algorithm.

To estimate the latent exploration tendency of each participant, we fitted a model with a random exploration parameter β to the first-phase choices. The model uses the same Bayesian learning algorithm we used to generate the recommendations, which returns an attraction score $q_{s,t}$ for each bandit based on the posterior expectations about the mean and standard deviation of the points generated by the bandit. Instead of using a deterministic choice rule, we use the *softmax* function with inverse temperature β to translate the attraction scores of the bandits into stochastic choices to account for decision errors of the participants. The probability of participant i to select bandit s in trial t is:

$$Pr(s) = \frac{e^{\beta_i q_{s,t}}}{\sum_k e^{\beta_i q_{k,t}}} \tag{2}$$

We use Bayesian multilevel modeling to estimate the inverse temperature parameter β_{i1} for each

participant in the first phase of the experiment.[9] To assure that the inverse temperature is positive, we use $\beta_1 = e^x$ and a normally distributed prior for the underlying variable x. We report the natural logarithm $log(\beta_1)$ of the inverse temperature in the first phase of the experiment to obtain an approximately normally distributed variation measure of participants' exploration tendencies.

We also fitted the same model to the choices participants make in the second phase of the experiment (tasks 4–6), which yields a parameter estimate β_2 for each participant. For the data of the control condition, comparing β_1 to β_2 allows us to assess the stability of the latent exploration tendency over the course of the experiment. For the treatments with algorithm recommendations, the model fits provide benchmarks for answering the question of whether accounting for the algorithm recommendations increases out-of-sample prediction accuracy.

To make reasonable comparisons between the exploration tendency of a participant and the exploration tendency of the algorithm, we also fitted the model to the algorithm recommendations. We did this on the individual level for each participant separately to obtain the algorithm's exploration tendency β_{alg} from the perspective of each participant. We use $log(\beta_1)$–$log(\beta_{alg})$ as a measure of difference between the exploration tendency of a participant and the algorithm, with positive values indicating that the participant has a stronger tendency toward exploitation.

We also fitted a two-parameter model with a random exploration parameter and a directed exploration parameter to the data of the first part of the experiment. The individual parameter estimates show that the variation in exploration behavior as measured by the number of switches stemmed from different degrees of random exploration. Participants generally avoided directed exploration, with little individual variation in this parameter. This can be attributed to our use of stationary bandits that discourage exploration in the later trials of each task.

For the treatment conditions with recommendations, we use a different model for the second-phase choices. In this model, we do not estimate the random exploration parameter for each individual. Instead, we fixed β to the mean of the posterior parameter distribution of each participant that we estimated based on their first-phase choices. This allows disentangling of the effect of algorithm recommendations from participants' exploration tendencies. To this end, we added a dummy $I_{s,t}$ to the equation of the attraction scores indicating whether the bandit was recommended by the algorithm, which yields:

$$q_{k,t} = \mu_{s,t} + \rho_i I_{s,t-1} \tag{3}$$

The parameter ρ_i of this dummy reflects the participant i's inclination to follow the recommendations of the algorithm. Positive values of ρ increase the probability that the recommended bandit is chosen. As in the model of first-phase choices, attraction scores are translated into choice probabilities using a *softmax* with inverse temperature fixed to the estimate of β_1.

The models were fit to the choice data using the Bayesian model fitting software Stan (Carpenter

[9]For brevity, we drop the subscript i from here onward. Keep in mind however, that β_1 takes different values between participants.

et al., 2017), accessed via rstan (Stan Development Team, 2022). For each treatment condition (exploration, balanced, exploitation, control) we fitted two Bayesian multilevel models: one for the first-phase choices and another for the second-phase choices. To approximate the posterior distribution of model parameters, we run four Markov Chains parallel with 3,500 iterations, of which 1,000 iterations were discarded as warm-up. Using this procedure, we obtain converged chains (max(Rhat) = 1.049) with informative posterior distributions (min(ESS) = 162) and no divergent transitions. Past research indicates that the models used meet the standards of parameter and model recovery (Danwitz et al., 2022).

For the population average of β, we use a log-normal prior with mean 0.5 and standard deviation of 0.1, and normal distribution with mean 0 and standard deviation of 0.01 censored at zero as prior for standard deviation of β in the population. For the model parameter ρ, we use a standard normal prior for the population mean and a censored standard normal distribution for the standard deviation of ρ in the population. Using these priors, we conducted a parameter recovery simulation showing that individual model parameters are recovered well from simulated data (see Figure C1 in Appendix C). Table D1 in Appendix D reports the posterior of the population mean and standard deviation of the model parameter for each fitted model, along with criteria for convergence and estimates of the out-of-sample prediction accuracy of each model (Vehtari et al., 2017). Comparing the out-of-sample prediction accuracy of the ρ models to the β models for the data of second-phase choices indicates that modeling choices based on recommendations improves out-of-sample prediction accuracy (exploitation: $z = -12.05, p < .001$; balanced: $z = -33.19, p < .001$; exploration: $z = -47.57, p < .001$).

4.2.1 Analyses of Participants' Latent Tendencies toward Exploration and Inclination to Follow the Algorithm.

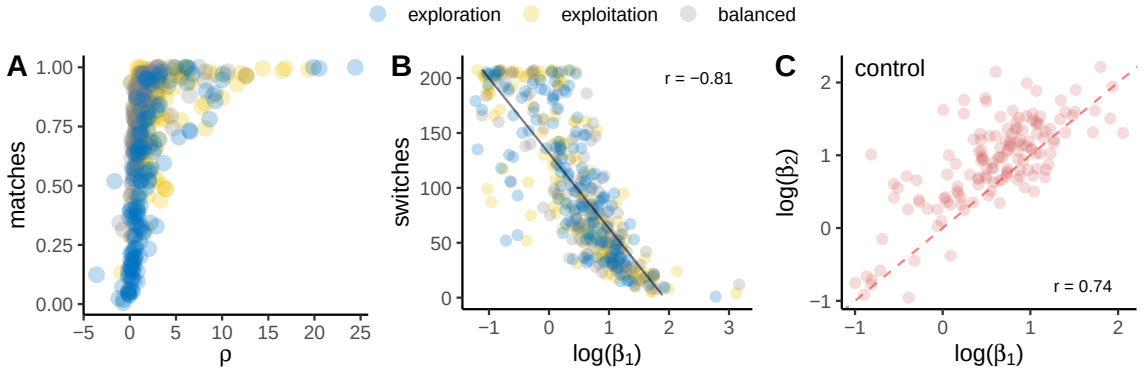
As a first step, we examine whether the estimated latent variables capture individual differences in the behavioral measures. As illustrated in Figure 4, the parameters show high correlations with the corresponding behavioral measures. Panel A shows the relationship between the parameter ρ, which measures the latent tendency to follow recommendations, and the number of times a participant chooses the bandit recommended by the algorithm in the second part of the experiment. Both variables are positively correlated, with Pearson's $r = 0.43$. The scatter plot also clearly shows the *softmax* function the model uses to translate recommendations into choice probabilities, with values of $\rho > 3$ resulting in more than 90% matches on the behavioral level. Panel B shows that the latent exploration tendency, as measured by $log(\beta_1)$, is negatively correlated with the number of switches in the first phase of the experiment. The larger a participant's latent exploration parameter, which indicates more exploitation and less exploration by the algorithm, the less frequently a participant switches between bandits (Pearson's $r = -0.81$, $t(440) = -28.87$, $p < .001$). Panel C of Figure 5 illustrates that the latent exploration tendency is fairly stable over the two phases of the experiment

for data of the control treatment (Pearson's $r = 0.74$, $t(143) = 13.00$, $p < .001$). There is a slight tendency toward less exploration in the second phase, indicated by the fact that most observations are above the 45° line.

Result 5a: The latent measures of a participant's exploration tendency and willingness to follow an algorithm correlate highly with the corresponding behavioral measures.

Result 5b: Participants' latent exploration tendencies are relatively stable over time.

