On the Decision to Explore New Alternatives:

The Coexistence of Under- and Over-Exploration

Kinneret Teodorescu and Ido Erev

Max Wertheimer Minerva Center for Cognitive Studies

The Technion- Israel Institute of Technology

Abstract

The decision between the "exploration of new alternatives" and the "exploitation of familiar alternatives" is implicit in many of our daily activities. How is this decision made? When will deviation from optimal exploration be observed? The current paper examines exploration decisions in the context of a multi-alternative choice task. In each trial, participants could choose a familiar option (the status quo) or a new alternative (risky exploration). The observed exploration rates were more sensitive to the common experience than to the average experience with exploration: participants exhibited under-exploration in "rare treasures" settings when the common outcome from exploration was disappointing and over-exploration in "rare mines" settings when the common outcome from exploration was attractive. This pattern can be captured with the assertion that the decision whether to explore new alternatives reflects reliance on small samples of past experiences. In addition, the findings highlight the value of a distinction between two types of exploration: forward-looking exploration, resulting from data collection tendencies, and backward-looking exploration, resulting from positive experiences with exploratory efforts in previous trials. We present a simple model based on these two motivations to explore new alternatives and demonstrate its high predictive value.

On the Decision to Explore New Alternatives and the Coexistence of Under- and Over-Exploration

Many important behavioral problems can be described as products of deviation from optimal exploration. The best-known examples are problems that have been depicted as reflections of insufficient exploration. One example is clinical depression. As noted by Seligman (1972) this disorder can be a result of learned helplessness: a state in which the organism does not explore enough. Such an interpretation of depression is supported by the observation that cognitive-behavioral therapy, one of the most effective treatments for depression, involves behavioral activation, a procedure in which the therapist encourages patients to participate in activities they no longer engage in, and to try new potentially rewarding activities (Beck, Rush, Shaw & Emery, 1979). Indeed, Jacobson et al. (1996) found that using only this behavioral activation component in therapy produced the same decrease in depression as full cognitive-behavioral therapy. In other words, enhancing active exploration may reduce depression.

Even among healthy individuals, researchers have found that people have a tendency to under-explore in a variety of domains. For instance, studies of performance in complex tasks reveal the value of training strategies which enhance exploration. For example, these training strategies were found to enhance performance among pilots (Gopher, Weil & Siegel, 1989; Seagull & Gopher, 1997), basketball players (see www.intelligym.com), as well as among experimental subjects in a multi-alternative choice task (Yechiam, Erev & Gopher, 2001).

Similarly, leading negotiation textbooks suggest that enhancing exploration of the parties' joint interests may help resolve social conflicts (e.g., Bazerman & Neal, 1992). For instance, studies of the fixed-pie bias show that negotiators tend to discard the possibility of a win-win result (Thompson & Hastie, 1990), and that encouraging them to explore the interests of the other side can lead to better agreements (e.g., Thompson, 1991).

These and similar studies suggest that without external guidance, people tend to exhibit insufficient exploration. The decision maker in such problems appears to select inefficient strategies, and to ignore the possibility that exploration may lead to the discovery of more effective ones. This common pattern can be described as an example of the status quo bias (Samuelson & Zeckhauser, 1988) or ambiguity aversion (Ellsberg, 1961).

There are times, however, when people exhibit the opposite bias, too much exploration. Unsafe sex and the exploration of untried illicit drugs are obvious examples (Bechara, 2005; Lowenstein, 1994). Another is extreme sports, which increasingly attracts participants who may not be fully cognizant of or prepared for the dangers involved (Palmer, 2002). Even exploring new paths while walking or hiking in certain parts of the world can be a suboptimal strategy, in view of the observation that thousands of civilians are injured or killed each year by landmines[1], and in other hiking accidents.

The co-existence of over- and under-exploration was explicitly studied in the context of consumer search behavior (Zwick, Rapoport, Lo & Muthukrishnan, 2003) and organizational strategy (Levinthal & March, 1993). Zwick et al. (2003) employed a simulated apartment purchasing task. At each stage, participants had to decide whether to accept the best available offer or continue to search. The participants did not search enough when searching had no cost, and searched too much when

[1] See the annual Landmine Monitor reports of the International Campaign to Ban Landmines, http://www.the-monitor.org/index.php/publications/display?url=lm/2010/es/Casualties_and_Victim_Assistance.html

searching was costly – even though they were given description of the task's incentive structure and were able to compute the "optimal stopping rule". Zwick and his co-authors proposed a behavioral decision rule that captures these findings. The rule assumes partial sensitivity to the factors that determine the optimal cutoff, and some sensitivity to other factors which are not correlated with the optimal search cutoff.

Levinthal and March (1993), considering exploration in the context of organizational strategy, suggested that organizations tend to exhibit insufficient exploration (e.g., do not invest enough in research and development) when their experience shows that most exploration efforts have failed. The opposite bias, over-exploration, occurs when most exploration efforts have seemed promising, but attempts to exploit these new technologies have led to disappointing outcomes. In this regard, Gavetti and Levinthal (2000) distinguished between two types of exploration: forward-looking and backward-looking exploration. The sensitivity to past experiences, suggested by Levinthal and March, is assumed to be a reflection of backward-looking exploration. Sensitivity to the future benefit from exploration is assumed to reflect forward-looking exploration.

The main goal of the current paper is to extend the study of the coexistence of over- and under-exploration to the context of individual behavior given limited information on the task's incentive structure. Specifically, we examine implicit exploration decisions in rudimental multi-alternative environments, in which information is attained through experience. We believe that this setting simulates situations in most real-world examples of over- and under-exploration, such as those presented above. For example, a depressed individual is not likely to decide explicitly between "exploration of new activities" and "exploitation of known activities". Rather, he or she selects between many alternative activities (e.g., eating one of many possible breakfasts, watching one of many TV shows, or visiting one of many web sites), where some imply exploration of new alternatives and others do not. In addition, the "potential explorer" in this and similar problems is not likely to have complete knowledge of the underlying payoff distributions.

Our analysis focuses on two possible explanations for the coexistence of insufficient and excessive exploration in such settings. The first explanation, henceforth referred to as the "mere noise" hypothesis, can be described as a generalization of Zwick et al.'s (2003) explication to the current context. It assumes that the coexistence of under- and over-exploration reflects a random component in the decision process. Such random behaviors, or "noise", can be driven by the stochastic nature of choice behavior (see Erev, Wallsten & Budescu, 1994; Thurstone, 1927) as well as an arbitrary or forward-looking search for information about the environment (Cohen, McClure & Yu, 2007; Daw, O'Doherty, Dayan, Seymour & Dolan, 2006). According to this explanation, although people's behavior is assumed to be generally guided by the optimal strategy for a given set of circumstances (Payne, Bettman & Johnson, 1988), noisy responses can lead to under-exploration when the optimal exploration level is very high and over-exploration when the optimal level is very low.[2]

It is important to emphasize that the mere noise hypothesis allows for the possibility that other factors besides noise contribute to the deviation from optimal exploration. For instance, a status quo bias may also give rise to insufficient exploration. Under this "status quo plus noise" scenario, insufficient exploration is the

[2] Under a simple abstraction of the effect of noisy responses, average exploration rates fall between the optimal rates, and the rates implied under random choice. Thus, the average rates reflect under-exploration when the optimal exploration level is very high and over-exploration when the optimal exploration level is very low. This statistical effect is commonly referred to as regression to the mean.

more common bias, but noise can still precipitate over-exploration when the optimal exploration level is extremely low. For example, a person might in general exhibit insufficient exploration of the sophisticated features of his new cell phone, yet still explore these features while driving – i.e., at a time when exploration is counterproductive.

