Journal of Behavioral Decision Making, *J. Behav. Dec. Making*, **27**: 109–123 (2014)
Published online 31 May 2013 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.1785

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, Technion–Israel Institute of Technology, Israel

ABSTRACT

The decision whether to explore new alternatives or to choose from familiar ones 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 "decisions from experience" 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 frequent outcome from choosing new alternatives than to the average outcome. That is, the implicit decision whether to explore a new alternative reflects underweighting of rare events: Over-exploration was documented in "Rare Disasters" environments, and insufficient exploration was evident in "Rare Treasures" environments. In addition, the results reveal a decrease in exploration of new alternatives over time even when it is always optimal and some exploration even when it is never reinforcing. These results can be captured with models that share a distinction between "data collection" and "outcome-driven" decision modes. Under the data collection mode, the decision maker collects information about the environment, to be used in future choices. Under the outcome-driven mode, the decision maker relies on small samples of previous experiences with familiar versus unfamiliar alternatives, before the selection of a specific alternative. The predictive value of a two-parameter "explorative sampler" quantification of these assumptions is demonstrated. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS implicit decisions from experience; underweighting of rare events; innovation; learning among cognitive strategies; learned helplessness; softmax; ε-greedy

The decision whether to explore new alternatives or to select familiar ones is implicit in many of our daily activities. For example, each time we surf the Internet, our behavior reflects decisions between entering familiar websites and trying new websites that we never tried before. Similarly, walking involves choices between familiar and new paths, and research implies selection between new and familiar questions and ideas. The current paper tries to shed light on these implicit exploration decisions.

The starting point of the present investigation is the observation that many important behavioral problems can be described as products of deviation from optimal exploration of new alternatives. The best-known examples are problems that have been depicted as reflections of insufficient exploration (sticking to familiar alternatives when it is better to explore new alternatives). One such 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 of new alternatives may be a key ingredient in reducing depression.

Even among healthy individuals, researchers have found that people have a tendency to under-explore new options in a variety of domains. For instance, studies of performance in complex tasks reveal the value of training strategies that enhance exploration of unfamiliar alternatives. One example is the "emphasis change" training protocol (Gopher, Weil, & Siegel, 1989), according to which the scoring rule is changed on a regular basis, forcing trainees to explore new ways to improve performance. This and similar training strategies were found to enhance performance among pilots (Gopher et al., 1989; Seagull & Gopher, 1997), basketball players (www.intelligym.com), and among experimental subjects in a multi-alternative choice task (Yechiam, Erev, & Gopher, 2001).

Similarly, leading negotiation textbooks suggest that enhancing exploration of new possible agreements 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; Erev & Grainer,) and that encouraging them to explore the interests of the other side can lead to better agreements (e.g., Thompson, 1991).

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

There are times, however, when people exhibit the opposite bias, too much exploration. Unsafe sex and the exploration of

*Correspondence to: Kinneret Teodorescu, Max Wertheimer Minerva Center for Cognitive Studies, Technion–Israel Institute of Technology, Haifa, Israel. E-mail: kinneret_w@yahoo.com

untried illicit drugs are obvious examples (Bechara, 2005; Loewenstein, 1994). Another is extreme sports, which increasingly attracts individuals 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 (Landmine Monitor reports, 2010) and in other hiking accidents. One indication for the significance of the tendency to exhibit over-exploration in these examples is fact that this tendency is often addressed by rules or laws designed to eliminate exploration.

The coexistence of over- and under-exploration of new alternatives (the tendency to exhibit over-exploration in some settings and under-exploration in others) 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 searching was costly—even though they were given a 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 search cutoff in conjunction with sensitivity to other factors, not correlated with the optimal cutoff.

Levinthal and March (1993), who considered exploration in the context of organizational strategy, suggested that organizations tend to exhibit insufficient exploration (e.g., do not invest enough in Research & 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.

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 important aspects of real-life examples of over- and under-exploration similar to the ones discussed earlier. For example, a depressed individual is not likely to decide explicitly between exploration of new activities and engagement in familiar 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 to the current context of explication by Zwick et al. (2003) of their results. It assumes that the coexistence of under- and over-exploration is the *sole* result of a random component in the decision process. Underlying such random behaviors, or "noise," can be the stochastic nature of choice behavior (Erev, Wallsten, & Budescu, 1994; Thurstone, 1927) as well as arbitrary search for information about the environment (Cohen, McClure, & Yu, 2007; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). According to this explanation, even though people's behavior is assumed to be generally guided by the optimal strategy for a given context (Payne, Bettman, & Johnson, 1988), noisy responses can still lead to under-exploration of new alternatives when exploration is always optimal and over-exploration when exploration of new alternatives is never optimal. 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.