Figure 4: Correlations between Parameter Estimates and Behavioral Measures

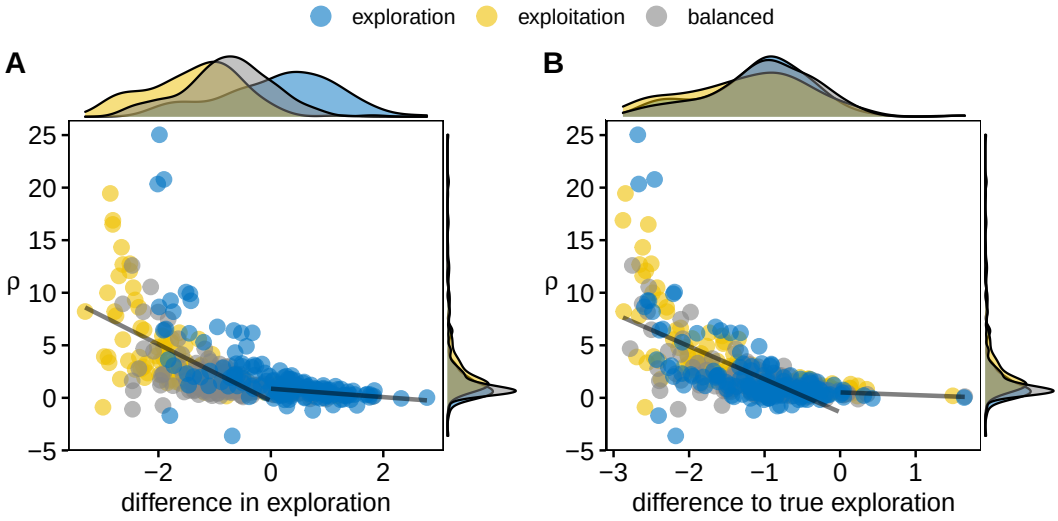


Note: Panel A shows the relationship between the latent tendency to follow recommendations ρ and the average frequency of choosing the recommended bandit in the second phase. Panel B shows the latent exploration tendency (higher $log(\beta_1)$ corresponds to less exploration) and the number of switches in the first phase. Panel C shows the correlation between the estimated latent exploration tendencies of the first and second phase of the control treatment.

Corresponding to the behavioral analyses, we use the latent parameters, estimated based on the cognitive modeling, to investigate how the similarity between the participants' and the algorithms' exploration tendencies relates to participants' inclination to follow the algorithm's advice. As illustrated in Panel A of Figure 5, we find a negative relationship between the estimated latent tendency to follow the recommendations of the algorithm and the difference in the exploration tendency of the participant and the algorithm. We use $log(\beta_1)$–$log(\beta_{alg})$ as a measure of difference between the exploration tendency of a participant and the algorithm, with negative values indicating that the participant explored more in phase 1 than the algorithm recommended in phase 2. The latent tendency to follow recommendations is largest for participants who explored more than the algorithm recommends. The tendency to follow the recommendation is substantially lower if the exploration tendencies of the participant and the algorithm were similar and stay at a very low level when the exploration tendency of the algorithm was higher than the participants' exploration tendency. On average, the tendency to follow recommendations is largest in the treatment with the exploitative

algorithm, which also features the most cases in which the algorithm was more exploitative than the participants.

Figure 5: Scatter Plot of the Differences in Exploration Tendencies and the Inclination to Follow the Algorithm



Note: The figure shows the relation of the parameter estimates of the latent tendency to follow recommendations ρ and the difference in the exploration tendency of the participant and the algorithm, with higher values indicating stronger exploitation (i.e., less exploration) by the participant than the algorithm. Panel A compares the participants' exploration tendency with the perceived exploration tendency of the algorithm ($log(\beta_1)$–$log(\beta_{alg})$) and Panel B with the uninfluenced exploration tendency of the algorithm ($log(\beta_1)$–$log(\beta_{exo})$. Colors indicate the different treatment conditions with recommendations. Parameter distributions are shown on the opposite axes respectively.

Given that the algorithm uses the feedback from the choices of a participant to generate a recommendation, participants' choice behavior can influence the degree to which the algorithm recommends exploration or exploitation and thus the exploration tendency displayed by the algorithm. We estimate the uninfluenced exploration tendency of the algorithm based on simulated recommendations and compare it with the exploration tendency estimated from the recommendations in the second phase. To this end, we simulated the recommendations the algorithm would have given if the participant had followed all recommendations and fitted the model for the first-phase choices to the simulated recommendations. The resulting exogenous exploration tendency β_{exo} allows us to study how the difference between the participant's and the algorithm's true exploration tendency $log(\beta_1)$–$log(\beta_{exo})$ is related to the latent tendency to follow the recommendations. Panel B of of Figure 5 shows a very similar negative, non-linear relationship between the estimated latent tendency to follow the recommendations of the algorithm and the difference between the participant's and the algorithm's unbiased exploration tendency. Table D2 in Appendix D shows that the non-linear relationship between the tendency to follow recommendations and the difference between the participant's and the algorithm's unbiased exploration tendency prevails in all treatments when controlling

for socio-demographic variables.

We also test whether the relationships (depicted in the two panels of Figure 5) are significant when controlling for consistency of recommendations, gender, and age. To this end, we regress the individual estimates of ρ on a dummy that indicates whether the participant explores more than the algorithm $I(log(\beta_1) < log(\beta_{exo}))$, the dissimilarity in the exploration tendency $|log(\beta_1)–log(\beta_{exo})|$, and their interaction, controlling for the consistency $log(\beta_{exo})$ of the algorithm. Table 3 shows that the dissimilarity between the exploration tendency of the participant and the algorithm is positively related to the inclination to follow the algorithm recommendations if the exploration tendency of the participant is stronger. This corresponds to the linear fits depicted for the data points below zero in Figure 5. For this data, the dissimilarity has a significant positive effect on ρ in both models, even when controlling for the consistency of the recommendations, which shows that relative differences in the exploration tendency matter.

Table 3: Effects of Dissimilarity and Algorithm Consistency on ρ

	diff exploration						diff true exploration					
	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
constant	0.72	0.42	0.084	1.92	0.43	$< .001$	0.25	0.62	0.690	1.96	1.16	0.091
I(more explorative)	-1.25	0.45	0.006	-0.38	0.45	0.391	-1.95	0.63	0.002	-1.92	0.63	0.003
dissimilarity	-0.41	0.41	0.319	-1.56	0.42	$< .001$	-0.39	0.99	0.693	-0.47	0.99	0.637
I(more explorative) ×dissimilarity	3.21	0.46	$< .001$	4.75	0.48	$< .001$	3.62	1.01	$< .001$	3.74	1.01	$< .001$
consistency	-	-	-	-1.75	0.24	$< .001$	-	-	-	-1.12	0.64	0.081
N	442			442			442			442		
R^2	0.40			0.46			0.46			0.46		

Note: Linear regressions of the latent tendency to follow algorithm recommendations, on a dummy that indicates whether the participant explores more than the algorithm (*I(more explorative)*), the *dissimilarity* in the exploration tendency $|log(\beta_1)–log(\beta_{alg})|$ of the participant and the algorithm, the interaction of the dummy, and dissimilarity. The second models additionally control for the consistency $log(\beta_{alg})$ of the algorithm. The right columns show the same models using $log(\beta_{exo})$ instead of $log(\beta_{alg})$. Age and gender are used as socio-demographic control variables.

Result 6: Participants are more inclined to follow the algorithm if it is less explorative and therefore more consistent over time than they are.

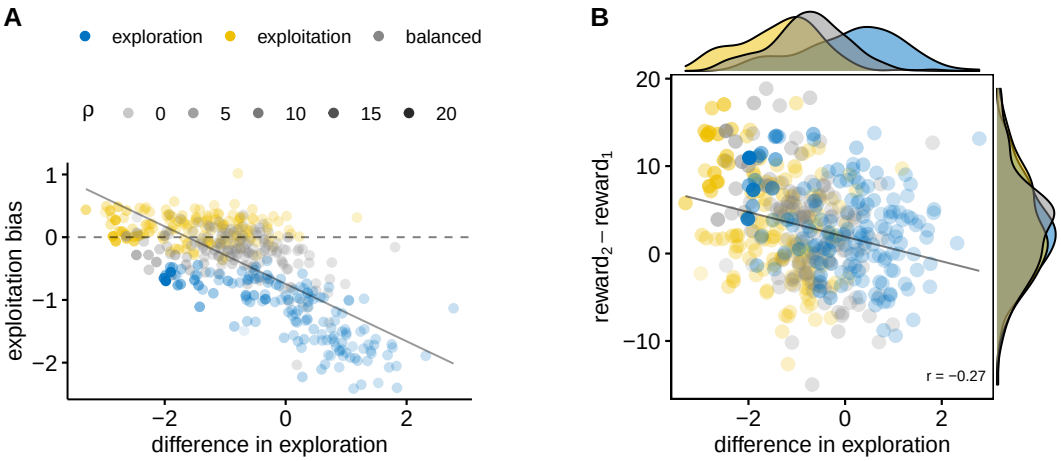
4.2.2 Biased Perceptions and Benefits of Recommendations.

