Graphical Abstract

When to Stop Searching in a Highly Uncertain World?
A Theoretical and Experimental Investigation of "Two-way" Sequential Search Tasks

Imen Bouhlel, Michela Chessa, Agnès Festré, Eric Guerci

Highlights

When to Stop Searching in a Highly Uncertain World?
A Theoretical and Experimental Investigation of "Two-way" Sequential Search Tasks

Imen Bouhlel, Michela Chessa, Agnès Festré, Eric Guerci

- In a highly uncertain world an exploration-exploitation trade-off combines with an accuracy-optimality one

- Experimental data supports optimal learning about when to stop in a search task

- High uncertainty disrupts the learning process

- Oversearch occurs when exploration is more costly

- Undersearch occurs when exploration is cheaper

When to Stop Searching in a Highly Uncertain World? A Theoretical and Experimental Investigation of "Two-way" Sequential Search Tasks

Imen Bouhlel[a], Michela Chessa[a], Agnès Festré[a], Eric Guerci[a]

[a]*Université Côte d'Azur, CNRS, Groupe de Recherche en Droit, Economie et Gestion, Nice, France,*

Abstract

When to stop exploring is crucial in contexts where learning to manage time and uncertainty is critical for carrying out successful initiatives (innovation, personnel recruitment, vaccine discovery, etc.). We investigate analytically and experimentally the exploration-exploitation trade-offs in such contexts. A "two-way" sequential search task is proposed where the classical exploration-exploitation trade-off in sequential decisions with finite-horizon is coupled with a further one about discovering the real value of each alternative. The longer the time spent for discovering an alternative, the higher the certainty but at a higher cost. Oversearch occurs when exploration is more costly, and undersearch when exploration is cheap. People learn better to stop the more certain the information is. A potential behavioral trap in the exploration of "two-way" search tasks is identified that brings towards local optima. We recommend thus policies that force people to reduce the time spent exploring the alternatives.

Keywords: optimal-stopping, exploration-exploitation trade-off, regret, experiment

1. Introduction

Knowing when to stop searching has become a key asset in our digital economies where the Web has transformed in the last two decades many experience goods such as finding a restaurant or even a mate into search goods (Klein, 1998). The more recent phenomenon of "Tinder-ization" (from the name of the well-known online dating application) has rendered the issue

even more pervasive, not to talk about its addictive potential due to instant connection, global reach, geo-locatization, visual interaction, likes and feedback, integration with Instagram and Spotify etc.

This question has been addressed by the literature in a twofold manner. On one hand, the optimal stopping problems literature approaches it as an exploration-exploitation issue (Chow et al., 1971; March, 1991; Gupta et al., 2006; Ferguson, 2007), that is, a trade-off between exploiting a safe known option/alternative, which may be sub-optimal, and exploring a new risky one, which may be much better, but may also be very unprofitable. The emblematic example of this literature is the so-called *Secretary problem* (Ferguson et al., 1989), which considers a context of sequential search where a manager needs to find the best applicant for a secretarial job[1]. On the other hand, the *multi-armed bandit* literature (Robbins, 1952) views the question from a different angle by accounting for an uncertainty in the value of each option/arm which can be reduced by the participant by spending more time in exploring such option. Our contribution is at the crossroad of these two strands of the literature[2]. More precisely, we extend the current literature on the optimal stopping problem in sequential search tasks in a direction which, to the best of our knowledge, has not yet been investigated. Is it optimal to explore more alternatives at the risk of higher uncertainty about their values in order to make an informed decision over a larger set of alternatives? What do people actually do? These two questions characterize our research work. Information gathering on alternatives/options in order to obtain a precise estimation of the alternative's value might be costly in real-world problems. The acquisition of more information is usually carried out at the cost of a reduced exploitation of potential gains. In our model, we emulate this realistic scenario by considering what we call *iterative exploitation* for determining the reward. The payoff depends not only on the value of the chosen alternative,

---

[1]The *Secretary problem* has become a field of study which has intrigued among others operational researchers and computational scientists as well as management and economics scholars for its potential real-world implications, namely, choosing an employee, an apartment (Zwick et al., 2003), a mate (Todd, 1997; Todd and Miller, 1999), or a restaurant, as well as dealing with a growing inventory (Hochman, 1973) or optimizing R&D and investment problems, through extensions of the original model. As observed by Hills et al. (2015), diverse domains have worked on exploration-exploitation issues largely in isolation (e.g., studies on spatial foraging), while the most recent trends across disciplines indicate that the formal properties of these problems share similar structures and similar solutions.

[2]See also the *infinitely many armed-bandits* literature (Berry et al., 1997).

but also on the remaining time that has not been used for exploration, that is, the value of the chosen alternative times the number of remaining trials. We propose thus in our model to include and investigate the exploration-exploitation trade-off between acquiring more information on the real value of the alternative at the cost of losing time for the reward exploitation of the alternative. We call our model "two-way" sequential search task, because we have two exploration-exploitation trade-offs. We thus differentiate it from the more classical "one-way" sequential version of the search task where the first exploration-exploitation trade-off is absent because the value of the alternative is certain.

In the "one-way" treatment we find an oversearching tendency in line with some studies of the literature Eriksson and Strimling (2010); Sandri et al. (2010); Juni et al. (2016); Sang et al. (2020). In particular, Eriksson and Strimling (2010); Sang et al. (2020) propose the most similar version of our "one-way" task, because the payoff includes the exploitation phase and allows for recall. Considering the vast "à la Secretary" literature, an undersearch tendency is mainly observed (Hey, 1987; Seale and Rapoport, 1997; Sonnemans, 1998; Seale and Rapoport, 2000; Bearden et al., 2006; Schunk and Winter, 2009; Oprea et al., 2009; Costa and Averbeck, 2013). Clear-cut conclusions have not been provided as (Zwick et al., 2003; Descamps et al., 2022) found both effects of under- and oversearch depending on the exploration costs.

Theoretical findings of our sequential search tasks suggest the existence of an optimal stopping threshold on the quality of an explored alternative in order for such alternative to be accepted. This threshold decreases with time, determining that it is optimal to become less exigent the more alternatives are explored. In the "two-way", the time spent to investigate the quality of a single alternative should always be as short as possible for a risk neutral agent, and it may become higher in case of risk aversion. But, in the "two-way" we provide experimental evidence that people tend to oversearch on the "first" exploration-exploitation phase, thus spending too much time on reducing the uncertainty on the value of the alternative at the expense of exploring more options. This latter aspect is found to encourage an undersearching tendency in the "second" exploration-exploitation trade-off unlike the "one-way". Furthermore, we find that people learn better when to stop in the "one-way" treatment where the information is certain. Therefore, "two-way" search tasks are behaviorally more challenging.

In our paper, we further develop our analysis by providing insights into

3

why policy recommendations encouraging people to reduce the time spent on gaining information about one specific alternative should be favoured. Indeed, we propose that many real-world problems can be framed in terms of "two-way" sequential search tasks. Starting from the originally inspiring problem of the literature, recruiting the "Secretary" remains a crucial example in which speed is essential for winning the war for talent, in particular in such a competitive environment. A strong feeling about the candidate is often crucial for the hiring process, more than a deep investigation about her/his skills. This "Tinder-ization" effect is strongly encouraged by digital environments, in particular on mobile devices, in which search mechanisms are often implemented via tap and swipe gestures to mimic "two-way" sequential search tasks. A tap implements the choice of getting more information, while a swipe that of passing to the following option. The speed of swiping provides a superficial hint of the available choices when facing huge sets of alternatives, as in the digital ecosystem.