The second explanation can be described as an adjustment of Levinthal and March's (1993) assertions to implicit individual decisions. It assumes that the coexistence of under- and over-exploration is a reflection of the tendency to rely on small sets of experiences from similar situations (Fiedler, 2000; Gonzalez, Lerch, & Lebiere, 2003; Hertwig, Barron, Weber, & Erev, 2004; Kareev, 2000; Hau, Pleskac, Kiefer, & Hertwig, 2008). Reliance on a small set of experiences implies overweighting of the frequent outcomes and underweighting of rare events: Naturally, rare outcomes are rarely included in a small sample; hence decisions are usually driven by the more frequent outcomes (Barron & Erev, 2003). Thus, it can lead to insufficient exploration when the probability of success (in a given exploration effort) is low and to excessive exploration when the probability of success is high (The exact relationship of the current hypothesis to Levinthal and March (1993) is clarified in the general discussion).

The two explanations considered here—mere noise and reliance on small samples—are expected to affect behavior in a wide set of situations, including complex natural settings as well as simple laboratory contexts. Here, we chose to compare them in a simple experimental environment that allows for precise manipulations of payoff structures and clear tractability of the optimal behavior. The simplified environments considered in the current experiments simulate situations in which exploration of new alternatives demands effort and/or entails satisfaction, whereas selection of familiar alternatives does not. To demonstrate this simplification, imagine a fisherman lifting rocks around a pond in search of live baits. In order to lift a rock for the first time, he has to loosen the rock out of the ground, thus investing effort. Under each rock, he may find a worm or nothing at all. However, each rock also holds the possibility of a hidden treasure (lost gold coin) or a hidden danger (an angry scorpion). Whatever is under a rock will not remain there once it has been discovered (a worm or gold will be collected by the fisherman

DOI: 10.1002/bdm

and the scorpion will sting the poor fisherman and either run away or be killed). Therefore, exploration of new alternatives (rocks) differs from revisiting alternatives that have already been explored.

The paper is organized as follows: In Studies 1 and 2, we examine environments that enable us to examine the two hypotheses described earlier. Then we describe a simple model that is able to account for the results while also making quantitative, *ex ante* predictions for other possible environments. In Study 3, a spectrum of payoff structures is examined, and the predictions of the model are tested. Last, our findings and their general implications are discussed.

STUDY 1: RARE TREASURES AND RARE DISASTERS ENVIRONMENTS

Method

Participants

Twenty Technion students (9 women and 11 men, with an average age of 24 years) participated in the experiment in return for a performance-based payment. They received a show-up fee of 20 NIS (New Israeli Shekels. Approximately \$5.5 at the time of the experiment). 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 multi-alternative choice task. The alternatives were 144 unmarked keys presented in a 12×12 matrix (Figure 1). In each trial, participants select one key, and their choice is followed by an immediate presentation of the trial's payoff on the selected key. The payoff associated with each key could be either a gain or a loss, as described subsequently, but only when the key is selected for the first time; 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.

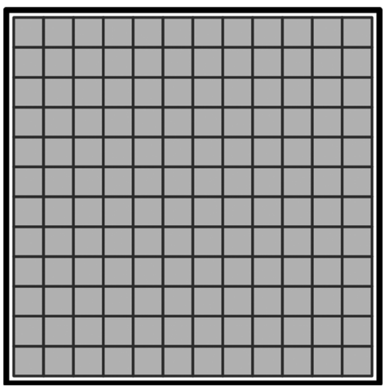

Figure 1. The matrix of keys presented to participants at the start of each trial

Two payoff structures were used. In the "Rare Treasures" condition, the initial payoff of each alternative produced a loss of 1 NIS (about \$0.28) with probability .9 and a gain of 10 NIS otherwise. Thus, the expected payoff from exploration of new keys was positive ($10(.1) - 1(.9) = +.1$). As the payoff from repeating a choice was 0, the optimal strategy was to explore new keys. In the "Rare Disasters" condition, the initial payoff of each key produced a gain of 1 NIS with probability .9 and a loss of 10 NIS otherwise. The expected payoff from exploration in this case was negative ($1(.9) - 10(.1) = -.1$); thus, exploration of new keys was costly in the long run. As noted, participants did not receive any description of the payoff structure and were blind to the title of each condition. Their only instruction was to select one out of the 144 keys in each trial while trying to maximize their payoffs.