In order to characterize how participants' choices bias the level of exploration exhibited by the algorithm, we conducted exploratory analyses that go beyond our preregistration and focus on the absolute bias $log(\beta_{alg}) - log(\beta_{exo})$ introduced by the choices of the participant. Panel A of Figure 6 relates the absolute bias induced by a participant's choices to the difference in the exploration tendency of the participant and the algorithm. Biases are mainly negative, which means that the participant's perception of the algorithm is distorted toward more exploration. The scatter plot in

Panel A of Figure 6 shows that differences in exploration explain the bias (Pearson's $r = -0.72$, $t(440) = -21.66$, $p < .001$). It further shows that positive differences in exploration, which are mainly observed in the exploration treatment, correspond to negative biases in the perception of the algorithm, which gives the impression that recommendations are more random. In contrast, the perception bias is smaller for participants with negative differences in exploration, who are also those that follow recommendations more frequently (see Figure 5).

Result 7: The perception of the algorithm's behavior is biased in the direction of more exploration by interacting with a human participant. This bias is particularly strong when the algorithm is more explorative than the human.

Figure 6: Biased Algorithm Perceptions and Heterogeneous Benefits



Note: A: Difference in the exploration tendency $log(\beta_1)$–$log(\beta_{alg})$ of the participant and the algorithm (positive values indicate more exploration by the algorithm than the participant) and bias in the perception of the exploration tendency of the algorithm $log(\beta_{alg}) - log(\beta_{exo})$ (negative values indicate a bias toward more perceived exploration). B: Difference in exploration tendency and difference in reward between the first and the second phase of the experiment.

Based on the individual parameter estimates, we can also answer the question of who benefits most from recommendations. Panel B of Figure 6 plots the differences in average reward over the two phases of the experiment against the difference in the exploration tendency of the participant and the algorithm. Most participants gain more in the second phase of the experiment with the help of recommendations, independent of the treatment condition. Panel B also illustrates that participants who explore the most also benefit the most from receiving recommendations (Pearson's $r = -0.27$, $t(440) = -5.86$, $p < .001$). As pointed out above, these participants are also the ones who exhibit the largest latent tendency ρ to follow the recommendations.

Result 8: Participants that have a high tendency to explore are most likely to follow the advice of an algorithm and also benefit most from receiving the advice.

4.3 Humans' Subjective Perceptions of the Algorithm

At the end of the experiment, we asked participants how useful they found the algorithm's advice, how much they relied on the algorithm and how much they followed the algorithm's advice. The answers to these three questions are sufficiently highly correlated (all $rs > .72$) that we have combined them into a single index of the perceived helpfulness of the algorithm. The subjective ratings are in line with the cognitive modeling results. The rating of the algorithm is positively correlated with participants' latent inclination to follow the algorithm ($r = .59$, $p < .001$). Participants also rated the helpfulness of the algorithm higher in the exploitative (M=.67, $SD = 0.21$) and balanced treatment (M=.65, $SD = 0.25$) than in the explorative treatment (M=.45, $SD = 0.29$, see Appendix A for more details).

Result 9: Subjective perceptions of the algorithm's usefulness correspond to the estimated inclination to accept its advice. They are in line with the findings from the cognitive modeling analysis as well as with the behavioral results.

5 Summary of Results and Discussion in Light of the Literature

5.1 Summary of Results

This study examines how human behavior in a stationary bandit task changes in response to receiving advice from algorithms with different exploration tendencies. In line with our expectations, we find that participants explore more when they obtain recommendations from a highly explorative algorithm than when the recommendations come from a balanced or more exploitative algorithm. In both the low-exploration and the balanced treatment conditions, participants explored less than in the control treatment, while participants in the control and in the high-exploration treatment conditions were on the same exploration level. Moreover, the participants exhibited the ability to use the algorithm's recommendation to enhance their performance: in all treatments with algorithmic advice, participants earned higher net rewards than in the control condition, while there was no significant difference among the treatments with recommendations. These results show that even in a probabilistic task, such as our multi-armed bandit, people are willing to accept advice from an algorithm and by doing so improve their performance. These results are in line with research

showing that when an algorithm performs well (Saragih and Morrison, 2022) and people have the opportunity to gain experience with it, they are willing to follow it (Filiz et al., 2021).

We also sought to test whether participants' willingness to follow the algorithm's advice depended on the exploration tendency of the algorithm and the similarity between it and their own exploration tendency. We hypothesized that the similarity of the advising algorithm and the participants regarding the inclination to explore impacts the usage of advice. This was operationalized in two ways: analyses of behavioral data and a cognitive modeling approach.

The two analyses reveal a similar pattern. On the behavioral level, we found that for those participants who individually would explore more than the algorithm would recommend, matches between participants' choices and the algorithm's suggestions increase with similarity. For those participants who would individually explore less than the algorithm recommends, there is no distinct relationship between switches and matches. Similarly, the cognitive modeling analyses show that participants' latent inclination to follow the algorithms' advice does not increase with similarity, but that participants are more willing to follow an algorithm's advice the more exploitative the algorithm is when compared to their own exploration tendency. While participants with a high exploration tendency often follow recommendations that are less explorative than they are, participants follow the advice when the algorithm suggests more exploration less frequently. Moreover, participants' subjective ratings of the algorithm's usefulness align with the behavioral and modeling analyses. Thus, presumably, following or not following algorithmic advice is a rather deliberate process.

In sum, we find no evidence that a similarity in exploration tendencies of the algorithm and the human advisee increases the willingness to follow the advice, but rather that the willingness to follow depends on the exploration preferences of the algorithm. Even though the quality of the advice, i.e., the algorithm's success rate, was equal, participants were only willing to frequently accept the advice when the algorithm was less explorative and thus gave more consistent advice over time. One reason could be that consistency is regarded as a signal of expertise (Falk and Zimmermann, 2017; Fehrler and Hughes, 2018), which is supported by participants' reports of a higher usefulness of the algorithm's advice in the exploitative and balanced treatments. Following this line of reasoning, participants may have hesitated to follow recommendations to explore because they interpreted exploration as randomness or a demonstrated lack of expertise. This line of reasoning confirms previous findings that humans lose their trust in algorithms rapidly once they observe algorithms committing errors (Dietvorst et al., 2015).

As pointed out above, most participants nevertheless enhanced their performance when they received algorithm recommendations. On average, participants in the explorative treatment also benefited from the advice. Our analyses show that this may be the case because even in the explorative condition some participants exhibited a stronger exploration tendency than the algorithm and were willing to follow it. Indeed, participants who individually explore extensively not only followed the recommendations the most, but also benefited the most from the recommendations in terms of

their earned net rewards. This shows that not only the exploration tendency of the algorithm but also the exploration tendency of the participant should be considered, and that highly explorative participants are the most likely to benefit from an algorithm's advice in tasks with exploration–exploitation trade-offs.

Lastly, our explorative analyses show that interacting with a human decision-maker can affect the behavior of the algorithm as the algorithm uses the participants' outcomes to inform its recommendations. The less explorative participants are in our setting, the stronger the algorithm's perceived bias toward giving explorative recommendations. While this effect is found in all three algorithm implementations, it is most prevalent for the high-exploration algorithm and least prevalent for the low-exploration algorithm. Continuously ignoring the algorithm's recommendation to explore provoked the algorithm to advise more exploration, which likely decreased participants' willingness to follow the algorithm even further. This should be taken into account in the design of algorithms for practical use cases.

5.2 Discussion in Light of the Literature

Our results inform two major debates in the current literature on taking advice from algorithms: (1) the debate on whether humans underutilize, rely too much on, or make appropriate use of algorithm recommendations, and (2) the debate on the importance of adviser–advisee similarity. On the practical side, our results also inform the development of algorithms to provide advice in decision support or recommender systems.