Our paper is organized as follows[3]. In Section 2, we present the analytical model as a class of sequential stopping rule problems, and we derive the analytical solutions to identify optimal decisions. In Section 3, we present the experimental protocol and conditions. Finally, in Section 4, we compare the theoretical predictions to the experimentally observed behavior and study the learning dynamics towards those predictions. We find in line with the literature, a tendency to oversearch when exploration is costly, and a tendency to undersearch when exploration is relatively cheap, as well as a more pronounced learning effect under certainty (referred as Treatment One-way in Section 3), that is when the exploration of a new alternative is more costly. We further investigate the stopping decision using a survival analysis in order to account for behavioral measures of regret and anticipation. Our main findings are partially in contrast with the literature (Zwick et al., 2003), as we find that anticipation leads to less exploration, while regret leads to more exploration.

---

[3]In Section Supplementary Material at the end of the paper, we provide the link to a rich Online Appendix. There you can find further analytical and experimental details and results.

2. Theoretical model

The theory of optimal stopping is concerned with the problem of choosing the time at which to take a given action in order to maximize an expected payoff, when faced with a sequence of random variables. Within this general formulation, we propose and study two models adopting the following assumptions. First of all, we consider a stopping rule problem with *finite horizon T*, i.e., with a known upper bound on the number of stages (alternatives from now on) at which one may stop. Secondly, we implement a classical exploration-exploitation trade-off in which, once an agent has decided to start exploiting, s/he cannot go back exploring. Thirdly, we assume that it is always possible to recall a past alternative to exploit its value and we do not discount this value. Finally, we do not count in the reward function the past alternatives of the exploration phase. We start enjoying a strictly positive payoff only when exploiting. Such assumption is typical of situations, such as the secretary problem, where we obtain utility only after her/his recruitment.

We note that with this formalization of the stopping problem, the implementation of the optimal stopping rule is conditional on the observed realization of the random variables. We report here the main theoretical findings. All proofs may be found in the Online Appendix A (see Section Supplementary material).

2.1. Model under certainty (MC)

The first model, referred to as MC, considers both the time horizon T and the number of alternatives n to be finite and equal, $n = T$. We will use the variable n to denote both quantities. Suppose that the the random variables X_i, with $i = 1, \ldots, n$, are i.i.d., $X_i \sim \mathcal{U}(\{1, \ldots, N\})$. The value x_i represents the intrinsic quality of alternative i, which is independently drawn from the discrete uniform distribution between a minimum of 1 and a maximum of N. This intrinsic value is fixed and is known with certainty by the agent if s/he explores it. At each $t \in \{1, \ldots, n-1\}$, the agent has already observed the quality of the first t alternatives and s/he decides whether to exploit one of them for the remaining time, or to explore the new alternative $t+1$. For each t, if the participant decides to exploit the current or a past alternative the decision is irreversible and s/he will not discover the values of the remaining $n - t$ alternatives.

The participant can exploit the best alternative found so far, times the remaining time horizon $n-t$. The strategy of exploiting the maximum known value dominates all other choices of exploitation. The optimal exploration-exploitation trade-off can be investigated by the method of backward induction. For "large" N, we may write the optimal stopping threshold as

$$\bar{m}_t \sim \frac{n-t-\sqrt{2n-2t-1}}{n-t-1}N. \tag{1}$$

The quantity $\bar{m}_t$ represents the threshold below which it is convenient to keep on exploring and above which it is convenient to start exploiting.

2.2. Model under uncertainty (MU)

We also propose and study a second optimal stopping rule model, referred to as MU, which differs from the previous one only in one relevant aspect, namely, the fact that each alternative needs to be repeatedly "explored" in order to gain information about its true value. The agent can benefit from this opportunity because s/he can reduce her/his time spent in one alternative either for exploring more alternatives or for exploiting it for a longer time. Therefore, the agent faces now two distinct explorations and associated trade-offs. The first one, as in the previous model, is about choosing the right alternative at which to stop the exploration phase. In the second one, during each alternative, the agent must decide how long the exploration time should be and consequently how precise s/he wants the information of its true value to be. In many real-world examples, we get an approximate estimation of the value of an alternative by exploring or testing it repeatedly (e.g., testing the same restaurant multiple times).

Formally, in the MU model the agent still has a maximum number of alternatives to explore equal to n and a time horizon of $T = n$, but s/he is not obliged at alternative i to spend a whole unit of time on this alternative in order to discover its real value x_i. Each alternative is divided into s sub-units of time and in each sub-unit of time s/he observes a certain number of normally distributed values $y_i = x_i + \epsilon_i$, where ϵ_i are realizations of some iid zero-mean random variables with finite variance σ^2. After a total of k observations of alternative i, the agent can estimate its intrinsic value computing the mean $\sum_{i=1}^{k} \frac{y_i}{k}$ with standard deviation $\sigma_M = \frac{\sigma}{\sqrt{k}}$. In particular, in our implementation we suppose that each sub-unit of exploration time on one alternative is twice as informative as the previous one exploring the same alternative.

At first, we consider a risk-neutral agent. We observe that, whenever deciding to explore an alternative l after the exploration of t alternatives for $h_1, \ldots, h_t$ sub-units respectively, with $h_i > 0$ for at least one $i = 1, \ldots, t$, such a strategy is dominated by the choice of exploiting alternative l after the exploration of t alternatives for 1 sub-unit each. Then, a priori for the agent it is optimal to decide to spend the minimum amount of time in exploring each alternative. Analogously to before and by backward induction, when N is "large", the threshold below which it is optimal to draw is

$$\bar{m}_t \sim \frac{ns - t - \sqrt{2ns - 2t - 1}}{ns - t - 1} N, \tag{2}$$

which is analogous to the threshold in (1), when the agent has already consumed t over a total of ns sub-units.

Now we take into account the risk-aversion of an agent, by assuming the possibility that s/he will always explore alternatives for a fixed number of sub-units in order to reduce the uncertainty till a given level of tolerance. Let $r = 1, \ldots, s$ be the parameter which defines the risk-aversion of an agent, where $r = 1$ corresponds to risk-neutrality, while $r = s$ corresponds to maximal risk-aversion. Intermediate levels of risk aversion provide an analogous threshold to (1) and to (2), which is now given by

$$\bar{m}_t^r \sim \frac{ns - rt - \sqrt{2ns - 2rt - 1}}{ns - rt - 1} N. \tag{3}$$

Observe how such a threshold is decreasing in r, meaning that the higher the risk aversion of an individual, the smaller the threshold for exploiting an alternative. As a consequence, risk-averse agents will, on average, explore less.
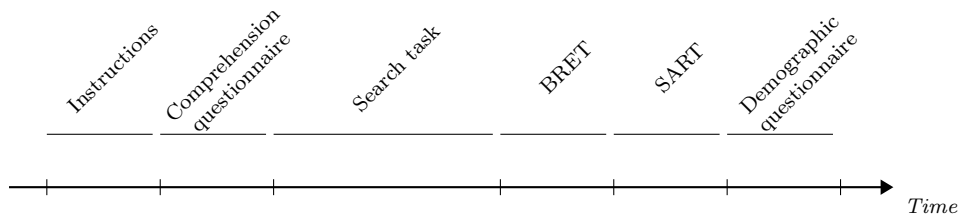
3. The Experiment

The two analytical MC and MU models presented in the previous section were conceived and investigated with a view to running an experiment in a laboratory with students to validate them empirically.
We recruited 77 participants of whom 46 were female and 48 were in the age group 20 to 25 years old. Participants were paid €5 for showing up to the experiment and a performance contingent bonus of €18 on average after

spending about one and a half hours in the lab. Participants were recruited using ORSEE (Greiner, 2015). Four experimental sessions were run with 15 to 23 subjects, and took place on September 26, October 24 and October 26, 2018 at the Laboratory.[4] The experiment was implemented using oTree (Chen et al., 2016).