The second explanation can be described as an adjustment of Levinthal and March's (1993) assertions to individual choices. It assumes that the coexistence of under- and over-exploration is a reflection of the tendency to rely on small sets of experiences in similar situations (see Fiedler, 2000; Gonzalez, Lerch, & Lebiere, 2003; Hertwig, Barron, Weber & Erev, 2004; Kareev, 2000). Reliance on a small set of experiences implies underweighting of rare events. Thus, it can lead to insufficient backward-looking exploration when the probability of success (in a given exploration effort) is low, and to excessive backward-looking exploration when the probability of success is high.[3]

Previous research suggests that the two explanations considered here – mere noise and reliance on small samples – affect behavior in a wide set of situations, including complex natural settings as well as simple laboratory contexts. Naturally, we chose to compare them in simple experiments. The paper is organized as follows: In Studies 1 and 2 we examine environments that enable us to qualitatively disentangle the two hypotheses described above. Then, we describe a simple model to account for the results and to drive quantitative predictions. In Study 3, a spectrum of payoff structures is examined and the a-priori predictions of the model are tested. Last, the general findings and its implications are discussed.

[3] The exact relationship of the current hypothesis to Levinthal and March (1993) is clarified in the general discussion.

Study 1

Method

Participants. Twenty Technion students (9 women and 11 men, with an average age of 24) served as paid participants in the experiment. They received a show-up fee of 20 NIS (about $5.5), and could win or lose up to 11 NIS depending on their performance in the experiment. The experimental session lasted about 10 minutes.

The task. This study used a multiple-alternative choice task. The alternatives were 144 unmarked keys presented in a 12x12 matrix (see Figure 1). In each trial, participants select one key and their choice is followed by immediate presentation of the trial's payoff on the selected key. The payoff associated with each key is either a gain or a loss, as described below, but only when the key is first selected; subsequent selection of any key always produces a status quo payoff (i.e., a payoff of 0). However, participants receive no prior information concerning the payoff structure, and so have to rely solely on their experience. Exploration of new alternatives, in this setting, is naturally defined as selecting a key that was not previously selected.

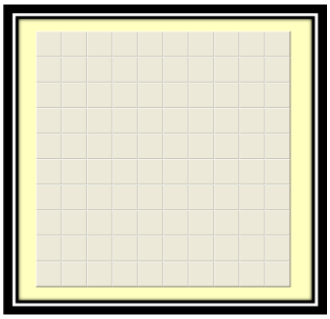

Figure 1. The computer screen presented to participants at the start of each trial.

Two payoff structures were used. In the "Rare Treasures" condition, 90% of the keys initially produced a loss of 1 NIS (about $0.3), and 10% produced a gain of 10 NIS. Thus, the expected payoff from exploration was positive (10(.1) -1(.9) = +.1). Since the payoff from repeating a choice was 0, the optimal strategy was to explore. In the "Rare Mines" condition, 90% of the keys initially produced a gain of 1 NIS, and 10% a loss of 10 NIS. The expected payoff from exploration in this case was negative (1(.9) -10(.1) = -.1), and exploration was costly in the long run.

Experimental design. The experiment used a within-subject design; each participant took part in the two conditions described above. The order of the two conditions was counterbalanced across participants. At the beginning of the experiment, participants were informed that they will play two distinct games of 100 trials each, and that their task is to select one key in each trial. The participants were further told that one of the trials would be randomly selected at the end of the experiment, and their payoff in that trial would be added to (or subtracted from) their show-up fee.

Predictions. The two hypotheses considered here lead to contradictory predictions about the experimental conditions. The mere noise hypothesis predicts more exploration in the Rare Treasures condition (where the optimal exploration level is high) than in the Rare Mines condition (where the optimal exploration level is low). The reliance on small samples hypothesis leads to the opposite prediction: Since rare experiences are less likely to be included in a small sample, participants' behavior is expected to be guided by the more common experience. Accordingly, this hypothesis predicts more exploration in the Rare Mines condition (where the common outcome of exploration is positive) than in the Rare Treasures condition (where the common outcome of exploration is negative).

Results

The data analysis was performed with respect to the percentage of trials that were exploratory (trials in which participants tried a new key divided by the total number of trials). In order to probe the learning process throughout the task, each condition was divided into 5 blocks, consisting of 20 trials apiece. For each participant, exploration rates (percentage of trials which were exploratory) were calculated for each block. Figure 2A presents the main experimental results.

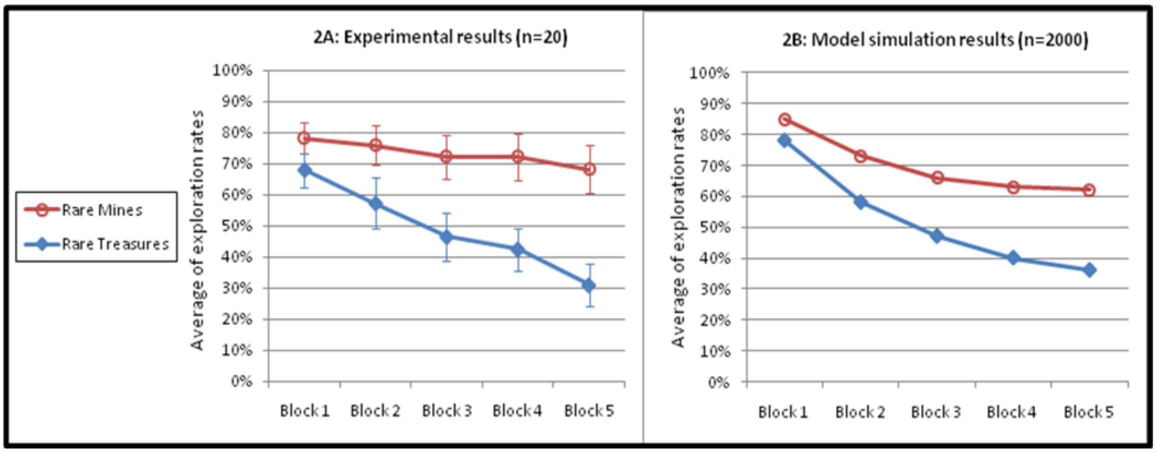

Figure 2. Average exploration rates by blocks of 20 trials. The left side (2A) displays the experimental results with SE bars, and the right side (2B) displays the results obtained from a simulation of the explorative sampler model, presented after the discussion of Study 2.