Although our approach does not include a control condition with human instead of algorithmic recommendations, it still informs the debate on algorithm aversion (Dietvorst et al., 2015; Dietvorst and Bharti, 2020; Logg et al., 2019). In their review on algorithm aversion, Jussupow et al. (2020) mainly identify the performance level of algorithms and information regarding their performance level as determinants of algorithm aversion or algorithm appreciation. In our setting, participants are biased toward the low-exploration algorithms, in relation to their individual tendencies, despite the fact that the low-exploration and high-exploration algorithms are equally good and participants obtain the same information regarding both. Hence, further investigations on algorithm aversion should take into account whether the algorithm under investigation is explorative or exploitative. If the asymmetry of participants' responses to low-exploration recommendations generalizes to different tasks and scenarios, it might systematically affect the exploration of teams of humans and algorithms: if humans rely on low-exploration recommendations while ignoring high-exploration recommendations, this will bias the tendency of human–algorithm teams away from exploration. This needs to be taken into account by the designers of such algorithms.

On the debate over the importance of similarity between adviser and the advisee, which has been mainly studied in settings where humans take advice from other humans (Yaniv, 2004), the question whether such findings generalize to taking advice from algorithms is gaining interest (Schemmer et al.,

2022; Pálfi et al., 2022; Himmelstein, 2022). Previous studies suggest that humans are more likely to accept advice from other humans that are similar to them or give similar advice (e.g., Gino et al., 2009; Minson et al., 2011; Tauni et al., 2019; Yaniv and Kleinberger, 2000). In the present study, we do not find any evidence that similarity in decision strategies increases willingness to accept an algorithm's advice. Future research is necessary to disentangle whether these findings stem from the nature of the advice giver, from the type of similarity considered, or from the sequential nature of our task as compared to one-shot judgment tasks. One indication of the importance of the task is Ihssen et al.'s (2016) finding that participants were more likely to follow a human with a more consistent strategy in a reversal learning task. This suggests that consistency is a key factor in participants' willingness to follow in tasks with sequential choices more generally.

Decision support systems often deal with well known, structured environments, for example, when clinicians diagnose diseases (Ganju et al., 2020). When applying decision support systems to situations where less information is available and trial-and-error learning might occur, such as in geological exploration or supply chain management, decision support systems are faced with the exploration–exploitation trade-off (Chaharsooghi et al., 2008). It has been shown that decision support systems can be especially beneficial for human performance in such less-predictable environments (Van Bruggen et al., 1998). However, our research indicates that it is not only important how the decision support system addresses this trade-off, but also how its chosen strategies account for the human factor. The concept of giving explorative recommendations has structural similarities to the concept of unexpected recommendations (Adamopoulos and Tuzhilin, 2014). Such recommendations that take the receivers' expectations into account have been shown to enhance the performance of recommendation systems. However, in our setup, we find the opposite effect.

Zhou et al. (2021) evidence that e-commerce product-search algorithms providing more refined, i.e., exploitative, recommendations increase the usefulness of the search algorithm. This is in line with our finding that recipients of algorithm recommendations value recommendations that are exploitative and consistent over time. Zhou et al. (2021) report that this effect is more pronounced for those people who already know what product they want. Exploratory consumers, on the other hand, engage more with the platform when search algorithms provide them with a broader range of recommendations. Similarly, in our research we can pinpoint the individual's exploration tendency as a reference point for the participants' perception of the algorithm. However, it is likely that our highly structured task gives rise to less curiosity and serendipity (Shani and Gunawardana, 2011) and therefore participants' exploration tendency and their interest in explorative recommendations might be lower than they would be in richer and less structured exploration settings.

6 Conclusion

We set out to study the effects of algorithmic advice, and the willingness of advisees to follow such advice, in a tightly controlled experimental exploration task. Our results show that participants benefit from the recommendations of algorithms in structured exploration–exploitation environments. When the algorithms are more exploitative than the advisee, the inclination to follow the algorithm increases, as does the advisee's subjective perception of the helpfulness of the algorithm. And while it is important to take the exploration tendency of the advisee into account, our research does not support the notion of a *per se* effect of similarity. Even though algorithmic recommendations cannot, and do not, always lead to good outcomes in every trial of the exploration of a probabilistic environment, most people are willing to accept the advice, unless it looks too explorative to them.

These results contributes novel insights to the rapidly expanding literature on how best to design algorithmic decision support systems (e.g., Adomavicius and Tuzhilin, 2005b; Panniello et al., 2016; Tauchert and Mesbah, 2019; Wang et al., 2022; Uotila et al., 2009). The importance of the consistency of advice relative to the advisee's own inclination toward exploration, which our results highlight, could be further investigated in future research on real customer recommendation systems, such as the one explored by Zhou et al. (2021).

References

Adamopoulos, P. and Tuzhilin, A. (2014). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(4):1–32.

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., and Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology*, 42(10):1931–1939.

Adomavicius, G., Huang, Z., and Tuzhilin, A. (2008). Personalization and recommender systems. In *State-of-the-Art Decision-Making Tools in the Information-Intensive Age*, pages 55–107. INFORMS.

Adomavicius, G. and Tuzhilin, A. (2005a). Personalization technologies: a process-oriented perspective. *Communications of the ACM*, 48(10):83–90.

Adomavicius, G. and Tuzhilin, A. (2005b). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749.

Brown, V. M., Hallquist, M. N., Frank, M. J., and Dombrovski, A. Y. (2022). Humans adaptively resolve the explore-exploit dilemma under cognitive constraints: evidence from a multi-armed bandit task. *Cognition*, 229:105233.

Capponi, A., Olafsson, S., and Zariphopoulou, T. (2022). Personalized robo-advising: Enhancing investment through client interaction. *Management Science*, 68(4):2485–2512.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).

Castells, P., Hurley, N., and Vargas, S. (2022). Novelty and diversity in recommender systems. In *Recommender systems handbook*, pages 603–646. Springer.

Chaharsooghi, S. K., Heydari, J., and Zegordi, S. H. (2008). A reinforcement learning model for supply chain ordering management: An application to the beer game. *Decision Support Systems*, 45(4):949–959.

Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F., and Peters, J. (2020). Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. *Elife*, 9:e51260.

Danwitz, L., Mathar, D., Smith, E., Tuzsus, D., and Peters, J. (2022). Parameter and model recovery of reinforcement learning models for restless bandit problems. *Computational Brain & Behavior*, pages 1–17.

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–879.

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

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

Ecken, P. and Pibernik, R. (2016). Hit or miss: What leads experts to take advice for long-term judgments? *Management Science*, 62(7):2002–2021.

Falk, A. and Zimmermann, F. (2017). Consistency as a signal of skills. *Management Science*, 63:2197–2210.

Fehrler, S. and Hughes, N. (2018). How transparency kills information aggregation: Theory and experiment. *American Economic Journal: Microeconomics*, 10.

Filiz, I., Judek, J. R., Lorenz, M., and Spiwoks, M. (2021). Reducing algorithm aversion through experience. *Journal of Behavioral and Experimental Finance*, 31:100524.

Ganju, K. K., Atasoy, H., McCullough, J., and Greenwood, B. (2020). The role of decision support systems in attenuating racial biases in healthcare delivery. *Management science*, 66(11):5171–5181.

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, 204:104394.

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278.

Gino, F., Shang, J., and Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational behavior and human decision processes*, 108(2):287–302.

Himmelstein, M. (2022). Decline, adopt or compromise? a dual hurdle model for advice utilization. *Journal of Mathematical Psychology*, 110:102695.

Hou, Y. T.-Y. and Jung, M. F. (2021). Who is the expert? reconciling algorithm aversion and algorithm appreciation in ai-supported decision making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–25.

Ihssen, N., Mussweiler, T., and Linden, D. E. (2016). Observing others stay or switch–how social prediction errors are integrated into reward reversal learning. *Cognition*, 153:19–32.

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

Kawaguchi, K. (2021). When will workers follow an algorithm? a field experiment with a retail business. *Management Science*, 67(3):1670–1695.

Laureiro-Martinez, D., Brusoni, S., Tata, A., and Zollo, M. (2019). The manager's notepad: Working memory, exploration, and performance. *Journal of Management Studies*, 56(8):1655–1682.

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

Madhavan, P. and Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, 8(4):277–301.

Mahmud, H., Islam, A. N., Ahmed, S. I., and Smolander, K. (2022). What influences algorithmic decision-making? a systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, 175:121390.

Meder, B., Wu, C. M., Schulz, E., and Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental science*, 24(4):e13095.

Minson, J. A., Liberman, V., and Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin*, 37(10):1325–1338.