The timeline of the experiment is described in Figure 1. The participants played the search task "game" after listening to the instructions and answering a comprehension questionnaire. The search task was the main experimental part where participants were involved in repeated sessions of the sequential stopping problem. This main task was followed by two control tasks: the *Bomb Risk Elicitation Task* (BRET - Crosetto and Filippin (2013); Holzmeister and Pfurtscheller (2016)) and the *Sustained Attention to Response Task* (Robertson et al., 1997); as well as a demographic questionnaire asking for the gender and the age class of the participant. All tasks except for the two questionnaires were paid, keeping on average similar monetary incentives. Experimental details as well as capture screens of each experimental part (instructions, tasks and questionnaires) are reported in the online Appendix E.

Figure 1: Timeline of the experiment

We used two treatments for the main task. In Treatment One-way (1w), we implemented the MC model, and in Treatment Two-way (2w), the MU model (see subsection 3.1 for more details). The search task was followed by two control tests and a short demographic questionnaire asking for the gender and the age class of the participant. The players repeated the search

---

[4]The name and location of the Laboratory have been removed in order to preserve anonymity.

task for 60 rounds, alternating sessions of both treatments (within-subject). In addition, the order in which these sessions were encountered was manipulated in a between-subjects design (see Table 1). In both treatment orders, the main task was preceded by a comprehension questionnaire of five questions, to which the correct answers were given after the full questionnaire was completed.

	Part 1		Part 2		Part 3
	10 rounds	**10 rounds**	**10 rounds**	**10 rounds**	**20 rounds**
Order 1w-2w:	One-way	Two-way	One-way	Two-way	Random combination
Order 2w-1w:	Two-way	One-way	Two-way	One-way	Random combination

Table 1: Treatment orders for the between- and within-subject design of the experiment.

3.1. Main task

The main task is thus composed of several sessions of the sequential MC and MU search models for a specific parameter configuration. The player faces $n = 10$ alternatives and each alternative in both MU and MC implementations contains 10 sub-units of time. The player thus has a total of 100 trials, or clicks (10 alternatives times 10 sub-units of time) for each treatment. In Treatment Two-way, the participant can freely[5] allocate the trials for the *exploration* of as many alternatives as s/he wishes, but s/he can also limit the number of trials for each alternative to increase the number of trials computed for the exploitation phase. Conversely, in Treatment One-way, the participant still has 100 trials, but we "force" her/him to spend ten trials on one alternative before passing to the next one or to exploit one of the ones already explored. Table 2 reports these treatment settings. This experimental design ensured a similar duration (60 sessions in total) and a coherent visual and interactive experience on the screen for the two models, while still guaranteeing the properties of the optimal decision of the MC model.

The exact values of the alternatives are discrete and drawn from a uniform distribution $\mathcal{U}(\{1, \ldots, N\})$ where $N = 100$. The values of the alternatives

---

[5]At the first trial, the player has only one possible action: to "see" the first alternative.

	Treatment One-way	Treatment Two-way
Alternatives	10	10
Trials	100	100
Minimum trials per alternative	10	1
Maximum trials per alternative	10	10

Table 2: Treatments

are hidden and can be discovered sequentially from left to right.[6]

Exploring an alternative.

As noted above, to explore/discover an alternative, the player clicks $s = 10$ times (Treatment One-way) or up to s times (Treatment Two-way) on the "See" button. By clicking $k \leq s$ times on the "See" button of an alternative, $\sum_{i=1}^{k} 2^i$ values are sampled from a normal distribution with mean the value of the alternative, and a standard deviation of $\sqrt{\sum_{i=1}^{s} 2^i}$, thus a standard deviation of around 45.23. At each click, these two values are computed and reported graphically by a dot, and an interval around that dot representing the uncertainty [mean $- 2 \cdot$ mean standard error, mean $+ 2 \cdot$ mean standard error]. After 10 clicks, the standard error of the mean is around 1, which enables a participant to discriminate clearly among integer-valued alternatives.[7] Finally, both treatments implement the mechanism that it is not possible to go back to *explore* an alternative again but only to select it for *exploitation*.

Exploiting an alternative.

The player can select only one single alternative to *exploit*. When s/he decides to *exploit* an alternative and thus to stop *exploring*, s/he needs to click on the "Choose" button. An alternative not explored cannot be chosen.

---

[6]The Online Appendix G shows different computer screens of the participant at different stages of the game.

[7]In the Instructions part of the experiment, the players were informed that the visual representation of a mean and of a standard deviation are provided and that they are obtained by samples of a normal distribution. We also told the subjects that at each click, the sample size doubled in such a way that at the last click (10th), they could discriminate with a 95% of certainty the real value of the alternative. Furthermore, players were informed that if the upper bound of the interval was greater than 100 (this can happen when adding a noise to an already high value), it was replaced by 100. If the lower bound of the interval was lower than 1, it was replaced by 1.

When choosing an alternative, the payoff is:

$$\frac{\text{Remaining trials} \cdot \text{Value of the selected alternative}}{10}$$

If the player spends all the trials on the *exploration* phase (i.e. does not choose any alternative), then her/his payoff is equal to zero.

3.2. Experimental research hypotheses

In line with previous literature, we have identified six research hypotheses that we aim to test. The first four focus on treatment effects, whereas the last two focus on behavioral aspects related to the search dynamics.

Treatment effects
H1: An oversearch tendency in Treatment One-way.
H2: An undersearch tendency in Treatment Two-way.
H3: A learning effect in Treatment One-way.
H4: A weaker learning effect in Treatment Two-way than in Treatment One-way.
In Treatment One-way, the exploration of a new alternative is much more costly than in Treatment Two-way where the cost can be lowered by accepting a higher uncertainty.

The literature has shown (see Section 1) that when sampling is relatively expensive, participants oversample and tend to learn over time to improve their search strategy. On the other hand, when it is relatively cheap, they undersample and fail to improve over time. We expect to observe the same behavioral pattern in our data, namely, a tendency to oversearch or undersearch in Treatments One-way and Two-way, respectively, and a stronger learning effect in Treatment One-way compared to Treatment Two-way.

Search determinants
H5: Anticipation favors exploration.
H6: Regret favors exploitation.
Regarding H5 and H6, we refer to the work of Zwick et al. (2003). They consider two empirical variables: the average rate of candidate arrival (AROCA)[8]

---

[8]An alternative is a candidate if it is the highest alternative observed so far.

at a given position in the sequence as a measure of anticipation and the number of periods (corresponding to alternatives in our case) since the last candidate (PSLC) as a measure of regret. In line with their findings, we expect to find a role played by anticipation and regret on the search dynamics. The intuition behind the hypothesis on anticipation is, as in Zwick et al. (2003), that an abundance of candidates will lead the participants to the erroneous optimistic belief that such a trend will continue ("*hot hand fallacy*"[9]), thus increasing the probability that they will explore a new alternative. As for regret, the intuition is that the longer the time since the last candidate was observed, the more disheartened the participants will get from continuing the search, and not profiting from that candidate. While in Zwick et al. (2003) the regret measure only considers time lag since the last candidate was encountered[10], we also take into account the loss incurred since that candidate.

In addition to theses six main research hypotheses, we test two additional hypotheses related to the individual differences in search behavior, which we only report in the Online Appendix D.

4. Results

Before reporting and discussing the experimental results, we describe how experimental outcomes for Treatment One-way and Treatment Two-way are studied and also compared to the predictions of the analytical models MC and MU, respectively.

Each participant in the main task plays 60 rounds of treatments One-way and Two-way (refer to Table 1). Each round comprises a series of ten independent alternatives (values from 1 to 100) sampled from the discrete uniform distribution which are different at each round and for each participant.[11] For each of the series used in the experiment, we have computed the predictions of the optimal decisions derived from our models MC and MU in order to track the performance of the participants' decisions. As far as Treatment

---

[9]Gilovich et al. (1985)

[10]In contrast to our experiment, their alternatives are ranked according to an ordinal scale.