Experimental design

The experiment used a within-subject design; each participant took part in the two conditions described earlier. The order of the two conditions was counterbalanced across participants. At the beginning of the experiment, participants were informed that they would play two distinct games of 100 trials each, and that their task was 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

As noted, the optimal strategy in the Rare Disasters condition is never to explore new keys and in the Rare Treasures condition is always to explore new keys. As this information is not given to participants ahead, optimal learning should lead them closer to the optimum as time passes. The two hypotheses considered here predict that participants will not reach the optimum even after accumulating many experiences. However, the mere noise hypothesis assumes that these deviations will be only as a result of noise, and therefore the general direction will be toward the optimal strategy. In contrast, the reliance on small samples hypothesis predicts that learning will be influenced more by the frequent experience than by the optimal strategy. Therefore, the two hypotheses lead to contradictory predictions about the relationship between the two experimental conditions: The mere noise hypothesis predicts more exploration in the Rare Treasures condition (where exploration of new keys is optimal) than in the Rare Disasters condition (where it is optimal to select familiar keys). The reliance on small samples hypothesis leads to the opposite prediction: As rare experiences are less likely to be included in a small sample, participants' behavior is expected to be guided by the more frequent experience. Accordingly, this hypothesis predicts more exploration in the Rare Disasters condition (where the frequent outcome from exploration of new keys is positive) than in the Rare Treasures condition (where the frequent outcome from exploration of new keys is negative).

DOI: 10.1002/bdm

Results

The data analysis was performed with respect to the percentage of trials in which participants tried a new key, henceforth referred to as "exploration rate." This choice of dependable variable constitutes a focus on "easy to observe" indications for exploration (it is possible, of course, that in some cases people can also explore the effect of reselecting a familiar alternative. Such exploration efforts are not captured by the current measure, but are discussed in the "Implications for descriptive models" section). In order to probe the learning process throughout the task, each condition was divided into five blocks, consisting of 20 trials apiece. For each participant, exploration rates (percentage of trials in which participants selected a new key) were calculated for each block. Figure 2(A) presents the main experimental results.

We conducted 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 Disasters versus Rare Treasures). The results revealed a main effect of condition, with significantly higher exploration rates in the Rare Disasters condition, in which it was not optimal to explore, compared with the Rare Treasures condition, in which exploration of new keys was the optimal strategy ($F(1, 19) = 1.63$; $p < .01$, $\eta_p^2 = .36$). 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 Disasters and Rare Treasures conditions respectively ($p < .001$). This pattern suggests that the coexistence of over- and under-exploration of new alternatives can be 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 ($F(4, 76) = 6.25$; $p < .001$, $\eta_p^2 = .25$): the observed exploration rates decreased with time. However, as can be seen in Figure 2(A), 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 Disasters condition. This interaction effect between condition and block was significant ($F(4, 76) = 3.28$; $p < .05$, $\eta_p^2 = .15$).

Figure 3(A) presents the individual exploration rates in the last block as a function of the average payoff from exploration of new keys in previous blocks. The results reveal an increase in exploration rates as a function of the experienced average payoff within each condition, but comparing the means 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 (+.08 vs. −.32), but exhibited lower exploration rates (31% vs. 69%), compared with the Rare Disasters condition. When the two outliers in the Rare Disasters condition