Orth, D. A., Buchanan, M., Amresh, A., Smith, C., Lematta, G., Cooke, N., Fouse, A., and Dubrow, S. (2021). Designing for exploration and exploitation in experimental search and rescue scenarios. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 65, pages 720–725. SAGE Publications Sage CA: Los Angeles, CA.

Panniello, U., Gorgoglione, M., and Tuzhilin, A. (2016). Research note—in carss we trust: How context-aware recommendations affect customers' trust and other business performance measures of recommender systems. *Information Systems Research*, 27:182–196.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203.

Pouya, P. and Madni, A. (2021). Performing active search to locate indication of ancient water on mars: An online, probabilistic approach. In *ASCEND 2021*, page 4024.

Prahl, A. and Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, 36(6):691–702.

Pálfi, B., Arora, K., and Kostopoulou, O. (2022). Algorithm-based advice taking and clinical judgement: impact of advice distance and algorithm information. *Cognitive research: principles and implications*.

Saragih, M. and Morrison, B. W. (2022). The effect of past algorithmic performance and decision significance on algorithmic advice acceptance. *International Journal of Human–Computer Interaction*, 38(13):1228–1237.

Schemmer, M., Hemmer, P., Kühl, N., Benz, C., and Satzger, G. (2022). Should i follow ai-based advice? measuring appropriate reliance in human-ai decision-making. *arXiv preprint arXiv:2204.06916*.

Schoonderwoerd, T. A., van Zoelen, E. M., van den Bosch, K., and Neerincx, M. A. (2022). Design patterns for human-ai co-learning: A wizard-of-oz evaluation in an urban-search-and-rescue task. *International Journal of Human-Computer Studies*, 164:102831.

Schultze, T., Rakotoarisoa, A.-F., and Schulz-Hardt, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment & Decision Making*, 10(2).

Schulz, E. and Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current opinion in neurobiology*, 55:7–14.

Shani, G. and Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer.

Soll, J. B., Palley, A. B., and Rader, C. A. (2022). The bad thing about good advice: Understanding when and how advice exacerbates overconfidence. *Management Science*, 68(4):2949–2969.

Speekenbrink, M. (2022). Chasing unknown bandits: Uncertainty guidance in learning and decision making. *Current Directions in Psychological Science*, 31(5):419–427.

Speekenbrink, M. and Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in cognitive science*, 7(2):351–367.

Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.5.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tauchert, C. and Mesbah, N. (2019). Following the robot? investigating users' utilization of advice from robo advisors.

Tauni, M. Z., Memon, Z. A., Fang, H.-X., Jebran, K., and Ahsan, T. (2019). Influence of investor and advisor big five personality congruence on futures trading behavior. *Emerging Markets Finance and Trade*, 55(15):3615–3630.

Thompson, W. R. (1933). Biometrika trust on the likelihood that one unknown probability exceeds another in view of the evidence of two samples.

Uotila, J., Maula, M., Keil, T., and Zahra, S. A. (2009). Exploration, exploitation, and financial performance: Analysis of s&p 500 corporations. *Strategic management journal*, 30(2):221–231.

Van Bruggen, G. H., Smidts, A., and Wierenga, B. (1998). Improving decision making by means of a marketing decision support system. *Management Science*, 44(5):645–658.

Van Swol, L. M., Paik, J. E., and Prahl, A. (2018). Advice recipients: The psychology of advice utilization.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27(5):1413–1432.

von Helversen, B., Mata, R., Samanez-Larkin, G. R., and Wilke, A. (2018). Foraging, exploration, or search? on the (lack of) convergent validity between three behavioral paradigms. *Evolutionary Behavioral Sciences*, 12(3):152.

Wang, Q., Huang, Y., Jasin, S., and Singh, P. V. (2022). Algorithmic transparency with strategic users. *Management Science*.

Wiehler, A., Chakroun, K., and Peters, J. (2021). Attenuated directed exploration during reinforcement learning in gambling disorder. *Journal of Neuroscience*, 41(11):2512–2522.

Wilson, R. C., Bonawitz, E., Costa, V. D., and Ebitz, R. B. (2021). Balancing exploration and exploitation with information and randomization. *Current opinion in behavioral sciences*, 38:49–56.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., and Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, 143(6):2074.

Wu, C. M., Schulz, E., Pleskac, T. J., and Speekenbrink, M. (2022). Time pressure changes how people explore and respond to uncertainty. *Scientific reports*, 12(1):1–14.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational behavior and human decision processes*, 93(1):1–13.

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

Zajkowski, W. K., Kossut, M., and Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *Elife*, 6:e27430.

Zettler, I., Thielmann, I., Hilbig, B. E., and Moshagen, M. (2020). The nomological net of the hexaco model of personality: A large-scale meta-analytic investigation. *Perspectives on Psychological Science*, 15(3):723–760.

Zhou, W., Lin, M., Xiao, M., and Fang, L. (2021). Exploitation and exploration: Improving search precision on e-commerce platforms. *Available at SSRN 3762144*.

Appendix

A Results From Post-Experiment Questions

In this section, we analyze the answers from the post-experiment questions and their relationship to the other behavior measurements. The summary of descriptive statistics is shown in Table A1. We analyze how the reward earned from the participants and the number of switches between boxes correlates with the self-reported exploration and exploitation tendencies in Table A2. Table A3 and A4 report the regressions of the explorative and the exploitative tendencies on the number of switches and the reward.

Table A1: Summary of Post-Experiment Questions

treatment	explorative	exploitive	algorithm usefulness	algorithm relying	algorithm following	algorithm rating
control	.78(0.22)	.88(0.18)				
exploration	.81(0.16)	.90(0.11)	.50(0.31)	.45(0.31)	.54(0.29)	.45(0.29)
exploitation	.74(0.22)	.87(0.15)	.69(0.21)	.65(0.24)	.66(0.25)	.67(0.21)
balanced	.82(0.19)	.90(0.13)	.67(0.25)	.62(0.28)	.67(0.26)	.65(0.25)

Note: The descriptive statistics of the post-experiment questions for each treatment. The numbers outside brackets are the mean of the reported value, and the numbers within brackets are the standard deviation. The algorithm rating is calculated from the average of three algorithm ratings.

Table A2: Correlations between Reward, Switch, and Explorative and Exploitative Tendency

	reward	switch	exploitative
switch	-.51(.26)		
explorative	.08(.01)	.07(.01)	
exploitative	.17(.03)	.11(.01)	.55(.30)

Note: The correlation between the explorative and exploitative tendency from the post-experiment questions and the number of switches and the amount of reward across the experiment. The numbers outside brackets are the correlation between variables, and the numbers within brackets are the R^2.

As the ratings on the algorithm were highly correlated, the three variables have been combined into a single variable, *algorithm rating*. The rating is significantly different between treatments ($F(2, 392) = 18.40$, $p < .001$, partial η^2=.086).

Table A3: Regression Table of Post-Experiment Explorative and Exploitative Tendency on Reward

	Coefficient	SE	t-value	p-value
Intercept	53.08	2.16	24.48	.000
explorative	6.02	1.34	4.45	.000
exploitative	-2.73	0.97	-2.82	.005

Table A4: Regression Table of Post-Experiment Explorative and Exploitative Tendency on Switch

	Coefficient	SE	t-value	p-value
Intercept	0.57	0.14	4.21	.000
explorative	-0.34	0.08	-3.99	.000
exploitative	0.22	0.06	3.68	.000

Table A5: Correlations between Reported Algorithm Usefulness, Relying, Following, Combined Rating, and Estimated ρ

	algorithm usefulness	algorithm relying	algorithm following	algorithm rating
algorithm relying	.84(.71)			
algorithm following	.72(.52)	.82(.67)		
algorithm rating	.92(.85)	.96(.92)	.91(.82)	
ρ	.52(.27)	.60(.36)	.52(.27)	.59(.35)

Note: The correlation between the algorithm usefulness, relying, following, and rating from the post-experiment questions and the estimated ρ. The numbers outside brackets are the correlation between variables, and the numbers within brackets are the R^2.