[11]In total 2315 series for Treatment One-way and 2305 series for Treatment Two-way. The last 20 rounds for each participant comprise treatments of type One-way and Two-way randomly selected.

One-way is concerned, for each explored alternative (the participant must spend 10 clicks on one alternative), the corresponding optimal threshold is computed according to equation (1). Using the precise experimental values, for each participant's series we evaluate at each alternative whether the highest observed alternative so far is higher or equal to the theoretical threshold, in which case the search is stopped and this alternative is selected as the optimal stopping time. Otherwise, the exploration continues. If at the penultimate alternative (the 9th), none of the observed alternatives was higher than the current threshold, then, the highest one is exploited. For Treatment Two-way, we apply a similar estimation procedure. We consider in this case the information obtained from the first of the ten clicks on the alternative, corresponding to the first mean value based on just two samples.[12] The theoretical thresholds are given by equation (2). In Treatment Two-way, it might be optimal to click on the last alternative, if all previous alternatives exhibited a mean value lower than the current threshold.

Experimental results are based on sample sizes of 34 subjects in treatments order 1w-2w, and 32 subjects in 2w-1w. Eleven subjects were discarded from the analysis based on two exclusion rules. The first rule (two people excluded) is having less than two correct answers out of five in the comprehension questionnaire preceding the main task. The second rule (nine people excluded) consists in not recalling and exploiting a "candidate" more than 90% of the time (i.e. six rounds) in Treatment One-way (see Figure 5 in the Online Appendix C for details).

Table 3 reports for each treatment (One-way and Two-way) aggregated experimental and optimal stopping times.

---

[12]We know from the optimal decision that it is optimal to stop at the first click for each alternative (see section 2).

	One-way		Two-way	
	Exper.	**Optimal**	**Exper.**	**Optimal**
Mean	$2.46^{a,b}$	2.35	3.95^{a}	5.37
Std Dev	1.73	1.48	3	3.23
Median	2	2	3	5

Table 3: Median, mean and standard deviation estimates of the optimal stopping times based on experimental and simulated (optimal) outcomes. [a] indicates a significant difference at 1% between the experimental and simulated (optimal) distributions of the same treatment (two-tailed two-sample paired t-test). [b] indicates a significant difference at 1% between the experimental distributions of the two treatments (two-tailed two-sample unpaired t-test).

The intuition that the average optimal stopping time, that is, the best alternative on average at which to stop, is higher in Treatment Two-way than in Treatment One-way is confirmed both theoretically and experimentally.

Furthermore, the results highlight the tendency to oversearch in Treatment Two-way and to undersearch in Treatment One-way with respect to optimal values. But, in order to adequately investigate all the research hypotheses (see Section 3.2), we introduce and adopt two main indicators: the *stopping time gap* (SG) and the *payoff gap* (PG). For each indicator and for each participant i we have 60 values, one for each round r:

$$SG_{ri} = \frac{\text{Experimentally observed stopping time}_{ri}}{\text{Theoretically predicted stopping time}_{ri}}$$

$$PG_{ri} = \frac{\text{Experimentally observed payoff}_{ri}}{\text{Theoretically predicted payoff}_{ri}}$$

The closer the SG_{ri} or the PG_{ri} are to 1, the closer the participant is to the optimal decision at round r. Values of SG higher than 1 are interpreted as an oversearch tendency whereas values lower than 1 are interpreted as undersearch. We define (geometric) average indicators SG_i and PG_i at the participant level (for more details, see Online Appendix C).

4.1. Testing the research hypotheses

Treatment One-way

Table 4 reports both participants' *payoff gap* and *stopping time gap* for both treatment orders and the different parts of the main task.[13]

In order to investigate **H1** about the tendency to oversearch in Treatment One-way, we refer to the last two columns of Table 4. In particular, we test if the mean stopping time gap (right part of the Table) of all the participants is greater than 1. We find statistically significant evidence of oversearch in the last column where all 60 rounds are considered except for the 2w-1w order. But, if we consider only the rounds of the last part of the main task (3rd column of th right part), it is no longer significant. On average, people tend to learn the optimal stopping alternative and in the last rounds of the task they all play on average optimally. To test **H3** about the presence of learning, we therefore compare columns 1 and 3 of the right part of Table 4 related to stopping times. We consider thus the first and last 10 rounds of the main task, for both indicators. We find strong evidence of learning for both indicators and all treatment orders. In particular, participants' performance gets closer to the optimal theoretical prediction throughout the game. This learning seems faster in order 2w-1w, since a significant difference is already observed between the first two parts (see Figures 1 and 2 in the Online Appendix C), while this is not the case in treatments order 1w-2w. This difference can be interpreted as the result of longer practice. Indeed, those who started with Treatment Two-way (2w-1w order) have an advantage since they have practiced 10 rounds more (from the other treatment though) before playing the first part of Treatment One-way than those who started the main task directly with Treatment One-way. This final observation might also explain why we did not observe oversearch in treatment order 2w-1w (last column).

Experimental findings about **H3** are further confirmed by different generalized linear mixed-effects models at the individual level where the dependent variables are either the payoff gap or the stopping time gap (see Online Appendix C).

---

[13]Figures 1 and 2 in the Online Appendix C show in detail the distributions of the participants' *payoff gap* and *stopping time gap* for Treatment One-way as well as the statistical comparisons. In the paper, we report parametric test results (t-test). The adopted significance threshold is 5%.

|		PG^{C1}	PG^{C2}	PG^{C3}	PG^{C}	SG^{C1}	SG^{C2}	SG^{C3}	SG^{C}
1w-2w	*Median*	0.926	0.966	0.985	0.947	1.197	1.184	1.042	1.172
	Mean	0.832^a	0.900	0.962	0.895^b	1.290^a	1.252	1.132	1.206^b
2w-1w	*Median*	0.921	0.976	0.986	0.954	1.090	0.964	0.973	1.052
	Mean	0.870^a	0.928	0.938^b	0.909^b	1.111^a	0.982	0.972	1.019
Both	*Median*	0.924	0.968	0.985	0.947	1.122	1.045	1.000	1.089
	Mean	0.850^a	0.913	0.950^b	0.902^b	1.203^a	1.121	1.054	1.115^b

Table 4: Treatment One-way: Mean and median values for the *payoff gap* (top part) and *stopping time gap* (bottom part) for each treatment order (rows) and for the different parts (columns). ^{a} is reported only in column 3 and indicates a significant difference at 5% between part 1 and 3 distributions (two-tailed two-sample test). ^{b} is reported only in columns 5 and 6 and indicates a significant difference at 5% from the mean value of 1 for either part 3 or all (two-tailed one-sample test).

To investigate the "search determinants" hypotheses (see Subsection 3.2), we perform a survival analysis by means of a Cox proportional hazards model (Cox, 1972) in order to identify which variables might influence the "exploit" decision and thus the exploration vs exploitation trade-off. We adopt as the time variable trials 11 (i.e., the first click on the second alternative), 21 (i.e., the first click on the third alternative), 31 (i.e., the first click on the fourth alternative), and so on, corresponding to the end of the group of ten trials for an alternative. For instance, at the beginning of trial 11, the participant has observed the value of alternative one and can decide whether to explore the second alternative or exploit the current one. It is worth remembering that, in Treatment One-way, participants can make a decision only in these trials.

We include the two measures of anticipation and regret as covariates in line with the literature (refer to Section 3.2): *AROCA* (*Average Rate Of Candidate Arrival*) as a measure of anticipation, and the payoff loss since the last candidate was encountered – which we will call "*Loss Since the Last Candidate*" (*LSLC*) – as a measure of regret. Both indicators are built on the notion of number of encountered candidates (a candidate is the highest alternative observed so far). More precisely, the *AROCA* when inspecting an alternative is computed as the number of previously encountered candidates divided by the total number of already inspected alternatives. The *LSLC* corresponds to the sum of the differences between the observed payoff for the last candidate and the observed payoff for the inspected alternatives since that candidate and until the last inspected alternative. These covari-

ates are computed at each time[14] the player inspects a new alternative or exploits one.[15] Finally, we control for the individual characteristics, namely, demographic or control task factors (gender, age, risk aversion level and impulsivity level).