We conducted a repeated measures ANOVA with two within-subjects factors: experience with the task as indicated by the block number (1-5), and the incentive structure condition (Rare Mines vs. Rare Treasures). The results revealed a main effect of condition, with significantly higher exploration rates in the Rare Mines condition, in which the optimal strategy was to exploit, compared with the Rare Treasures condition, in which the optimal strategy was to explore (F(1,19)=10.63; p=0.004). Tukey's post-hoc test showed the largest gap between the two conditions to be in the final block, with exploration rates of 69% and 31% in the Rare Mines and Rare Treasures conditions respectively (p<0.001). This pattern suggests that the coexistence of over- and under-exploration is better described as a reflection of reliance on small samples than as a reflection of mere noise.

In addition, the results reveal a main effect of the block (F(4,76)=6.25; p<0.001): the observed exploration rates decreased with time. However, as can be seen in Figure 2A, this decline is mostly due to the dramatic decrease in exploration rates in the Rare Treasures condition, a pattern that was almost absent in the Rare

Mines condition. This interaction effect between condition and block was significant (F(4,76)=3.28; p=0.015).

Figure 3A presents the individual exploration rates in the last block as a function of the average payoff from exploration in previous blocks. The results reveal an increase in exploration as a function of the average payoff within each condition, but comparing the mean of the two conditions reveals the opposite pattern for the typical subject. In the Rare Treasures condition, participants experienced on average a higher mean payoff from exploration (+0.08 versus -0.32), but exhibited lower exploration rates (31% versus 69%), compared with the Rare Mines condition.

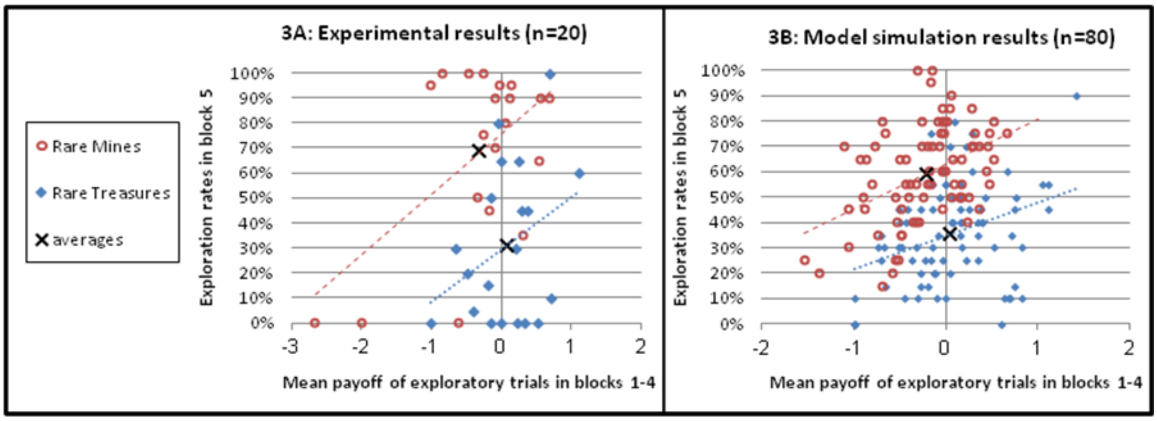

Figure 3. Exploration rates in the last block as a function of the mean payoff from exploration in previous blocks. Each dot represents one participant, the Xs represent the averages over participants in each condition, and the lines represent the fitted linear (regression) trends. The left side (3A) displays the experimental results and the right side (3B) displays the results obtained from a simulation of the explorative sampler model, presented after the discussion of Study 2.

Discussion

In this study we examined two payoff structures in which relying on the common outcome from exploration leads to suboptimal exploration levels (defined according to the expected value from exploration). The results suggest that exploration decisions reflect higher sensitive to the common experience than to the average experience. Participants exhibited insufficient exploration when the common experience with exploration was disappointing (the payoff "-1" in the Rare Treasures condition), and explored too much when the common experience was reinforcing (the payoff "+1" in the Rare Mines condition). This difference between the two conditions favors the reliance on small samples hypothesis over the mere noise hypothesis.

It is important, however, to note that the current results do not negate the possibility that the responses include a noisy (i.e., random) component. Indeed, important features of the current results are consistent with this assumption. The high exploration rates that were observed in early blocks, as well as the decrease in exploration rates with time, may be the product of a certain percentage of random choices: Random choices imply high exploration rates in early trials (when most options are new), and a reduced probability of selecting new keys over time (when the proportion of new keys is lower).

One way to explain this random-choice-like pattern is the assumption that it reflects the role of forward-looking exploration (see Gavetti & Levinthal, 2000). This kind of exploration is more likely to drive behavior at the beginning of each task (when being forward looking is important) than toward the end (see a similar assumption in Hariskos, Leder & Teodorescu, 2011). As noted above, in the current setting this form of exploration can be approximated as a random choice between all available alternatives. Study 2 was designed to improve our understanding of this assumption by examining forward-looking choices distinctively. This study focuses on extremely simple environments, in which backward-looking decision making will always lead to the optimal strategy. In such environments, any deviations from the optimal strategy can only be attributed to forward-looking choices.

Study 2

Method

Participants. Twenty Technion students (6 women and 14 men, with an average age of 24) who did not take part in Study 1 served as paid participants in the experiment. They received a show-up fee of 30 NIS and could win an additional 1 NIS or lose up to 10 NIS, depending on their performance in the experiment (average total payoff of 29 NIS, about 8 USD). The experimental session lasted about 20 minutes.

The task and experimental design. The same basic paradigm as in the first study was used, only this time with 120 alternatives (a 12X10 matrix). The following four simple environments were examined within participants (the order of the environmental conditions was counterbalanced across participants):

Condition ‘All zero’: All keys always yield a payoff of zero, whether they represent a new alternative or one that was previously selected. In other words, the trial’s payoff is always zero. In this condition, there is no optimal strategy, and backward-looking decisions will always result in a random choice between exploration of new alternatives and exploitation of familiar alternatives.

Condition ‘Explore+1’: Selection of a new alternative results in a payoff of +1, while selection of a familiar alternative (i.e., a key that has been selected in a previous trial) results in a payoff of zero. In this condition, the optimal strategy is to keep exploring new alternatives, and since each strategy will always produce the same payoff (+1 for new keys, 0 for familiar ones), backward-looking decisions will always lead to the optimal strategy.

Condition ‘Explore-1’: Selection of a new alternative results in a payoff of -1, while selection of a familiar alternative results in a payoff of zero. Again, backward-looking decisions will always lead to the optimal strategy, which in this condition is to select a familiar alternative.

Condition ‘Explore-10’: Selection of a new alternative results in a payoff of -10, while selection of a familiar alternative results in a payoff of zero. This condition is similar to condition ‘Explore-1’, in which the optimal strategy is to select a familiar alternative. We included this variation to examine the possibility that forward-looking behaviors depend upon the magnitude of the loss from exploration. Will participants apply the optimal strategy more quickly when exploration is more costly (compared with the 'Explore-1' condition)?

Results and Discussion

A repeated measures ANOVA revealed highly significant main effects of both condition and block, as well as a significant interaction (F(3,57) = 22.24; F(4,76)=65.63; F(12,228)=6.14 respectively, with p<0.001 for all effects). The average exploration rates in the four conditions are presented on the left side of Figure 4.