Table A6: Tukey HSD on Rating of Algorithm

treatments		mean difference	p-value	95% lower bound	95% upper bound
exploration	exploitation	-0.17	.000	-0.24	-0.10
	balanced	-0.15	.000	-0.22	-0.08
exploitation	balanced	0.02	.813	-0.05	0.09

B Behavioral Results

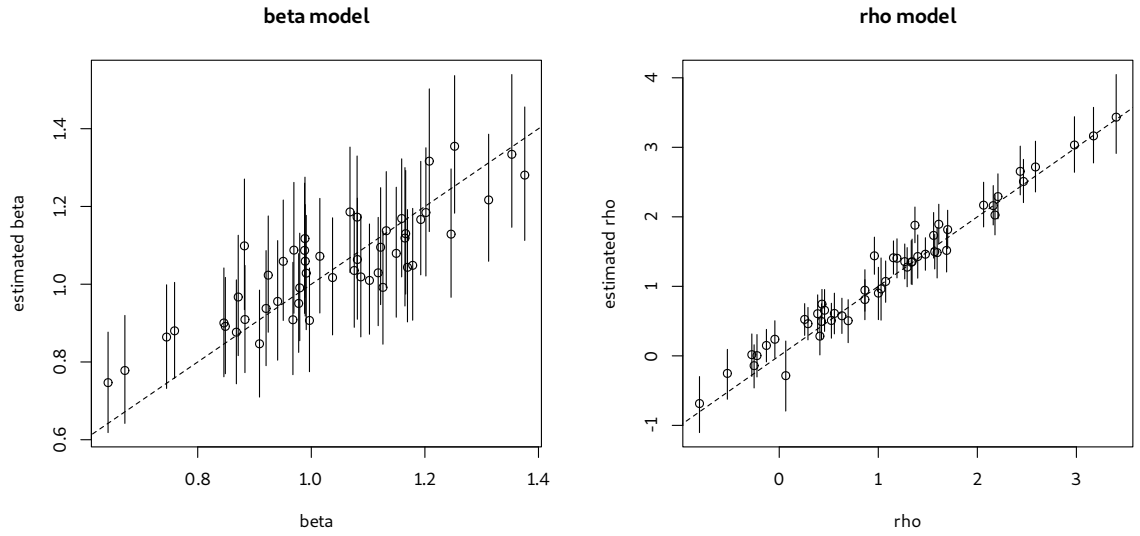
Table B1: Regressions of Matching on the Difference in the Exploration Tendency between the Participants and the Algorithm

	all			exploitation			balanced			exploration		
	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
constant	131.70	4.93	$< .001$	173.20	9.81	$< .001$	155.90	4.53	$< .001$	98.14	13.77	$< .001$
difference	0.38	0.02	$< .001$	0.23	0.07	.002	0.41	0.04	$< .001$	0.43	0.07	$< .001$
difference^2	0.00	0.00	$< .001$	0.00	0.00	.097	0.00	0.00	$< .001$	0.00	0.00	.024
exploitation	25.10	5.17	$< .001$	-	-	-	-	-	-	-	-	-
balanced	22.77	4.91	$< .001$	-	-	-	-	-	-	-	-	-
N	442			146			148			148		
R^2	0.55			0.09			0.40			0.46		

Note: Linear regressions of number of matching in phase 2 on the difference in the exploration tendency between the participants and the algorithm. Difference in the exploration tendency is the number of switches by a participant in phase 1 minus the number of recommended switches in phase 2 from the algorithm. The variable difference^2 is squared difference. The "al" regression uses all data from the non-control treatments and the other three regressions use data from individual treatments. Treatment exploration is the base group for treatments exploration, exploitation and balanced. All models also include age and gender as socio-demographic control variables.

C Parameter Recovery

Figure C1: Parameter Recovery



Note: Results of a parameter recovery simulation for the two models. Scatter plots of true individual parameters (x-axis) versus estimated individual parameters (y-axis). Whiskers indicate 95% highest density credibility intervals for each parameter estimate.

D Supplementary Results: Cognitive Modeling Analyses

Table D1: Summary of Fitted Models

Model	Treatment	Tasks	Mean	Std	min(neff)	max(Rhat)	div	max(tree)	LOO
β	control	1-3	0.58	0.26	1218	1.0044	0	4	-46355
β	control	4-6	0.89	0.26	994	1.0012	0	4	-37882
β	exploitation	1-3	0.47	0.30	960	1.0039	0	4	-47161
β	exploitation	4-6	1.32	0.23	1901	1.0024	0	4	-24789
ρ	exploitation	4-6	3.55	1.38	480	1.0123	0	6	-23273
β	balanced	1-3	0.61	0.28	1165	1.0042	0	4	-45252
β	balanced	4-6	1.17	0.18	2665	1.0004	0	4	-28524
ρ	balanced	4-6	2.22	1.44	330	1.0107	0	5	-24136
β	exploration	1-3	0.42	0.28	997	1.0049	0	4	-50562
β	exploration	4-6	0.89	0.19	2343	1.0009	0	4	-38641
ρ	exploration	4-6	1.97	1.98	162	1.0494	0	6	-29862

Note: Population mean and standard deviation of the model parameter for each fitted model. The remaining columns report the minimum effective sample size, maximum Rhat, number of divergent transitions and the maximum tree-depth for each model. LOO is an estimate for the out-of-sample prediction accuracy of a Bayesian Model (Vehtari et al., 2017).

Table D2: Regression Models for Inclination to Follow Recommendations (ρ)

	all			exploitation			balanced			exploration		
	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
constant	0.03	0.32	.931	0.86	0.67	.202	0.21	0.29	.476	-1.63	0.90	.071
difference	0.29	0.31	.347	0.43	0.49	.384	-0.09	0.35	.796	0.36	0.68	.595
difference^2	1.30	0.12	$< .001$	1.47	0.18	$< .001$	0.78	0.14	$< .001$	1.50	0.27	$< .001$
low explore	0.49	0.25	.052	-	-	-	-	-	-	-	-	-
optimal	-0.34	0.25	.183	-	-	-	-	-	-	-	-	-
N	442			146			148			148		
R^2	0.56			0.69			0.50			0.47		

Note: Linear regression of the latent tendency to follow algorithm recommendations on the difference and the squared difference between the participant's exploration tendency and the algorithm's true exploration tendency $log(\beta_1)$–$log(\beta_{exo})$. Models for the data of all treatments with age and gender as socio-demographic control variables.

E Post-hoc Tests

This section summarizes post-hoc pairwise comparisons between the treatments using Tukey HSD corrections in phase 2.

Participants are more explorative in the exploration and control treatments than in the exploitation and balanced treatments. Between the control and the exploration treatment, and between the exploitation and the balanced treatment, no significant difference could be detected. Regarding rewards, participants in the exploitation treatment earn on average more than subjects in the control treatment. The number of choices that match with the algorithm's recommendation is higher in the balanced and exploitation treatments than in the exploration treatment.

Table E1: Tukey HSD on Average Number of Switches

treatments		mean difference	p-value	95% lower bound	95% upper bound
exploitation	control	-7.72	$< .001$	-11.93	-3.51
	balanced	-2.12	0.559	-6.31	2.06
	exploration	-11.39	$< .001$	-15.58	-7.20
balanced	control	-5.59	0.004	-9.79	-1.40
	exploitation	2.21	0.559	-2.06	6.31
	exploration	-9.27	$< .001$	-13.44	-5.09
exploration	control	3.68	0.110	-0.52	7.87
	exploitation	11.39	$< .001$	3.51	11.93
	balanced	9.27	$< .001$	5.09	13.44

Table E2: Tukey HSD on Average Rewards

treatments		mean difference	p-value	95% lower bound	95% upper bound
exploitation	control	1.58	$< .001$	0.38	2.77
	balanced	-0.41	0.814	-1.59	0.78
	exploration	0.46	0.749	-0.73	1.65
balanced	control	1.98	0.004	0.79	3.17
	exploitation	0.41	0.814	-0.78	1.59
	exploration	0.87	0.233	-0.31	2.05
exploration	control	1.12	0.074	-0.07	2.30
	exploitation	-0.46	0.749	-1.65	0.73
	balanced	-0.87	0.233	-2.05	0.31

Table E3: Tukey HSD on Number of Matches

treatments		mean difference	p-value	95% lower bound	95% upper bound
exploitation	balanced	2.98	0.252	-1.44	7.40
	exploration	22.17	$< .001$	17.75	26.59
balanced	exploitation	-2.98	0.252	-7.40	1.44
	exploration	19.19	$< .001$	14.78	23.59

F Changes Over Tasks

Figure F1 shows changes over tasks. In general, average switches and entropy decrease over tasks and average rewards increase over tasks. The exploitation treatment and the balanced treatment have a higher decrease in switches in comparison to the other two treatments.