We report the results with a forest plot in Figure 2. Detailed results are reported in Table 4 in the Online Appendix C. A hazard ratio above 1 indicates that the covariate is positively associated with the "exploit" decision probability, and thus negatively associated with the duration of exploration phase.
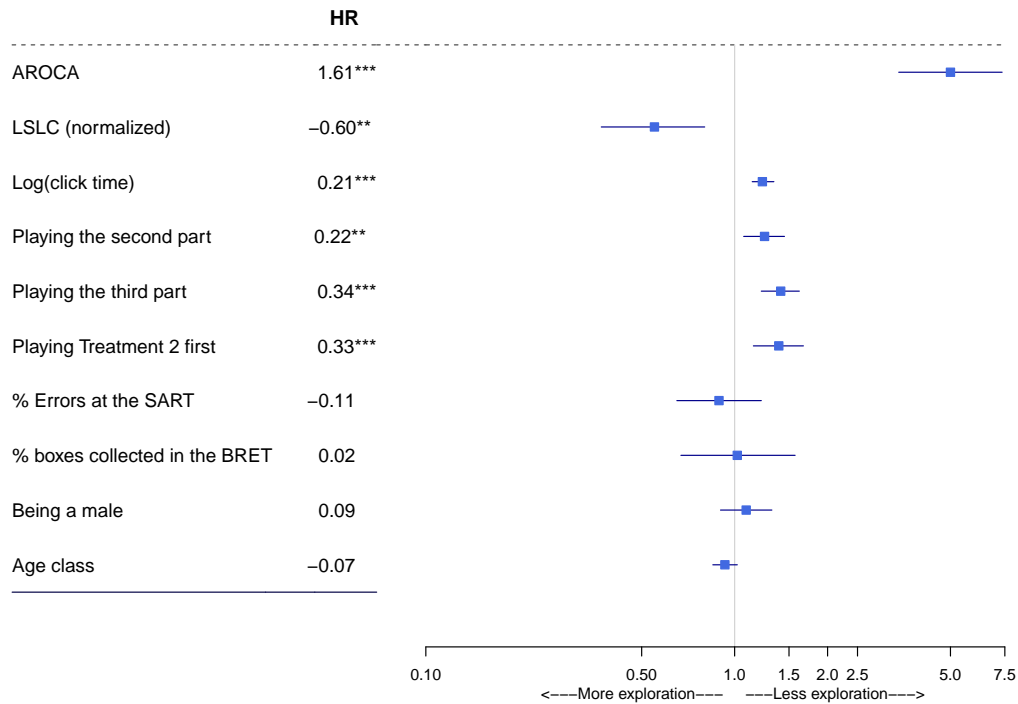
Figure 2: Forest plot of Cox regression model of time to exploit in One-way with robust standard errors
Note: $^*p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

The results highlight the crucial role of experience as well as the "candidate"

_____

[14]We introduce time-varying covariates in the Cox model.

[15]The first inspected alternative is considered a candidate. At the first trial, the *AROCA* and the *LSLC* are set to 0.

as an empirical notion for building measures of anticipation and regret. Indeed, the variables related to the temporal structure of the game (part of the game and treatment order) as well as the measures of anticipation and regret are statistically significant and relevant thus highlighting once more how learning is central in this context. The results reject hypotheses **H5** and **H6**, as we find opposite correlations of the anticipation and the regret measures than in Zwick et al. (2003). In particular, our results show that anticipation favors exploitation, while regret favors exploration. Finally, the variable "Log(click time)" is significant. This covariate measures the "impulsivity" or conversely the "reflectiveness" estimated as the log of the time (updated at each trial) spent so far before the decision to click for a new alternative or for exploiting a previous one. The more the time spent in deciding to click (reflecting more), the less the exploration. In general, we might conclude that irrespective of individual attitudes or demographic characteristics and despite a relatively uncertain environment, the game provides relevant cues through anticipation and regret signals to learn from for all participants. Participants exploit such information to improve their performance as previously shown. If we look more precisely at the different effects, we highlight that the more experience (the part of the game) the less the exploration. People in early periods tend to oversearch. This attitude might derive from an early tendency to explore more an uncertain environment, but progressively they learn to stop more optimally, thus maximizing their payoff. This is consistent with the findings in Table 4. Anticipation and regret counterbalance a potential tendency to oversearch and undersearch in this game. At a specific time/alternative t, the higher the level of regret is, the more likely it is that people will continue to explore. Conversely, at a specific time/alternative t, the higher the level of anticipation is, the more likely it is they will start to exploit. We interpret these findings as a realization of a so called *gambler's fallacy*, that is, the erroneous belief that if a particular event has occurred often (or rarely) in the recent past, it is less (or more) likely to happen in the future. We believe that these two competing signals are useful for learning the optimal stopping time throughout the game. Despite the uncertainty of the values of subsequent alternatives, participants use and trust signals from the environment based on past and current values.

Treatment Two-way
We report in Table 5 both participants' *payoff gap* and *stopping time gap* for

18

the different parts and treatment orders.[16]

To test the validity of **H2** about the tendency to undersearch in Treatment Two-way, we consider again the stopping time gap in the last two columns of the right part of Table 5. This time, we investigate whether the values are significantly lower than 1. For all treatment orders and for both columns (last part and all parts) **H2** is confirmed, except for treatment order 2w-1w at the aggregate level. But again, it is worth looking at learning dynamics to understand why order matters. As in Treatment One-way, a statistically significant difference between the first and the last parts of the treatment is observed for both treatment orders. These results confirm **H4** and thus the relevance of learning in these kinds of tasks. Contrary to Treatment One-way, in Treatment Two-way participants significantly deviate from the optimal stopping times in the last part of the experiment for both treatment orders, since the average stopping time gap gets further from 1. Repetition of the task worsen the average performance in terms of stopping time but, apparently in contradiction, the payoff gap increases and thus improves, approaching the value of one. This latter important issue is addressed in detail at the end of this section, whereas in the following we focus on SG analysis. As for the treatment One-way, we confirm the robustness of these findings
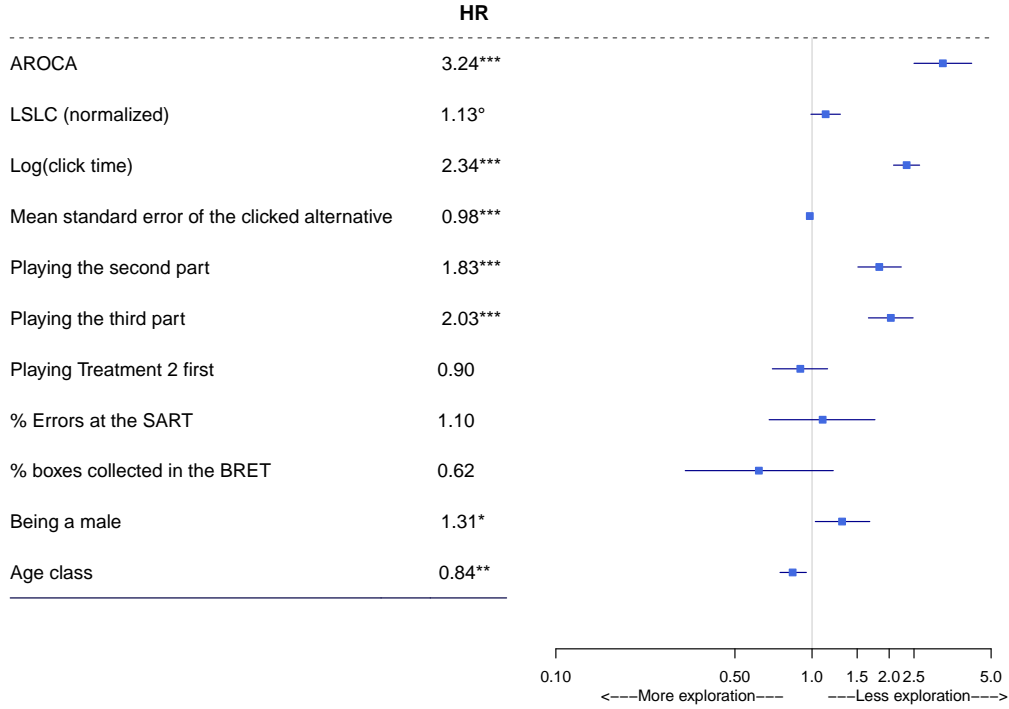
		PG^{U1}	PG^{U2}	PG^{U3}	PG^{U}	SG^{U1}	SG^{U2}	SG^{U3}	SG^{U}
1w-2w	*Median*	0.940	0.966	0.943	0.951	0.801	0.752	0.786	0.845
	Mean	0.850^a	0.922	0.946	0.901	0.985^a	0.843	0.877^b	0.888^b
2w-1w	*Median*	0.837	0.943	0.954	0.907	0.928	0.760	0.728	0.780
	Mean	0.770^a	0.874	0.891	0.841	1.077^a	0.907	0.789^b	0.895
Both	*Median*	0.907	0.948	0.949	0.935	0.819	0.760	0.764	0.796
	Mean	0.811^a	0.899	0.919	0.872	1.029^a	0.874	0.834^b	0.891^b