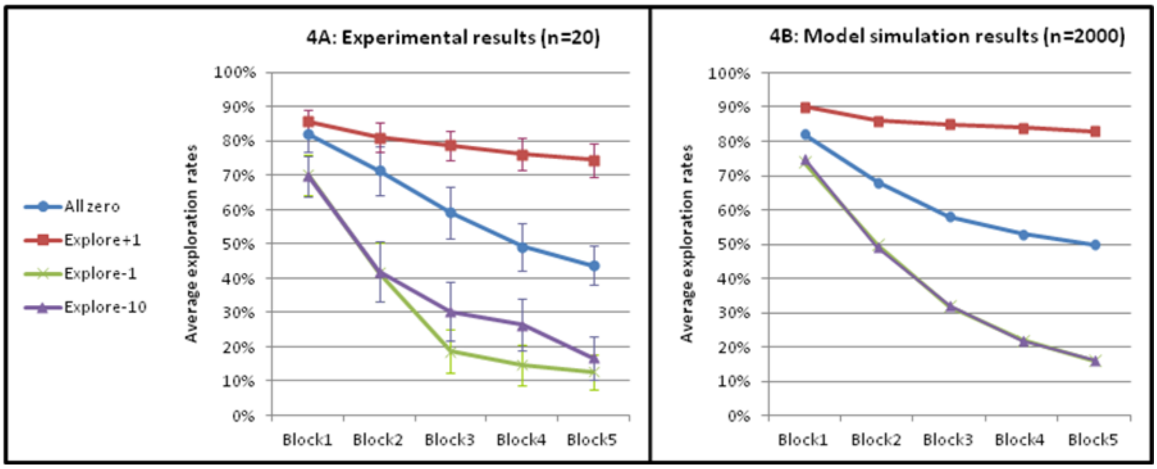

Figure 4. Average exploration rates by blocks of 20 trials. The left side (4A) displays the experimental results with SE bars, and the right side (4B) displays the results obtained from a simulation of the explorative sampler model, presented after the discussion of Study 2.

As can be seen in the graph, in the first block, high exploration rates (above 50%) were observed in all conditions. However, by the last block exploration rates fell to 74.5% (from 85.75%) in the 'Explore+1' condition, 43.75% in the 'All zero' condition, 16.75% in the 'Explore-10' condition and only 12.75% in the 'Explore-1' condition. Tukey's post-hoc test of the mean exploration rates showed that while the 'Explore-1' condition differed highly significantly from the 'All zero' and 'Explore+1' conditions ($p<0.001$ for both), there was no significant difference between the 'Explore-1' and 'Explore-10' conditions ($p=0.85$).

The lack of significant difference between the 'Explore-1' and 'Explore-10' conditions is consistent with previous findings that show limited sensitivity of decisions from experience to payoff magnitude (see review in Erev & Barron, 2005). At the same time, the participants were highly sensitive to whether or not exploration was efficient. They learned to keep exploring when exploration was effective ('Explore+1') and to explore much less when exploration was costly ('Explore-1' and 'Explore-10'). These results are consistent with any reasonable model of backward-looking decision making, as well as with animal studies, which show that variability of responses can be reinforced (Neuringer, 2002).

It is important to note, however, that exploration rates were still far from the optimal level. For example, in the 'Explore+1' condition, backward-looking choices should lead to exploration of new keys in 100% of the trials. Yet, in the last block, the average participant explored new keys in only 74.5% of the trials, which in the current context, can be referred to as insufficient exploration. Here, neither sampling biases in particular nor any other backward-looking mechanism in general can explain this deviation from the optimal exploration level. However, stochastic responses will cause asymmetric noise in cases where the optimum is very high or very low. This explanation was addressed in the mere noise hypothesis. Although noise by itself cannot explain the results of Study 1, where deviation from the optimal exploration level was also the result of backward-looking choices, the results of Study 2 show that regression to the mean also plays a role and that random choices are evident not only in early blocks, but also in later choices.

An Explorative Sampler Model

The results obtained in the two studies above can be reproduced with a simplified variant of the explorative sampler model (Erev, Ert & Yechiam, 2008).[4] The modified model distinguishes between two motivations for exploration: forward-looking exploration, resulting from data collection tendencies, and backward-looking exploration, resulting from positive experiences with exploratory efforts in previous trials. The two motivations for exploration are captured with the assumption that each decision in a multiple-alternatives environment is made in one of two modes: "forward-looking mode" or "backward-looking mode". The forward-looking mode implies a random choice between all the alternatives. The probability of using the forward-looking mode decreases with trials, and depends on the expected length of the experiment (T): it diminishes quickly when *T* is small, and slowly when *T* is large (Carstensen, Isaacowitz, & Charles, 1999). The probability of using the forward-looking mode at trial t is $P(Forward_t) = F_i^{\frac{t-1}{T}}$, where 0> F_i >1 is a trait of participant i that captures the tendency to collect information under the forward-looking mode.

Under the backward-looking mode, participant i samples (with replacement) M_i past experiences with each strategy (M_i> 0 is a trait of the participant), and selects the strategy with the highest sample mean (exploration of new alternatives vs. exploitation of familiar alternatives) or randomly in the case of a tie. The choice

[4] The current model generalizes a restricted variant of the explorative sampler model. The restrictions imply linear value function, no recency effect, and complete sensitivity to small samples. They were introduced to clarify the analysis. Relaxing these restrictions can only improve the fit of the model.

among the alternatives themselves (i.e., which familiar or unfamiliar alternative to select) follows a similar logic, but the details of this choice do not affect the results considered here (see Figure 5 for illustration).

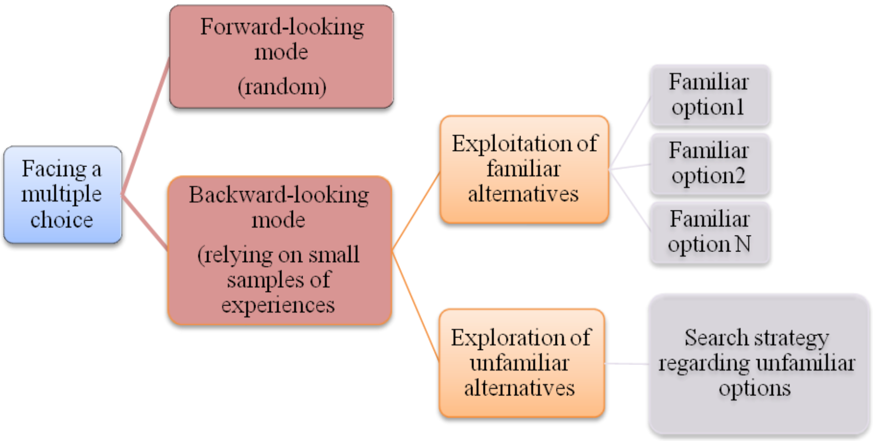
Facing a multiple choice
Forward-looking mode (random)
Backward-looking mode (relying on small samples of experiences
Exploitation of familiar alternatives
Exploration of unfamiliar alternatives
Familiar option1
Familiar option2
Familiar option N
Search strategy regarding unfamiliar options

Figure 5. An illustration of the modified explorative sampler model.