Figure F1: Average Number of Switches, Average Rewards and Entropy over Tasks

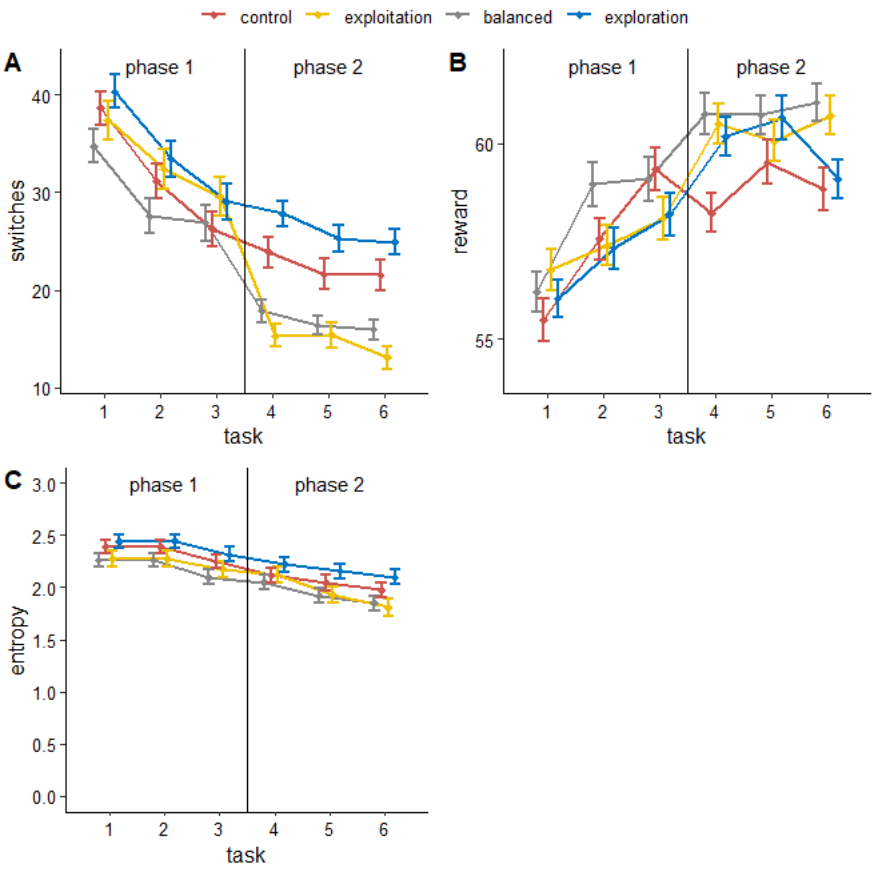
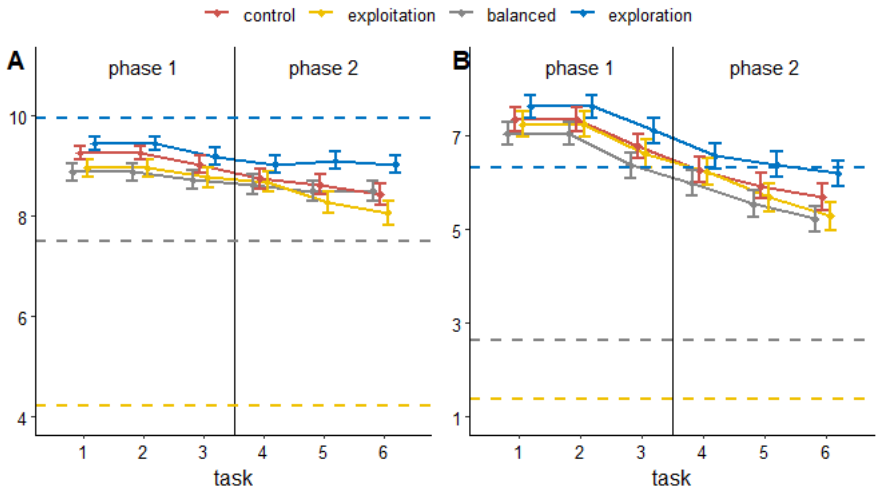


Figure F2 shows how the average number of bandits explored changes over tasks. Panel A includes all trials, while Panel B includes the last 50 trials when participants' behavior is stabilized. We simulate with the three exploration parameters $\phi = 0.4$, $\phi = 1.4$ and $\phi = 3.3$ and calculate the average number of bandits that the algorithm would explore when it is exploitative, balanced, and

explorative. The dotted lines in the figure are the simulation results. For all trials, participants explore on average more bandits than a balanced algorithm and an exploitative algorithm, but they explore less than an explorative algorithm. Looking at the last 50 trials, participants are more explorative than an explorative algorithm in phase 1. Participants in treatments other than exploration explore fewer bandits than an explorative algorithm in phase 2.

Figure F2: Average Number of Bandits Explored.

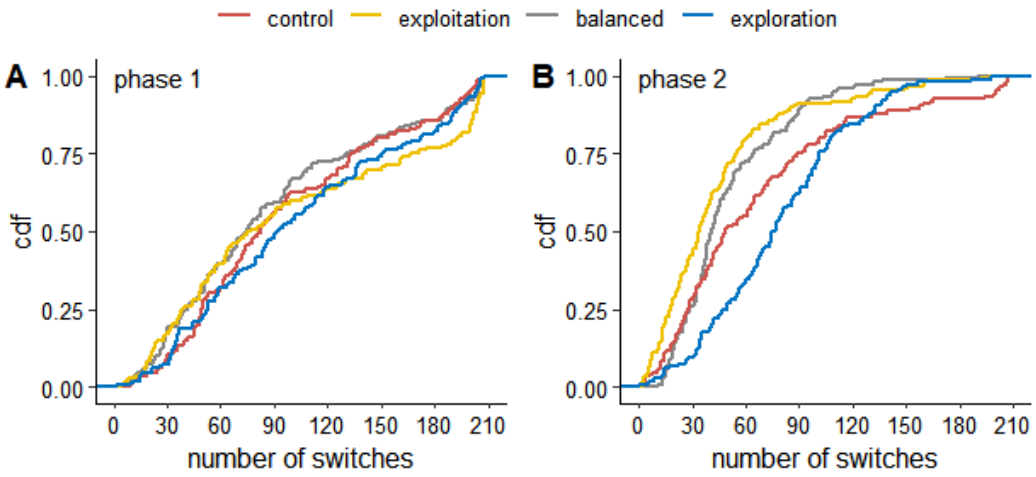


Note: Panel A includes all the trials. Panel B includes only the last 50 trials. The dotted lines are the average number of bandits explored by algorithm based on simulation.

G Empirical CDF

The two panels in Figure G1 are empirical CDF for all treatments in phase 1 and phase 2. There is not much difference in phase 1. In phase 2, participants in the exploration and control treatments are less explorative than those in the exploitation and balanced treatments. The control treatment first order statistically dominates the exploitation treatment.

Figure G1: Empirical Cumulative Distribution Curve



H Instructions for the Experiment

H.1 Consent Form

Dear Participant:

This is an academic study undertaken by the Faculty of Psychology at the University of Bremen, Germany. It seeks to advance knowledge about how people make decisions and how they learn to decide. Participation in this study is entirely voluntary, and you may cease participation at any time. If you decide to participate in this study, you must affirm that you understand the terms below and consent to participate. However, receiving compensation requires that you complete the experiment in its entirety.

The study takes about 25 minutes on average. You will see ten boxes on the screen and have to pick one of the boxes. Each box contains a different number of points. Your goal is to maximise the number of accumulated points. As compensation, you will receive a base payment of £2.5 plus an additional payment depending on how many points you accumulate through the choices you make. The amount of this additional payment can be between £0 and £11. However, receiving £11 is very unlikely. If you always choose the best or the second-best option, you will in most cases earn a bonus of about £3. At the beginning of this study, you will be asked to enter your Prolific ID, which is required to receive the payment. Except for your Prolific ID, no information will be tracked or stored that could be used to reveal your identity. The decisions you make and your answers during the study are pseudonymised. After you receive the payment, we will delete your Prolific ID from the data. From that point forward, we will no longer be able to identify your data or delete your data on request. We will use the collected data in the course of academic communication, which includes making it publicly available on the platform of the Open Science Framework (osf.io). There are no known physical, mental, social, or legal risks involved in the study. We kindly ask that you give the tasks your full attention and exit the full-screen mode only after completing the study. Please make sure that your participation is free from any distractions (e.g., mobile phone, television, music).