Table 5: Treatment Two-way: Mean and median values for the *payoff gap* (top part) and *stopping time gap* (bottom part) for each treatment order (rows) and for the different parts (columns). ^{a} 5% significance difference between part 1 and 3 distributions (two-sample test). ^{b} 5% significance difference between either part 3 or all and the mean value 1 (one-sample test).

by testing different generalized linear mixed-effects models at the individual level where the dependent variables are either the payoff gap or the stopping

---

[16]Figures 3 and 4 in the Online Appendix C show in detail the distributions of the participants' *payoff gap* and *stopping time gap* for Treatment Two-way as well as the statistical analysis.

time gap (see Online Appendix C).

	HR	
AROCA	3.24***	
LSLC (normalized)	1.13°	
Log(click time)	2.34***	
Mean standard error of the clicked alternative	0.98***	
Playing the second part	1.83***	
Playing the third part	2.03***	
Playing Treatment 2 first	0.90	
% Errors at the SART	1.10	
% boxes collected in the BRET	0.62	
Being a male	1.31*	
Age class	0.84**	

0.10   0.50   1.0   1.5 2.0 2.5   5.0
<---More exploration---   ---Less exploration--->

Figure 3: Forest plot of Cox regression model of time to exploit in Treatment Two-way
with robust standard errors
*Note: °p<0.1; *p<0.05; **p<0.01; ***p<0.001*

We again perform a survival analysis to study the drivers of the "exploit"
decision. Here, we consider all the trials (except the very first one where
no decision is required) since, unlike in Treatment One-way, the player is
not forced to spend a certain number of trials exploring a given alternative.
We fit the Cox proportional hazards model (see details about model speci-
fication in the Online Appendix C) where we keep the same covariates, but
add the mean standard error of the clicked alternative, a real-time variable
monitored by the participants on the screen which reports the uncertainty
in the estimation of the alternative value that they see. It is a kind of mea-
sure of the risk they are accepting. We report here in Figure 3 the results
only with a forest plot (Detailed results are reported in Table 5 in the On-
line Appendix C). As in Treatment One-way, anticipation ($AROCA$) plays a
relevant and significant effect in pushing participants to stop earlier in the

20

exploration. But this time, in contrast to Treatment One-way, the regret variable ($LSLC$) is not significant. This outcome might highlight a conscious behavioral attitude. Participants decide the level of uncertainty (variance). They can acknowledge this deliberate choice of accepting greater uncertainty at the cost of a regret feeling. They cannot use an exact value to estimate potential loss or benefits for counterfactual reasoning. Therefore, regret is a milder signal that does not allow participants to delay the choice to exploit and thus to counterbalance the undersearch tendency.

As in Treatment One-way, progress in the game variables (i.e., playing the second or the last part of the treatment) show once more that learning is occurring, especially in favor of a sooner exploitation. However, unlike in Treatment One-way, the treatment order does not have any significant effect, as previously discussed. To summarize, these results allow us to reject hypotheses **H5** and **H6**. Finally, we detect also the coherent effect of the variable "Mean standard error of the clicked alternative" that implies that the greater the uncertainty on the value of the alternative, the greater the probability to explore.

As previously noted, we will now address the important and apparently discordant issue about Table 5 where a divergent dynamics from the optimal SG coexists with a convergent dynamics towards an optimal PG.

Figure 4 represents the two directions, or ways of the exploration-exploitation trade-off on the two axes. On the y-axis the classical *à la Secretary* trade-off on the average stopping alternative is reported. The x-axis gives the trade-off relating to the average information sampled on the alternatives, that is, the average number of clicks by alternative. Circles correspond to two scatter plots where participant averages are reported over two distinct parts of the experiments, namely yellow circles for part 1 and red circles for part 3. We aim to follow the learning dynamics. We draw also the two regression lines. The two lines are almost parallel. The red one, corresponding to the last rounds of the experiment, is below the yellow one. This implies that throughout the experiment participants reduce the average number of clicks per alternative and/or the stopping time. This is consistent with previous findings about the decrease of the SG. The point is why the PG increases towards one, thus improving but not achieving it. We should detect in some sense an optimal strategy by reducing the average number of clicks per alternative, because we know that people are undersearching even in part 1 and thus decreasing further the average stopping time, as we observe, shouldn't be profit-maximizing. We note that the red line in the central interval $x \in [2, 6]$
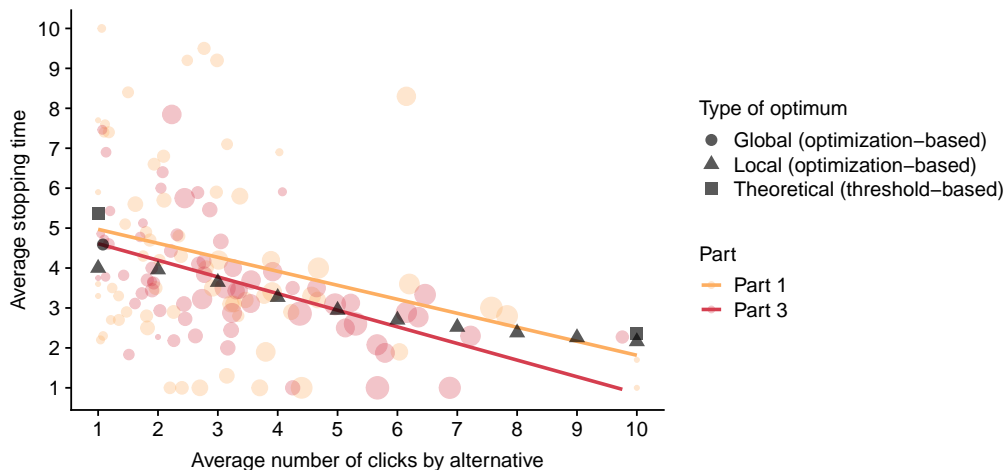
Figure 4: Scatter plot of individual averages of stopping times and number of clicks over the first part (yellow circles) and the last part (red circles) of the experiment.