The right-hand boxes in Figures 2-4 present the predictions of the model under the assumption that the traits are drawn from uniform distributions: F_i from u(0, ϕ) and M_i from {1, 2, ,,,, μ}. The two free parameters were set (to fit the data) at $\phi = .24$ and $\mu = 8$. In addition, the derivation assumed an accurate recollection of the length of the experiment (T = 100). The predictions were derived using a computer simulation. The results show that the model reproduces the following observations: (1) over-exploration in the Rare Mines condition; (2) insufficient exploration in the Rare Treasures condition; (3) higher exploration rates in the Rare Mines than in the Rare Treasures condition; (4) a decrease in exploration with time; (5) a sharper decrease with time in the Rare Treasures condition; (6) higher sensitivity to the common payoff than to the average payoff; and (7) a similar exploration pattern in the 'Explore-1' and 'Explore-10' conditions.

In summary, the current two-parameter model captures all seven qualitative phenomena documented above. In addition, the model appears to provide a good quantitative fit for the mean exploration rates. It is important to recall, however, that the quantitative assumptions and the values of the two free parameters were post-hoc fitted to the behavioral data, and it is possible that the model's apparent success is a reflection of overfitting (Roberts & Pashler, 2000). Study 3 was designed to address this possibility. It examines the predictive value of the explorative sampler model in a broader set of payoff structures, in which the congruence between the common experience with exploration and the average experience is varied.

Study 3

Method

Participants. Forty Technion students (23 women and 17 men, with an average age of 25) who did not take part in the first two studies served as paid participants in the experiment. They received a show-up fee of 15 NIS, and could win up to 40 NIS depending on their performance in the experiment (average total payoff of 45 NIS, about $13). The experimental session lasted about 30 minutes.

The task. The same basic paradigm was used. However, in this study, we examined a spectrum of payoff structures and used a quasi-random algorithm to select the paradigm's parameters and determine the settings (a detailed description of the algorithm is presented in Appendix 1). Figure 6 presents the ten conditions randomly chosen according to this algorithm:

Condition		Exploitation of familiar alternatives	Exploration of new alternatives	Expected value from exploration
C95: Costly 0.95	Rare Mines	9.5	(**10**, 0.95; -10)	9
C75: Costly 0.75		11.5	(**15**, 0.75; -1)	11
C50: Costly 0.50		4.5	(11, 0.50; -3)	4
C25: Costly 0.25		8.5	(50, 0.25; **-6**)	8
C5: Costly 0.05		5.5	(100, 0.05; **0**)	5
E95: Effective 0.95		11.5	(**13**, 0.95; -7)	12
E75: Effective 0.75		0.5	(**4**, 0.75; -8)	1
E50: Effective 0.50		2.5	(8, 0.50; -2)	3
E25: Effective 0.25	Rare Treasures	5.5	(27, 0.25; **-1**)	6
E5: Effective 0.05		1.5	(116, 0.05; **-4**)	2

Figure 6. The ten randomly selected conditions. In each condition, exploration is either costly or effective in the long run (represented by C or E in the condition name) and there is a 0.95 to 0.05 probability of getting a higher payoff from exploration than exploitation (represented by the digits in the condition name). The columns show the payoffs and probabilities for exploitation and exploration and the expected value from exploration; in the middle column, the common experience with exploration is shown in bold. For example, in the last condition – E5 – pressing a familiar key (exploitation) always results in a payoff of 1.5, and exploring a new key results in a payoff of +116 with probability 0.05 and -4 otherwise. The expected value from exploration in this condition is 2. Thus, exploration is effective in the long run, although the common experience with exploration is disappointing (-4 compared with 1.5).

In the condition names, C and E reflect whether exploration is costly or effective according to expected values, while the number represents the probability of getting a higher payoff from exploration compared with the constant payoff obtained from exploitation. Where this number is higher than 50, the common outcome from exploration (shown in bold in the figure) is better than from exploitation; where it is lower than 50, the common outcome from exploration is worse.

Notice that in conditions C95 and C75, the common outcome from exploration is better than that from exploitation, but exploration is costly in the long run. Therefore, these conditions are different versions of the Rare Mines condition from Study 1. Similarly, conditions E25 and E5 are different versions of the Rare Treasures condition from the first study, since the common outcome from exploration is worse than the outcome from exploitation, but exploration is effective in the long run.

Each condition consisted of 60 alternatives and 50 trials. Each participant experienced all ten conditions, with the order counterbalanced across participants. Between conditions, participants were informed that although the instructions for playing the following game (condition) are the same, the payoff structure would be different. Unlike the first study, the performance-based payment in this experiment was calculated on an accumulated basis, meaning that participants accumulated their payoffs rather than receiving a payoff for one trial chosen randomly (as in Studies 1 and 2). Participants were informed about this procedure at the start of the experiment.[5]

Predictions. The left-hand columns in Figures 7 and 8 present the predictions of the explorative sampler model to the current study. Figure 7A shows the predicted exploration rates over trials, and Figure 8A shows the predicted rates in two blocks of 25 trials. The predictions were derived using a computer simulation in which 2000 virtual agents that behave in accordance with the model (with the parameters that best fitted the results of Study 1 and 2), participate in the 10 conditions of Study 3.

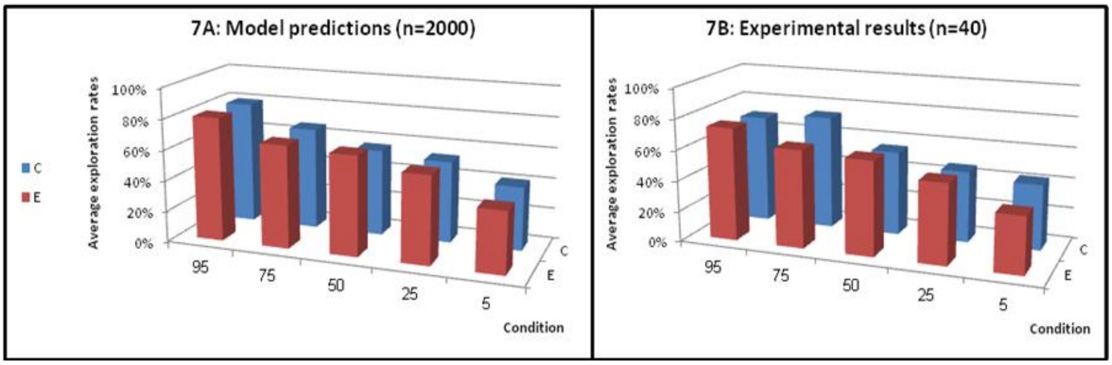

Figure 7. The left side (7A) displays the model's a-priori predictions based on 2000 simulations with the same value for the trait parameters as in the first two studies (without any fitting procedure). The combination of the z and the x axis produce the ten conditions: The z axis represents the probability to receive a higher payoff from exploration and the x axis represent whether exploration is costly or effective in the long run. For example, condition E95 is represented by E on the z axis and 95 on the x axis. The right side (7B) displays the average exploration rates across all subjects (n=40) in each of the ten conditions.

[5] The purpose of using the one-trial payment procedure in the first two studies was to avoid "wealth" issues – situations in which the subject feels that he has earned enough money and so no longer needs to pay attention to the task. Since the one-trial payment is often criticized for being unrealistic, in Study 3 we used the accumulated procedure, but without presenting the accumulated-sum to participants (in order to relax "wealth" issues).