When you participate, you affirm that:

1. You will solve the tasks diligently and without the help of any tools.

2. You are over 18 years of age.

3. You understand that your participation is voluntary and that you may refuse to participate in this study or discontinue your participation at any point. Receiving the payment, however, is contingent on finishing the study.

4. You agree to the use of your data for the purpose of research and it being published anonymously for academic communication.

5. If you would like to receive a complete description and rationale for this study, or if you have questions regarding any concerns that arise from participating in this study, please direct your request to hslin@uni-bremen.de.

6. If you have any ethical concerns, please direct them to pbanik@uni-bremen.de.

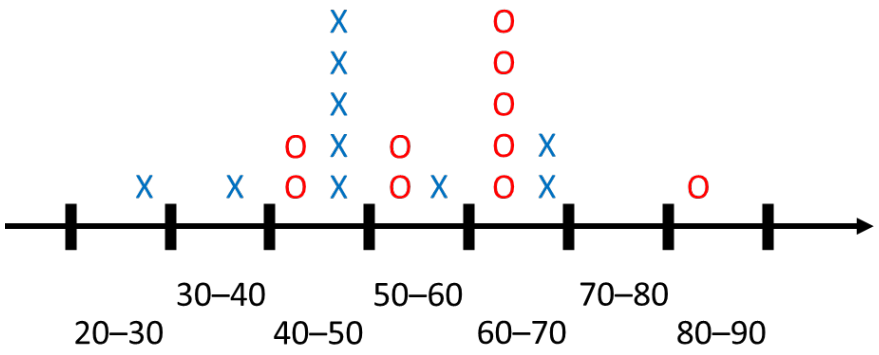
If you agree with these terms, please select 'continue' to proceed.

H.2 Instruction

To earn the highest possible amount during this study please carefully read this introductions page. Approximately half of your payment depends on your individual decisions you make later.

This experiment consists of two sessions, each consisting of a training block followed by three real decision-making blocks. In each block you will be asked to make repeated choices between ten boxes that are presented on the screen. In each instance you may choose one box to be opened by clicking on it. Once the box is chosen, a message below the box will show the number of points the box contained. Your goal is to maximise the number of points (i.e., the reward you receive from a box) over all decision-making blocks.

Each time you open a box, the reward that you receive may differ. Each box generates the reward from a fixed average in each block. This average reward is randomly selected from a normal distribution with a mean of 50 and standard deviation of 10. Each time a box is clicked, the reward is drawn from a normal distribution with the mean of the average reward of the clicked box and a standard deviation of 10. This means that, if you select a box with an average reward of 50 multiple times, the average of the generated reward will be close to 50. Half of the generated rewards will be between 43 and 57 points. For example, imagine you click ten times on a high-paying box (red circle, with an average reward of 60) and ten times on a low-paying box (blue cross, with an average reward of 45). The high-paying box may generate the rewards 60, 41, 81, 51, 69, 42, 54, 60, 60, 61, and the bad box may generate the rewards 48, 37, 49, 60, 29, 44, 65, 45, 48, 54. Thus, as the figure below illustrates, the low-paying box may sometimes result in rewards that are higher than the high-paying box. On average, however, you will earn a higher reward when clicking on the high-paying box.



During each block, the program will display the status of the current block on the top of the screen. It contains the number of choices you have made, the total amount of choices you can make in the current block, and the number of points you have earned in the current block.

Between each choice, you have unlimited time to decide which box you want to click on. Please carefully consider your choices.

For participating, you will receive a base payment of £2.5. In addition, you will receive a payment that depends on your choices during the experiment.

The additional payment is based on the average points in the boxes you opened across the six actual decision-making blocks. If you choose randomly and earn on average 45 points, you will receive no additional compensation. For every one point above 45, you will receive a bonus of 20 pence. For example, if you scored on average 55 points per trial, you will receive a bonus payment of £2. This bonus is capped at 100 points, i.e., you can receive a maximum of £11 in addition to the base payment of £2.5. However, it is unlikely that you reach a point average of 100 points. Typically, if you always choose the best or the second-best box available, you will likely reach a point average of about 60, which would result in a bonus payment of £3.

If you're ready, please select 'continue' to proceed.

H.3 Control Questions

To ensure you fully understand the instructions, please answer the following questions by selecting your answers with your mouse:

1. What's your goal in the task?
    a. Get as many points as possible.
    b. Explore as many boxes as possible.
    c. Choose a box as quickly as possible.
2. Which statement about the number of points you receive when choosing a box repeatedly is correct?
    a. The number of points in each box changes from trial to trial.
    b. Each box always gives the same number of points.
3. How should you achieve higher points in the experiment?
    a. By clicking the high-paying box.
    b. By clicking random boxes.
4. How is your payment determined?
    a. I receive only a base payment of £2.5.
    b. My payment depends on my choices. In addition to the base payment, I will receive a payment that depends on the average number of points I collect from the boxes.

Once you have finished answering the questions, please press 'continue' to confirm your answers.

H.4 Instruction Before First Training Block

The next block is a training block. Please familiarise yourself with the procedure. The points acquired during the training block do not count toward your final reward.

If you're ready for the training block, please press 'continue' to proceed.

H.5 Instruction After Training Block

The training block is over.

All the boxes will be reset to random states. You cannot carry over what you learned to the next phase.

If you're ready, please select 'continue' to proceed with the decision-making block.

H.6 Instruction After Each Testing Block

In this block, you acquired xxx points per trial!

All the boxes will be reset to random states. You cannot carry over what you learned to the next block.

If you're ready, please select 'continue' to proceed.

H.7 Instruction Before Second Training Block

H.7.1 If Participants Are In the Control Treatment.

The next session will be the same as the previous one. You will first see a training block and then three actual decision-making blocks. Please remember that the points acquired during the training block do not count toward your final reward.

If you're ready for the training block, please press 'continue' to proceed.

H.7.2 If Participants Are In the Algorithm Recommendation Treatments.

The next session follows the same general procedure as the previous session. However, in these blocks you will receive recommendations from a cutting-edge artificial intelligence (AI) algorithm, which will suggest which box to open next for each trial. The AI makes the recommendations based on the same information that you have. Thus, it knows about the general distribution of rewards in the boxes and it will learn to make better recommendations based on the information it receives by observing your choices within a block. The AI does not know the rewards from the unopened boxes.

The AI will suggest to you which box to open next by presenting a yellow square around the box. In each trial, you can choose whether you want to follow the suggestion from the AI or choose a box on your own. You will first be confronted with a training block and then three actual decision-making

blocks. Please familiarise yourself with the procedure. The points acquired during the training block do not count toward your final reward.

If you understand the newly added AI and you're ready, please press 'continue' to proceed.

H.8 Instruction After the Second Phase

The testing is now over. Across the six decision-making blocks, you acquired xxx points per trial, which is converted to £yyy additional payment.

Before we finish the experiment, we would like to ask you a few questions.

If you're ready, please select 'continue' to proceed with the questions.

H.9 Questions

H.9.1 Effort Question.

Please give us your honest and subjective assessment of how much effort you put into maximising your profit. Your answer will NOT have any negative consequences. Please respond by clicking on the slider below.

H.9.2 Data Quality.

Is there any reason we should doubt your data quality (for example, because you were distracted or were not concentrating)? Please answer honestly; your answer will NOT affect the amount of reward you receive or have any negative consequence.

My data may be flawed.

My data is fine.

H.9.3 Exploitation Tendency.

How much did you try to keep updated about how good all the boxes are? Please respond by clicking on the slider below.

H.9.4 Exploration Tendency.

How much did you try to choose the box in each trial with the highest average profit? Please respond by clicking on the slider below.

H.9.5 Algorithm Usefulness.

How useful were the suggestions from the AI? Please respond by clicking on the slider below.

H.9.6 Relying On the Algorithm.

How much did you rely on the suggestion from the AI in general? Please respond by clicking on the slider below.

H.9.7 Following the Algorithm.

How much did you follow the suggestions from the AI when you were unsure about which box to choose next? Please respond by clicking on the slider below.

H.9.8 Further Comments.

If you have any comments or remarks about the study, you can leave them anonymously in the box. If not, you may leave the textbox empty. Once you have finished, please press 'continue' to proceed.