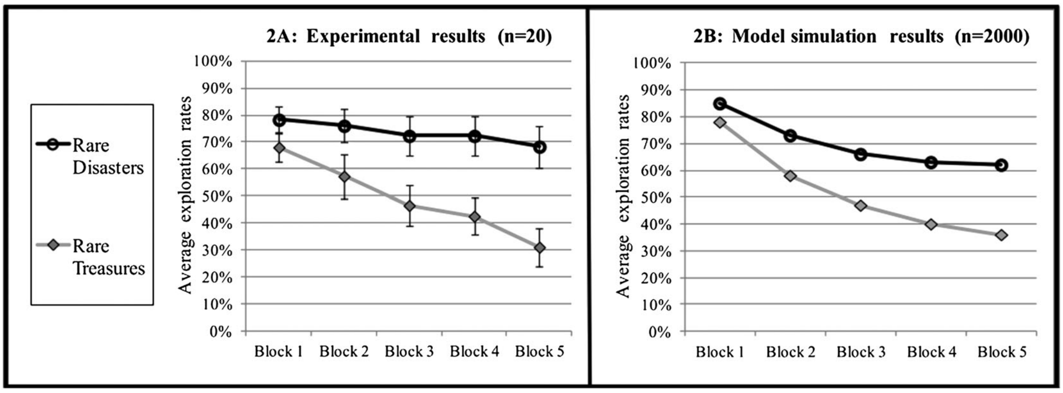

Figure 2. Average exploration rates by blocks of 20 trials. The left side (A) displays the experimental results with SE bars, and the right side (B) displays the results obtained from a simulation of the explorative sampler model, presented after the discussion of Study 2

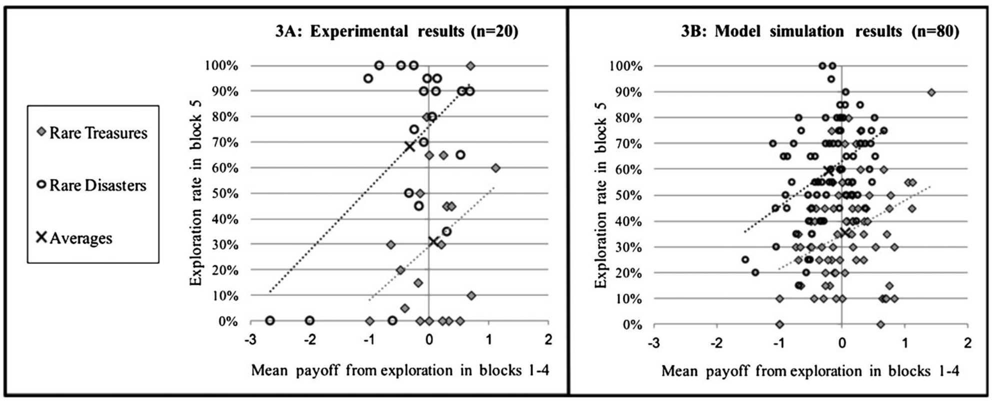

Figure 3. Exploration rates in the last block as a function of the mean payoff from exploration in previous blocks. Each dot represents one participant, the *X*s represent the averages over participants in each condition, and the lines represent the fitted linear (regression) trends. The left side (A)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

DOI: 10.1002/bdm

(participants who experienced a mean payoff of −2 and lower) are removed, the qualitative result remains: Now participants in the Rare Disasters condition experienced on average mean payoff of −.1 and exhibited exploration rates of 76%.

Figure 4 presents the exploration rates of each of the 20 participants in five blocks of 20 trials in each problem. The results reveal relatively gradual changes over time. That is, there is no indication for a common use of "all or none" strategy (exploring new alternatives on all trials at the beginning and at a certain point abandoning exploration and switching to selection of familiar keys for the remaining trials). In addition, Figure 4 shows that most participants exhibited the aggregated pattern: 15 of the 20 participant explored more alternatives in the Rare Disasters than in the Rare Treasures condition. Two participants exhibited the opposite pattern (subjects 2 and 13), and three participants exhibited similar exploration rates in both conditions (subjects 9, 16, and 19).

Finally, Figure 4 shows that the order of playing the two problems did not have a clear effect. The exploration rates of the subjects that played the Rare Disasters Problem first (the odd numbers) is similar to the exploration rates of the subjects that played the Rare Treasure problem first (the even numbers).