As Figure 7A shows, the model implies that the co-existence of over- and under-exploration, documented in Study 1, is expected to emerge in the current study as well. The model predicts higher exploration rates in conditions C95 and C75 (Rare Mines environments) than in conditions E25 and E5 (Rare Treasures environments). As noted before, in these conditions, greater sensitivity to the common outcome compared with the average outcome will result in suboptimal exploration rates: that is, under-exploration in Conditions E5 and E25, and over-exploration in Conditions C75 and C95.

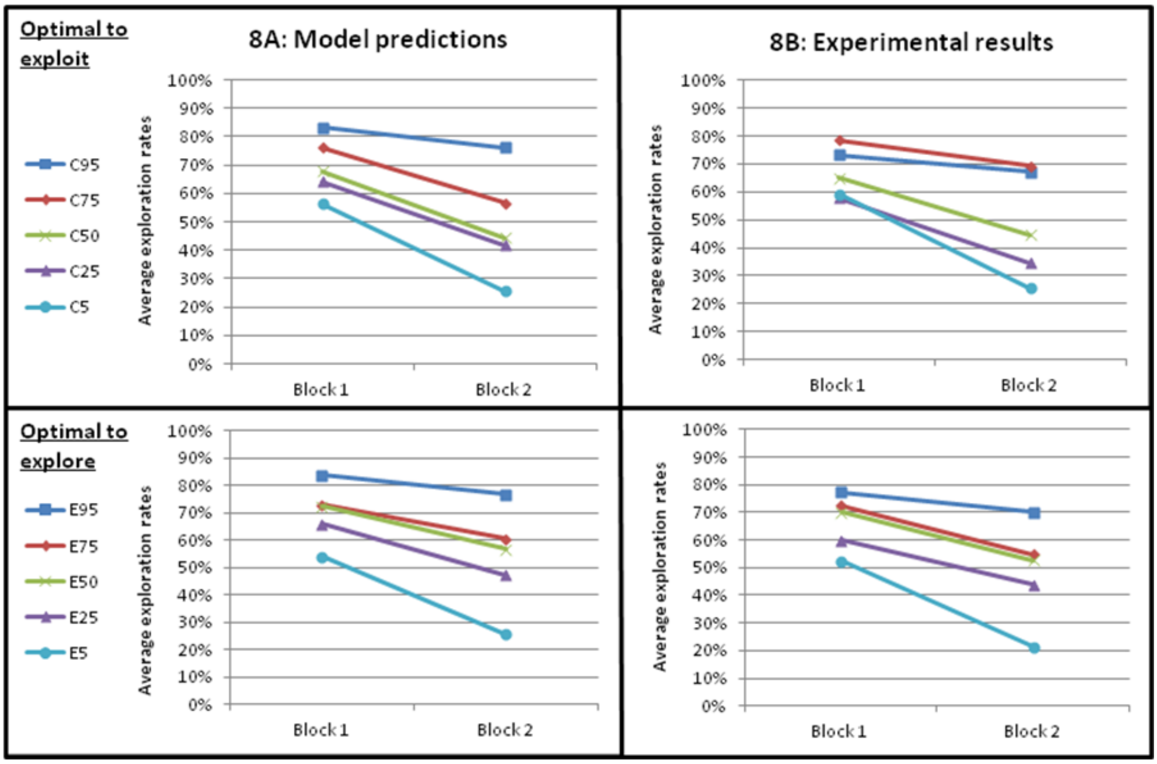

Figure 8. Average exploration rates divided into two blocks of 25 trials each. The upper graphs show conditions C95-C5, in which exploration is not optimal, and the lower graphs show conditions E95-E5, in which exploration is the optimal strategy. The left side (8A) presents the model's a-priori predictions based on 2000 simulations with the same value for the trait parameters as in the first two studies. To the right (8B) are the experimental findings (n=40).

As seen in Figure 8A, the model implies that the contingent decrease in exploration over time, discussed above, is expected to occur in the current setting too. Comparison of the predicted exploration rates in the first and second blocks of 25 trials reveals a predicted decrease in exploration with time in all ten conditions. In addition, the model predicts a greater decrease when the common outcome of exploration is disappointing.

Results

A repeated measures ANOVA revealed highly significant main effects of condition and block as well as a significant interaction effect (F(9,351)=15.85; F(1,39)=115.64; F(9,351)=6.42 respectively, with p<0.001 for all effects). The right-hand side of Figure 7 (7B) presents the observed exploration rates for each of the ten conditions. The results reveal high correspondence to the model's predictions. The correlation between the model predictions and the observed rates (using average exploration rates in each condition as a unit of analysis) is 0.94.

As predicted, higher exploration rates were observed when the common experience with exploration was rewarding than in conditions where the common experience was disappointing, regardless of the optimal exploration level. More specifically, participants explored new keys in 70.25% of the trials in condition C95, when it was not optimal to explore but the common outcome from exploration was reinforcing (an extreme version of the Rare Mines condition), and explored new keys in only 36.9% of the trials in condition E5, in which exploration was optimal but the common outcome from exploration was disappointing (an extreme version of the Rare Treasures condition). Tukey's post-hoc test showed that this gap (33.35%) was highly significant (p<0.001). In addition, the gap between exploration rates in conditions C75 and E25 (moderate versions of the Rare Mines and Rare Treasures conditions, respectively) reached 22% and was also highly significant (p<0.0001).

The right-hand side of Figure 8 (8B) presents the observed exploration rates in two blocks (25 trials per block, for each 50-trial game). The results correspond to the model's predictions: higher exploration rates were observed in the first block than in the second block for all conditions. Similarly, in the Rare Treasures environments (conditions E25 and E5), we observed a dramatic decrease in exploration rates even though the optimal strategy in these conditions is to explore, and in the Rare Mines environments (conditions C95 and C75) participants continued to explore new keys (with a very small decrease in exploration rates) even though exploiting familiar keys was optimal.

To clarify the meaning of the current results, maximization rates for the last blocks were calculated as a function of the congruence between a strategy driven by the common outcome from exploration and the average outcome. For the congruent conditions (C25, C5, E95, E75), the average maximization rate was 66.15%, while in the incongruent conditions (C95, C75, E25, E5) the average maximization rate reached only 32.22% (the model simulation results were about the same, with 67% and 35% respectively). Therefore, it seems that strong deviation from the optimal exploration level is evident in cases where the common outcome from exploration is misleading with respect to the optimal exploration level.

Alternative Models and Equivalent Number of Observations

The value of the current model, and of the analysis that supports it, could be questioned on the grounds that the high agreement between the observed and predicted exploration rates (a correlation of 0.94) does not mean much. It is possible that other models, including models that assume very different cognitive processes, fit the data better. Apparent support for this critique comes from analysis of a simple "probability matching model" which predicts that exploration rates will be identical to the probability that exploration is successful (e.g, p=0.05 in conditions E5 and C5). The correlation between the prediction of this model and the observed exploration rates is 0.96. Thus, when correlation is used to evaluate accuracy, the probability matching model outperforms the explorative sampler model.