(where there are more samples) approaches the triangles. Triangles are the outcomes of a new set of optimization-based simulations where we investigate the presence of local optima. The local optima that we consider are the ones that are determined by fixing the number of clicks per alternative. For each series encountered by the participants we compare the outcome of exploring one, two, three or ... alternatives, and estimate the average optimal stopping time. These simulations determine different local optima, one for each value of the number of clicks per alternative. Behaviorally, we might assume that participants constrain themselves throughout the experiment to the same average number of clicks per alternative adopted in part 1 and optimize accordingly. Let's suppose that a participant in part 1 performs on average 4 clicks on each alternative and we extrapolate that for instance s/he has explored on average 4 alternatives. We know by estimation that the global optimum[17] corresponds to an average number of clicks close to 1

---

[17]The global optima are estimated by optimizing in the two dimensional space. This implies that in almost 94% of the series participants should adopt a one click per alternative strategy. More details about these simulations are given in the Online Appendix B. The black dot is not on $x = 1$, whereas the closest triangle is exactly on $x = 1$, because we constrain the optimization on the subspace of points with $x = 1$. This justifies the discrepancy between these two points.

and an average stopping alternative of almost 5. The black dot reports this global optimum which should be approached by every efficient learner. S/he should increase her/his average exploration time to approach the value of 5 and simultaneously decrease the number of clicks per alternative. Both SG and PG values would increase. In general, all red circles should concentrate around the black dot if all people where efficient learners. But, with the triangles we highlight that there is in any case an opportunity to increase the PG by reducing at the same time the SG value as we observed in Table 5. It is indeed a sub-optimal strategy, but it is still locally a profit maximizing strategy. Why should people identify and favor this strategy instead of moving towards the global optimum? First of all, we assume that only in very few cases can people be perfectly rational by deducing the global optimum at an early stage of the experiment. These perfectly rational participants would already be playing close to the black dot in part 1. But this is not the case: we don't have yellow circles very close to the black dot, but we have a few red ones. We indeed have some participants in part 1 (yellow circles) already playing at an average number of clicks per alternative close to one, but not close to the optimal value. These participants should have an easier task to approach the global optimum or the corresponding local optimum. The other participants prefer at the beginning of the search task to spend more time on each alternative to reduce the uncertainty on the value of the alternative represented visually as a variance. The great majority of these participants are actually undersearching with respect to the global optimum (circles are below the global optimum on the y-axis). But we know that the regret feeling (LSLC) which is an oversearch signal is not driving the choice of stopping time in this context. The anticipation signal (AROCA), on the other hand, might push participants to focus on reducing the average stopping time. This in order to opens up an interesting and general insight on "two-way"search tasks. We propose then that in "two-way" sequential search tasks this is actually a trap. We will discuss this in the following sections.

4.2. Summary of results

Table 6 summarizes and compares the results of both treatments. The major results of our experiment so far are that: 1) we find clear evidence for learning throughout the game; 2) in Treatment One-way, participants learn on average to play optimally; and 3) they reinforce throughout time a tendency to undersearch in Treatment Two-way, because a learning dynamics

towards local optima takes place.

	Treatment One-way (Costly exploration)	Treatment Two-way (Cheaper exploration)
Theoretical stopping time	$\simeq 2$	$\simeq 5$
Observed search behavior	Oversearch	Undersearch
Learning	Strong	Weak
Regret	Increases exploration	No effect
Anticipation	Decreases exploration	

Table 6: Summary of the results of the two treatments

We report in Appendix also a final treatment comparison analysis to investigate if it is more difficult to learn in one environment than the other. We study this effect at the individual and not pooled level.

5. Discussion and conclusions

In this paper, we have studied both theoretically and experimentally an original optimal stopping problem, the "two-way" sequential search task, by comparing it to the more classical one-way version. We have discussed the analytic solutions for both models, which are characterized by a decreasing threshold search strategy, meaning that the minimum value that the individual should be ready to accept decreases over time. With respect to our modelling framework, finite horizon and common total number of trials in both treatments, the theoretical results show that the search in terms of number of explored alternatives is longer under uncertainty. The exploration of a new alternative is more costly since to gain information about the value of the alternative one needs to spend realistically more effort, thus trials, in the exploration of the alternative.

We have run a human subject experiment to test the empirical validity of theoretical and simulated outcomes. In line with the existing experimental literature on optimal stopping problems, we find that participants tend to oversearch and learn better when exploration is costly, and to undersearch

when exploration is relatively cheap.

Furthermore, we find that anticipation (based on an expectation formed in past periods) about the possible presence of better values in the next alternatives tends intuitively to reduce exploration, as opposed to the results in Zwick et al. (2003). Conversely, we find that regret, when present, counterbalances this tendency by increasing exploration. Indeed, with particular information (one-way search tasks) both anticipation and regret are active signals based on the notion of a candidate (the best alternative encountered so far). The participants can trust this information and form proper expectations or feelings of regret. As a result, they learn the search task correctly. When the information about the value of the alternative is not trustworthy, the participants use the notion of a candidate in terms of rankings (ordinal terms) as find Zwick et al. (2003), but not in cardinal terms (i.e., precise payoff losses). They thus advance on their search with an excess of exploration effort.

To summarize, our study provides some elements showing that in a highly uncertain world people actually underexplore alternatives because they prefer to put effort on achieving a higher level of certainty on their values. This result informs firms and policy makers of the need to encourage people to browse their environment to discover more options even in an approximate way. This can be achieved by designing proper management strategies. In particular, when referring to the context of the digital economy, our findings are in line with a tendency emerged within digital platforms for recommender systems or social media of designing interfaces in laptop or mobile applications with click and scroll mechanisms or with tap and swipe gestures that mimic our "two-way" sequential search tasks. In a tap and swipe designed application, in fact, tapping (i.e., the choice of getting more information) may be costly, while swiping (i.e., the option of having a fast look to the available options) can be achieved at almost zero cost. In general, our findings raise the issue of providing correct incentives, instructions or devices such as nudges or boosts to orientate people in sequential search.

Appendix A. Result comparison

We define a measure of strategic stability for an individual with respect to the optimal stopping time by using the dispersion of her/his SG values in the set of consecutive rounds determining the different parts of the game. We adopt as a measure of dispersion the geometric standard deviation (SD)

factor. This multiplicative factor is equal to 1 if all SG values are identical in consecutive rounds. This might occur either if a subject has learned the optimal strategy and therefore all SG values are stable and close to 1 or simply because the subject has learned and adopted a stable undersearch or oversearch behavior. In this latter case, the dispersion can also be very low even if the average (geometric mean) stopping time gap is not equal to 1.

Figure A.5 reports the distributions of the SD factors for the participants for both treatments and in each part of the experiment. We also plot six transparent distributions which show the results of a simulation study reproducing a worst case scenario of purely random subjects. The distributions are obtained by estimating the SG values as if the participants played randomly each round, adopting a uniform distribution for every choice they make. It is worth noting that, in the simulation, we have reproduced the same experimental conditions faced by the subjects in the experiment regarding the number of players, the number of rounds and the values of the alternatives. The comparison between the transparent and non-transparent distributions is thus informative of the capability of real subjects to understand the nature of the game and to define a strategy which is conditional to the specific round.

We can easily confirm previous findings about a learning effect. In Treatment One-way (upper panel) subjects throughout the game learn to become more stable since both the mode and the shape of the distributions move towards one. This is a statistically significant effect for all parts compared. Even at the end of the experiment (last two parts) subjects still improve strategic stability. This is not the case for Treatment Two-way (lower panel): on average they stop learning after the second part of the experiment.