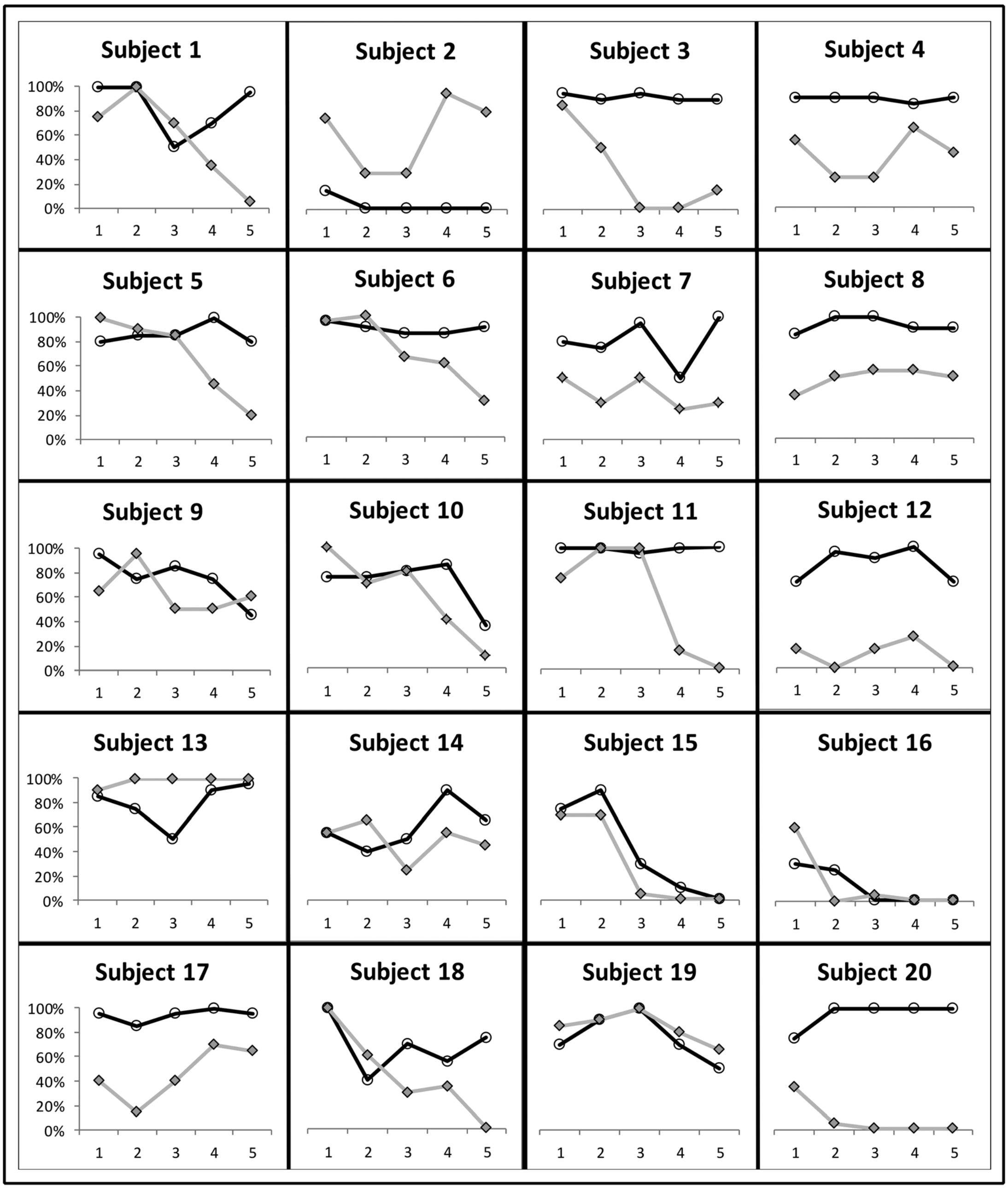

Figure 4. Individual exploration rates over the five blocks of each condition. Diamonds with gray lines represent exploration rates in the Rare Treasure condition, and circles with the black lines represent exploration rates in the Rare Disasters condition. Participants with odd numbers (the first and third columns from the left) played the Rare Disasters game first and the Rare Treasure game second, whereas participants with even numbers (the second and fourth columns from the left) experienced the opposite order of games

DOI: 10.1002/bdm

Discussion

In this study, we examined two payoff structures in which reliance on the frequent outcome from exploration of new alternatives (the more probable outcome) leads to suboptimal exploration levels (defined according to the expected value from exploration of new keys). The results suggest that exploration decisions reflect higher sensitivity to the frequent outcome than to the average outcome. Participants exhibited insufficient exploration of new keys when the frequent outcome from exploration was disappointing (the payoff "−1" in the Rare Treasures condition) despite the higher mean payoff of exploration and explored too many new keys when the frequent outcome was reinforcing (the payoff "+1" in the Rare Disasters condition) despite the lower mean payoff of exploration. 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 among the available alternatives: As such, 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).

Study 2 was designed to improve our understandings of the indications for random choices. Its main goal is to examine if these indications are stable over trials or decrease with experience. A decrease is expected under the natural assumption that the indications for random choice are reflections of an attempt to collect data and the need for data collection decreases with time (see a similar assumption in Hariskos, Leder, & Teodorescu, 2011). To this end, Study 2 focuses on simple deterministic environments, in which the payoff resulting from exploration of new keys is constant. In such simple environments, choices which are driven by previously obtained outcomes from exploration imply maximization even if people rely on small samples; thus, any deviations from the optimal strategy can only be attributed to random noise and/or data collection behaviors.