We have two answers to this critique. First, correlation is a poor measure of quantitative accuracy. As was suggested by Erev, Roth, Slonim, and Barron (2007), a more appropriate measure is the model's ENO (equivalent number of observations). The ENO of a model is an estimate of the size of the experiment that has to be run to obtain predictions that are more accurate (in term of mean squared deviation) than the model's prediction.[6] The ENOs for the current model are 29.3 and 47.1 for the first and second blocks respectively. In contrast, the ENOs for the probability matching model are only 0.62 and 2.64 for the first and second blocks respectively (larger is better).

A second answer to the current critique starts with the observation that the cognitive processes implied by the probability matching model discussed above are not very different from the processes assumed by the explorative sampler model. Specifically, like the explorative sampler model, the probability matching model implies the following: (1) backward looking choices; (2) a two-stage decision under backward looking choices (the first between exploration and exploitation, and the second between the alternatives themselves); (3) reliance on a small sample (probability matching implies reliance on a single observation). Our attempts to fit the current results with models that do not share these assumptions have yielded very poor results.

[6] For example, assume that we want to predict the exploration rate of one subject in one condition, and we can use two measures: the ex-ante prediction of the model and the mean of the observed exploration rate over 20 other subjects. If the ENO of the model is 20, the two predictors are expected to be equally accurate.

General Discussion

Previous research suggests that many behavioral problems can be described as the product of deviations from optimal exploration. Some problems appear to reflect insufficient exploration, and other problems appear to reflect excessive exploration. The current analysis tries to improve our understanding of the decision to explore in an attempt to clarify the conditions that lead to over- and under-exploration of new alternatives. Study 1 shows that the co-existence of these contradictory biases can be the product of a tendency to underweight rare events: Under-exploration was documented when the typical outcome from exploration was disappointing (even when exploration was effective on average), and over-exploration was documented when the typical outcome from exploration was reinforcing (even when exploration was counterproductive on average). Study 2 shows a decrease in exploration with experience. A decrease was observed even when a 100% exploration rate was the best strategy and the size of the sample was irrelevant.

These results can be captured with an "explorative sampler" model that quantifies two basic assumptions. The first is a distinction between two cognitive modes that lead to exploration: forward and backward looking. Forward-looking exploration decreases with time, and implies data collection that can be approximated (in the current setting) as random choice. Backward-looking exploration implies sensitivity to past experiences. The second assumption states that the outcome of the backward-looking process reflects reliance on a small sample of past experiences. Study 3 shows that the model provides useful ex-ante predictions of behavior in a wide set of situations.

Organizational Strategy and Implicit Exploration Decisions by Individuals

The basic properties of the decision to explore by individuals, suggested here, are surprisingly similar to the basic properties of the decision to explore by firms (Levinthal & March, 1993, and Gavetti & Levinthal, 2000). Levinthal and his co-authors suggest that like our participants, firms explore in two modes, forward and backward looking, and rely on small samples.

We believe that this similarity reflects two common features of typical exploration problems. The first is the fact that performance tends to improve when the explorer (an individual or a firm) considers the future and learns from the past. Thus, the attempt to improve performance implies the co-existence of backward and forward looking exploration. The second is the fact that there are many reasons for reliance on small samples (Hertwig and Erev, 2009). These reasons include objective constraints (when the event is extremely rare almost any sample size is likely to be too small), cognitive limitations (retrieving large samples is more demanding), and the assumption that the environment is dynamic (when the environment can be in one of many states, reliance on the small set of experiences that occur in similar situations can enhance performance).

The main difference between the current results and the assumed properties of exploration by organization involves the relative importance of the different reasons for reliance on small samples. The organizational learning literature emphasizes the objective constraints. It suggests that rare events are underweighted because most organizations never face them (Levinthal and March, 1993). The leading organizational learning models imply contingent weighting of experienced rare outcomes: Attractive rare outcomes are underweighted even when they are experienced, but unpleasant rare events are overweighted. This pattern, referred to as the hot stove effect (Denrell & March, 2001), is a result of the assumption that extreme negative payoffs dramatically decrease any further updating of beliefs and thus loom larger than positive outcomes.

Our results suggest that the hot stove effect is not very strong in the current context: We observed similar sensitivity to positive and negative rare events. This pattern is captured here with the assumption that the tendency to rely on small samples is a property of the learning process resulting from cognitive limitations and/or beliefs that the environment is dynamic. As a result, the present model implies similar weighting of positive and negative rare events.

Implications for Mainstream Behavioral Decision Research

To clarify the implications of the current results for behavioral decision research, it is constructive to focus on the decisions made in the Rare Mine environments. These decisions involved a choice between the safe status quo (a constant payoff from repeating a previous choice) and a risky gamble with a lower expected value. The leading models of choice behavior predict a tendency to prefer the status quo option. This preference is consistent with many popular theoretical concepts. Examples include: (1) maximization of expected value; (2) risk aversion; (3) loss aversion and the status quo bias (Kahneman, Knetsch & Thaler, 1991); (4) inertia (Cooper & Kagel, 2008); (5) familiarity (Huberman, 2001); and (6) the possibility effect (Kahneman & Tversky, 1979). The present results suggest that these concepts do not provide a good prediction of behavior in the current context. The assumption that the decision to explore reflects reliance on small samples provides more accurate prediction.

Decisions from Experience and Reliance on Small Samples

Most previous studies of the tendency to rely on small samples of experiences focus on binary choice tasks (see Hertwig & Erev, 2009; Ungemach, Chater, & Stewart, 2009). These studies show that a simple abstraction of this tendency facilitates the derivation of learning models with surprisingly high predictive value. Indeed, the large advantage of sampling models over other learning models is one of the clearest outcomes of two recent choice prediction competitions (Erev, Ert, & Roth, 2010; Erev, Ert, Roth et al., 2010).

The current analysis extends this research to address choice in multi-alternative settings. The results highlight the value of a distinction between two modes of exploring new alternatives: forward-looking exploration, which reflects a random data collection process, and backward-looking exploration, which reflects reliance on small samples of past experiences. The analysis of the suggested explorative sampler model shows that both motivations to explore are necessary to capture the experimental findings.

Practical Implications

At first glance, the current results appear to be inconsistent with empirical analyses of exploration by individuals. While our results suggest that excessive exploration is not necessarily less common than insufficient exploration, it is much easier to find empirical demonstrations of insufficient exploration. One simple explanation might be that "Rare Treasure"-like environments are more common in real life than "Rare Mine"-like environments. However, there is another explanation for this apparent asymmetry: Many of the behaviors that reflect too much exploration have been outlawed. The examples of illicit drugs and landmines considered here demonstrate this point.

Better understanding of the decision to explore can be extremely important when law-based solutions are insufficient. Overconsumption, one of the most important problems of our time (Botsman & Rogers, 2010), is an interesting example. Exploration as defined here ("trying a new alternative") incorporates real-world activities such as buying a new product. Moreover, many consumption decisions are similar to the rare mines problem. The common outcome is that a new product will benefit the buyer in some way, but in certain cases we buy products that we do not use, thereby losing money, time and space in our home.[7]

The main argument of this paper is that when deciding whether to explore new alternatives, people often rely on small samples of experiences, which usually consist of the common outcomes. Accordingly, deviations from the optimal exploration level are expected to be observed when behaviors which are driven by the common outcome from exploration are not in line with the optimal strategy.