In all conditions, we can highlight that the distributions are in general multi-modal. Despite the majority of subjects improving their strategic stability throughout the experiment thus approaching SD factor values of 1, few subjects clearly approach the optimal strategy. These observations suggest the in order to that Treatment Two-way is more difficult to play. The optimal strategy in Treatment Two-way is more difficult to detect and to comply with.
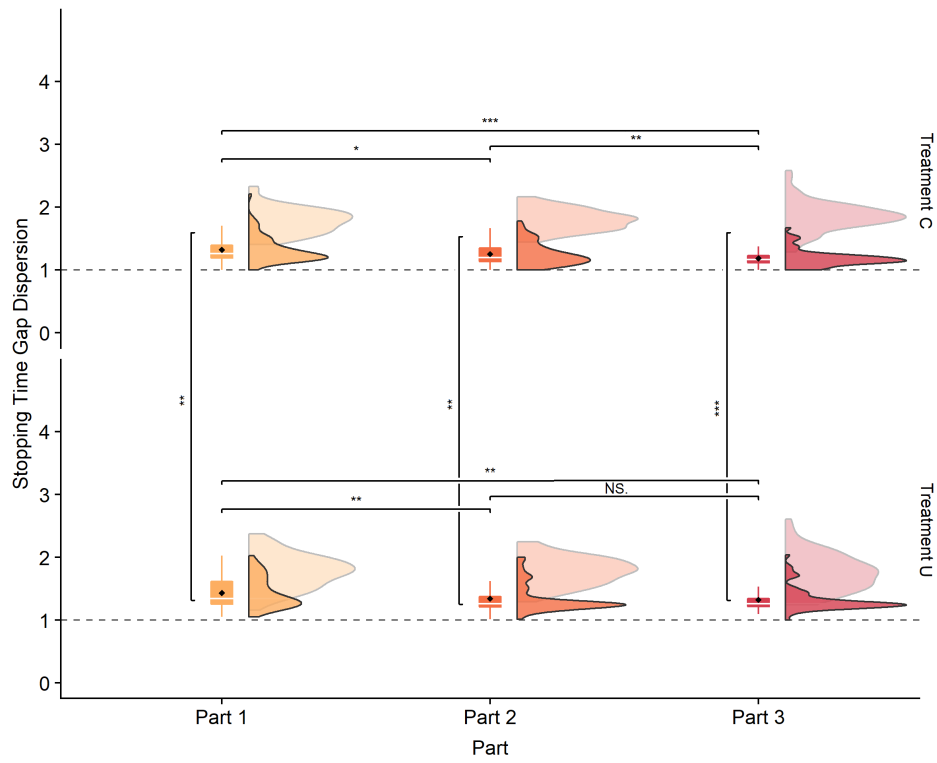
Figure A.5: Evolution of the dispersion of the stopping time gap through the different parts by treatment.

Boxplots (Black diamonds: mean) and Kernel probability density plots. Horizontal comparisons: Two-tailed paired t-test; Vertical comparisons: Two-tailed unpaired t-test

Note: $^{NS.}p{\geq}0.1;\ ^{\circ}p{<}0.1;\ ^{*}p{<}0.05;\ ^{**}p{<}0.01;\ ^{***}p{<}0.001$

Supplementary material

Supplementary material associated with this article can be found, in the online version, at ...

References

Bearden J.N., Rapoport A., Murphy R.O., 2006. Sequential observation and selection with rank-dependent payoffs: An experimental study. Management Science 52, 1437–1449.

Berry D.A., Chen R.W., Zame A., Heath D.C., Shepp L.A., et al., 1997. Bandit problems with infinitely many arms. The Annals of Statistics 25, 2103–2116.

Chen D.L., Schonger M., Wickens C., 2016. oTree—an open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance 9, 88–97.

Chow Y.S., Robbins H.E., Siegmund D., 1971. Great expectations : the theory of optimal stopping. Boston Mass by Houghton Mifflin.

Costa V.D., Averbeck B.B., 2013. Frontal–parietal and limbic-striatal activity underlies information sampling in the best choice problem. Cerebral cortex 25, 972–982.

Cox D.R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) 34, 187–202.

Crosetto P., Filippin A., 2013. The "bomb" risk elicitation task. Journal of Risk and Uncertainty 47, 31–65.

Descamps A., Massoni S., Page L., 2022. Learning to hesitate. Experimental Economics 25, 359–383.

Eriksson K., Strimling P., 2010. The devil is in the details: Incorrect intuitions in optimal search. Journal of Economic Behavior & Organization 75, 338–347.

Ferguson T.S., 2007. Optimal Stopping and Applications. Online book.

Ferguson T.S., et al., 1989. Who solved the secretary problem? Statistical science 4, 282–289.

Gilovich T., Vallone R., Tversky A., 1985. The hot hand in basketball: On the misperception of random sequences. Cognitive psychology 17, 295–314.

Greiner B., 2015. Subject pool recruitment procedures: organizing experiments with orsee. Journal of the Economic Science Association 1, 114–125.

Gupta A.K., Smith K.G., Shalley C.E., 2006. The interplay between exploration and exploitation. Academy of Management Journal 49, 693–706.

Hey J.D., 1987. Still searching. Journal of Economic Behavior & Organization 8, 137–144.

Hills T.T., Todd P.M., Lazer D., Redish A.D., Couzin I.D., 2015. Exploration versus exploitation in space, mind, and society. Trends in cognitive sciences 19, 46–54.

Hochman E., 1973. An optimal stopping problem of a growing inventory. Management Science 19, 1289–1291.

Holzmeister F., Pfurtscheller A., 2016. oTree: The "bomb" risk elicitation task. Journal of Behavioral and Experimental Finance 10, 105–108.

Juni M.Z., Gureckis T.M., Maloney L.T., 2016. Information sampling behavior with explicit sampling costs. Decision 3, 147.

Klein L.R., 1998. Evaluating the potential of interactive media through a new lens: Search versus experience goods. Journal of Business Research 41, 195–203.

March J.G., 1991. Exploration and exploitation in organizational learning. Organization science 2, 71–87.

Oprea R., Friedman D., Anderson S.T., 2009. Learning to wait: A laboratory investigation. The Review of Economic Studies 76, 1103–1124.

Robbins H., 1952. Some aspects of the sequential design of experiments. Bulletin of the American Mathematical Society 58, 527–535.

Robertson I.H., Manly T., Andrade J., Baddeley B.T., Yiend J., 1997. Oops!':
performance correlates of everyday attentional failures in traumatic brain
injured and normal subjects. Neuropsychologia 35, 747–758.

Sandri S., Schade C., Musshoff O., Odening M., 2010. Holding on for
too long? An experimental study on inertia in entrepreneurs' and non-
entrepreneurs' disinvestment choices. Journal of economic behavior & or-
ganization 76, 30–44.

Sang K., Todd P.M., Goldstone R.L., Hills T.T., 2020. Simple threshold
rules solve explore/exploit trade-offs in a resource accumulation search
task. Cognitive Science 44, e12817.

Schunk D., Winter J., 2009. The relationship between risk attitudes and
heuristics in search tasks: A laboratory experiment. Journal of Economic
Behavior & Organization 71, 347–360.

Seale D.A., Rapoport A., 1997. Sequential decision making with relative
ranks: An experimental investigation of the "secretary problem". Organi-
zational behavior and human decision processes 69, 221–236.

Seale D.A., Rapoport A., 2000. Optimal stopping behavior with relative
ranks: The secretary problem with unknown population size. Journal of
Behavioral Decision Making 13, 391–411.

Sonnemans J., 1998. Strategies of search. Journal of economic behavior &
organization 35, 309–332.

Todd P.M., 1997. Searching for the next best mate. In: Simulating social
phenomena, 419–436. Springer.

Todd P.M., Miller G.F., 1999. From Pride to Prejudice to Persuasion: Sat-
isficing in Mate Search, 287–308. Evolution and Cognition. New York:
Oxford University Press.

Zwick R., Rapoport A., Lo A.K.C., Muthukrishnan A., 2003. Consumer
sequential search: Not enough or too much? Marketing Science 22, 503–
519.