STUDY 2: EXPLORATION IN DETERMINISTIC SETTINGS

Method

Participants

Twenty Technion students (6 women and 14 men, with an average age of 24 years) 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, approximately \$8). 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 12×10 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 "one optimal strategy," and decisions which are based on outcomes from the current environment (driven by the frequent and/or the average outcome) will always result in a random choice between exploration of new alternatives and selection of familiar alternatives (keys that were selected in the previous trials).

Condition "Explore+1": Selection of a new alternative results in a payoff of +1, whereas selection of a familiar alternative results in a payoff of zero. In this condition, the optimal strategy is to keep exploring new alternatives, and as each strategy has a constant payoff (+1 for new keys, 0 for familiar ones), reliance on outcomes from the current experimentation will always lead to the optimal strategy.

Condition "Explore−1": Selection of a new alternative results in a payoff of −1, whereas selection of a familiar alternative results in a payoff of zero. Again, reliance on previous experiences will always lead to the optimal strategy, which in this condition is to select a key that was already selected in previous trials.

Condition "Explore−10": Selection of a new alternative results in a payoff of −10, whereas 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 random choices and/or data collection tendencies 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 ($F(3, 57) = 22.24$, $\eta_p^2 = .54$; $F(4, 76) = 65.63$, $\eta_p^2 = .77$; $F(12, 228) = 6.14$, $\eta_p^2 = .24$; respectively, with $p < .001$ for all effects). The average exploration rates in the four conditions are presented on the left side of Figure 5.

As can be seen in the graph (Figure 5(A)), 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 although the Explore−1 condition differed highly

DOI: 10.1002/bdm

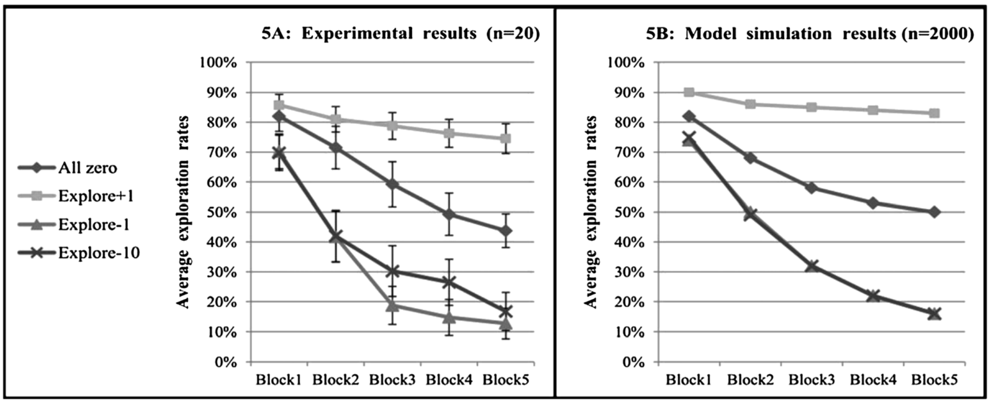

Figure 5. Average exploration rates by blocks of 20 trials. The left side (A) displays the experimental results with SE bars, and the right side (B) displays the results obtained from a simulation of the explorative sampler model, presented after the discussion of Study 2

significantly from the All zero and Explore+1 conditions ($p < .001$ for both), there was no significant difference between the Explore−1 and Explore−10 conditions ($p = .85$).

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 a review in Erev & Barron, 2005). At the same time, the participants were highly sensitive to whether or not exploration of new keys was optimal. They learned to keep exploring new keys when exploration was beneficial (Explore+1) and to explore much less new keys when it was costly (Explore−1 and Explore−10). These results are consistent with animal studies, which show that variability of responses (selecting a different response option each time) 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, any outcome-driven 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 outcome-driven mechanism in general can explain this deviation from the optimal behavior. This deviation and the decreased exploration rates observed in all conditions, however, could be an indication for a certain amount of random choices. These random choices (in later trials) can be a result of two main factors: data collection tendencies, resulting from examination of the possibility that the payoff structure has been changed (believes about a dynamic nature of the environment), and/or errors resulting from misjudgments about the selected key (whether it was a new key or a familiar one). Although random choices by themselves cannot explain the results of Study 1 (where deviation from optimal exploration could be accounted for by outcome-driven choices), the results of Study 2 show that regression to the mean also plays an important role in the current setting.