It is important to note that the suggested explanation for the co-existence of over- and under- exploration does not rule out the influence of other factors that might be involved in exploration decisions (e.g., genetic disposition, social environment, personality characteristics, etc.). To the contrary, it may even be that such factors have a large influence on the payoff structures perceived by decision makers. For example, while one person may find exploration of new medicines rewarding most of the time but leading to severe side effects on rare occasions, another might experience the reverse payoff structure (experiencing light side effects most of the time and, rarely, extraordinary relief). The experimental paradigm used in the current paper overcomes this issue by presenting the same payoff structures to all participants. Our findings

[7] For example, it is estimated that Australians alone spend on average ~9.99 billion USD every year on goods they don't use (i.e., that never even make it out of the box). That is an average of 1,156 USD for each household (Botsman & Rogers, 2010).

show that participants were more sensitive to the common experience with exploration than to the average experience, a general bias which may be reflected in different situations for different people.

In summary, we believe that the co-existence of insufficient and excessive exploration of new alternatives can be a product of the tendency to base backward-looking exploration decisions on small samples of experiences. The two biases appear to contribute to extremely important social problems. We hope that better understanding of the decision to explore new alternatives can help address these problems.

References

Bazerman, M., & Neal, M. (1992). *Negotiating Rationally*. New York: Free Press.

Bechara A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience, 8,* 1458–1463.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York: Guilford.

Botsman, R., & Rogers, R. (2010). *What's Mine is Yours: The Rise of Collaborative Consumption*. New York: HarperCollins.

Carstensen, L. L., Isaacowitz, D., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist. 54,* 165–181.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B, 362*, 933-942.

Cooper, D. J., & Kagel, J. H. (2008). Learning and transfer in signaling games. *Economic Theory, 34,* 415-439.

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

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science, 12,* 523-538.

Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics, 75,* 643-669.

Erev, I. and Barron, G. (2005). On adaptation, maximization and reinforcement learning among cognitive strategies. *Psychological Review, 112(4), 912-931.*

Erev, I., Ert, E., & Roth, A. E. (2010). A choice prediction competition for market entry games: An introduction. *Games, 1,* 117-136.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S., Hau, R. Hertwig, R., Stewart, T., West, R., & Lebiere, C. (2010). A choice prediction competition, for choices from experience and from description. *Journal of Behavioral Decision Making, 23*, 15-47.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making, 21*, 575-597.

Erev, I., Roth, A. E., Slonim, R. L., & Barron, G. (2007). Learning and equilibrium as useful approximations: Accuracy of prediction on randomly selected constant sum games. *Economic Theory***,** 33 **(**1), 29-51.

Erev, I., Wallsten, T. S., & Budescu, D.V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519-527.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*, 659–676.

Gavetti, G., & Levinthal, D. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly, 45*, 113-137.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in real-time dynamic decision making. *Cognitive Science, 27*, 591–635.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to training of complex skills. *Acta Psychologica, 71***,** 147–179.

Hariskos, W., Leder, J., & Teodorescu, K. (2011). A commentary on the market entry competition 2010. *Games, 2*, 200-208.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534–539.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences,* 13, 517-523.

Huberman, G. (2001). Familiarity breeds investment. *Review of Financial Studies, 14,* 659−680.

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E., & Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Consulting and Clinical Psychology, 62,* 295-304.

Janowsky, D. S., El-Yousef, M. K., Davis, J. H. & Sereke, H. S. (1972). A cholinergic adrenergic hypothesis of mania and depression. *Lancet, 19* , 675–681.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5(1)*, 193-206.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263-291.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107*, 397–402.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal, Winter Special Issue, 14,* 95–112.

Loewenstein, G. F. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116(1),* 75–98.

Neuringer, A. (2002). Operant variability: Evidence, functions, theory. *Psychonomic Bulletin & Review, 9*, 672–705.

Palmer C. 2002. ‘Shit happens’: The selling of risk in extreme sport. *The Australian Journal of Anthropology*, *13(3)*, 323–336.

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 534–552.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358–367.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*, 7–59.

Seagull, F. J., & Gopher, D. (1997). Training head movement in visual scanning: An embedded approach to the development of piloting skills with helmet-mounted displays. *Journal of Experimental Psychology: Applied, 3,* 163–180.

Seligman, M. E. (1972). Learned helplessness. *Annual Review of Medicine, 23,* 407–412.

Thompson, L. (1991). Information exchange in negotiation. *Journal of Experimental Social Psychology, 27*, 161–179.

Thompson, L., & Hastie, R. (1990). Social perception in negotiation. *Organizational Behavior and Human Decision Processes, 47,* 98–123.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted, when rare outcomes are experienced (rarely)? *Psychological Science, 20*, 473–479.

Yechiam, E., Erev, I., & Gopher, D. (2001). On value and limitation of emphasis change and other exploration enhancing training methods. *Journal of Experimental Psychology: Applied, 7(4)*, 277–285.

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

Appendix 1:

The basic multi-alternative paradigm described in studies 1 and 2 can be summarized by four main parameters: Low (the lower payoff obtained from exploration), Plow (the probability of getting the payoff Low when choosing to explore), High (the higher payoff obtained from exploration) and Exploit (the constant payoff obtained from exploitation of a familiar key).

In Study 3, we used a random-selection algorithm of the paradigm's parameters to determine the settings of the experiment. We first cast lots of the Exploit and Low parameters (Exploit = uniform distribution between 1 and 12; Low= uniform distribution between 0 and -10) to avoid the possibility of negative or small total payoffs, and set the Plow parameter to range between 0.05 and 0.95. Then, the High parameter was determined such that the expected value from exploration would be equal to the Exploit parameter (H = round [(Exploit-Low*Plow)/(1-Plow)]). Ten games in which the above constraints were met were randomly chosen. Then, to ensure an optimal exploration level, we added 0.5 to the Exploit value in the first five games and subtracted 0.5 from the Exploit value in the other five games. This way, for each Plow value (0.05, 0.25, 0.5, 0.75, 0.95), there was one game in which the optimal strategy was to exploit and one game in which the optimal strategy was to explore.

Accordingly, game no.1, in which Plow=0.05 and the optimal strategy is to exploit, was an extreme version of the Rare Mines condition in Study 1, and game no.10, in which Plow=0.95 and the optimal strategy is to explore, was an extreme version of the Rare Treasures condition in Study 1.

Game	1	2	3	4	5	6	7	8	9	10
Condition's name	C95	C75	C50	C25	C5	E95	E75	E50	E25	E5
Plow	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
Low	-10	-1	-3	-6	0	-7	-8	-2	-1	-4
High	10	15	11	50	100	13	4	8	27	116
EV_explore (=Exploit)	9	11	4	8	5	12	1	3	6	2
Noise	+0.5	+0.5	+0.5	+0.5	+0.5	-0.5	-0.5	-0.5	-0.5	-0.5
Exploit final (Exploit+Noise)	9.5	11.5	4.5	8.5	5.5	11.5	0.5	2.5	5.5	1.5
Optimal strategy	Exploit	Exploit	Exploit	Exploit	Exploit	Explore	Explore	Explore	Explore	Explore