As noted before, random choices between all keys can lead to a decrease in exploration rates (as there are more new keys at the beginning than toward the end of the task). Therefore, a fixed random choice rate over trials can cause to a decrease in exploration rates over time. However, it is important to note that the observed decline in exploration rates (especially in the Explore−1 and the Explore−10 conditions, in which the selection of new keys cannot be attributed to outcome-driven choices) is greater than what is expected under a fixed random choice rate. These steep declines in exploration rates suggest that the indication for random choices decreases with time.[1]

IMPLICATIONS FOR DESCRIPTIVE MODELS

Most descriptive models of learning abstract exploration as a stochastic component in the decision mechanism. The common ways to capture exploration are the "softmax" and the "ε-greedy" rules (Daw et al., 2006). Under the softmax rule, people tend to select the alternatives that lead to the highest average reinforcement, but in some cases, they "explore" one of the other alternatives. The probability of exploration depends on the magnitude of the difference between the average reinforcement scores. Thus, exploration is high when all alternatives lead to similar average reinforcements and low when one of the alternatives is much better. The ε-greedy rule assumes fixed or diminishing probability of random exploration that does not depend on the average reinforcement.

These abstractions provide a useful approximation of behavior when the number of alternatives is small, but they fail to capture behavior in the current setting. For example, these abstractions cannot capture the large difference between conditions Explore−1 and Explore+1 in Study 2: As all the alternatives were associated with similar reinforcements, the softmax abstraction implies similar exploration rates in the two conditions. The ε-greedy abstraction fails because it implies fixed (or decreased) exploration rate independently of the payoffs.

We chose to capture the results with a model that quantifies the two hypotheses mentioned earlier that were supported by the experimental results: reliance on small samples and decreasing

[1] Assuming a fixed rate of random choice, the ratio between the exploration rate at trial $t[E(t)]$ and the rate of unfamiliar alternatives at that trial $UF(t)$ should not change after $t = 2$. The results show clear deviation from this prediction. For example, in conditions Explore−1 and Explore−10, $E2/UF2 = .95/(119/120) = .96$, whereas $E100/UF100 = .15/(86/120) = .2$.

DOI: 10.1002/bdm

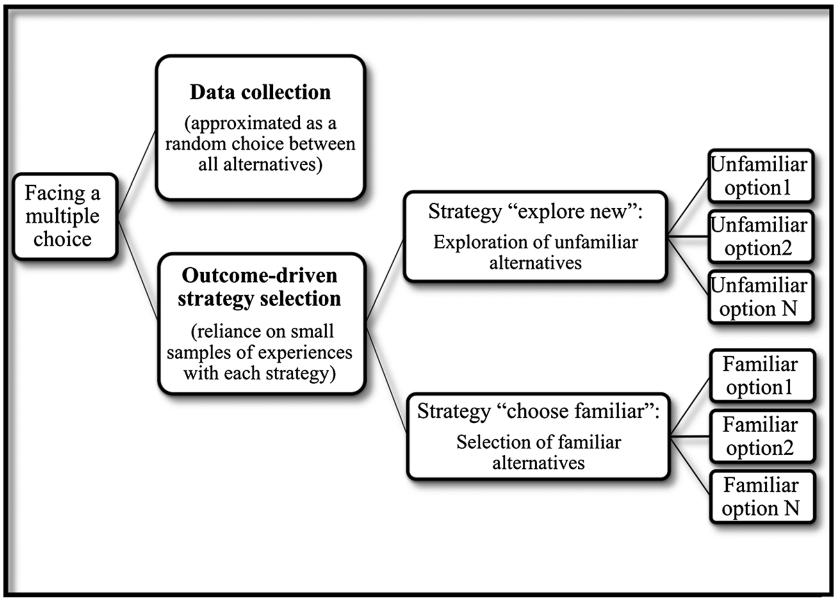

Figure 6. An illustration of the three-stage explorative sampler model

random choices. The decreasing randomness hypothesis is similar to the ε-greedy abstraction (with decreasing exploration rate); however, in the current model, exploration is driven not only by a random component but also by reliance on small samples. The reliance on small samples hypothesis implies that the tendency to explore is determined by previous outcomes from exploration. Accordingly, when decision makers rely on previous outcomes, they first consider experiences with exploration of new alternatives versus selection of familiar ones, followed by a choice between the alternatives themselves (see similar ideas of selection between strategies in Busemeyer & Myung, 1987; Yechiam et al., 2001; Rieskamp & Otto, 2006; Koehler & James, 2010).

In the following section, we present a two-parameter abstraction of these hypotheses. The current model is a modification of a simplified variant of the explorative sampler model (Erev, Ert, & Yechiam, 2008) [2], which provided the best prediction of behavior in two-alternative choice tasks under an "experience-repeated" setting, examined in the Technion choice prediction competition (Erev et al., 2010b). The modification involves the addition of the initial sub-decision between new and familiar alternatives.

A three-stage explorative sampler model

The model assumes that decision makers first choose between two cognitive modes: "data collection" or "outcome-driven" modes (see Figure 6 for illustration). The probability to use the data collection mode 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 exact probability is $P(Collect)_t = C_i^{\frac{t-1}{T}}$, where $0 > C_i > 1$ is a trait of participant i that captures the tendency to collect data. The choice among the alternatives under the data collection mode is approximated as random choice. The abstraction of data collection behaviors as a random choice is a simplification assumption motivated by an attempt to reduce the number of free parameters. Evaluation of the experimental results shows that most subjects exhibited, at times, sequences of systematic search (e.g., from the top left of the matrix to the top right). It is also possible that some subjects returned to familiar keys by mistake, or in order to collect data and/or to check if the payoff has been changed. In the current context, when the payoffs distribution associated with the keys were not correlated with the key's location and the payoffs change only once, the random search simplification is not too costly; it captures the joint effect of these behaviors. Under one abstraction of the motivation for data collection, it reflects older experiences (prior to the current experiment) that demonstrate the value of this effort.

When the agent relies on previous experiences with the current task (under the outcome-driven mode), she or he draws (with replacement) S_i past outcomes from exploration of new keys versus S_i outcomes from selection of familiar keys ($S_i > 0$ is a trait of the participant that represent the size of the sample) and selects the strategy with the highest sample mean (and randomly in the case of a tie). The choice among the alternatives themselves (i.e., which familiar or unfamiliar alternative to select) follows a similar logic. However, the result of this choice does not affect the exploration measure considered here (exploration rate of new alternatives).

The right-hand boxes in Figures 2, 3, and 5 illustrate the predictions of the three-stage model under the assumption that the traits are drawn from uniform distributions[3]: C_i from

[2]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.

[3]The assumption that the traits are drawn from a uniform distribution is a simplification that was used in the original version of the model. It might be that other distributions can also account for the results and perhaps provide more insights into individual differences. Indeed, Figure 3 suggests that the proposed model under-predicts the variance in the data. However, a precise account of individual differences is beyond the scope of the current paper.

DOI: 10.1002/bdm

$U(0, \phi)$ and S_i from $\{1, 2, \ldots \mu\}$. The predictions of the model were derived using a computer simulation, and the two free parameters were set (to fit the data) at $\phi = .25$ and $\mu = 8$. The results show that the three-stage model reproduces the following observations: (i) over-exploration of new keys in the Rare Disasters condition; (ii) insufficient exploration of new keys in the Rare Treasures condition; (iii) higher exploration rates in the Rare Disasters than in the Rare Treasures condition; (iv) a decrease in exploration of new keys over time; (v) a sharper decrease in exploration of new keys with time in the Rare Treasures condition; (vi) higher sensitivity to the frequent payoff than to the average payoff; and (vii) a similar pattern of exploration rates in the Explore−1 and Explore−10 conditions.

Notice that the current model reflects the joint effect of the two factors considered earlier: stochastic choice and reliance on small samples. In order to evaluate the contribution of each factor, we compared the full model with simpler variants assuming that only one of the factors drives behavior. The first variant is an abstraction of the mere noise hypothesis. It is captured by the three-stage model with the constraint that the sample size is very large (i.e., $S_i \geq 1000$). When the sample is large, only the random component can cause to deviations from optimal exploration, in accordance with the mere noise hypothesis. This restricted version of the model (with $S_i \geq 1000$ and only one free parameter—C_i) produces less than 50% exploration rate in the last block of the Rare Disaster condition and therefore cannot account for aforementioned observations 1, 3, 5, and 6.

The second restricted model eliminates the data collection mode and assumes that all decisions are based on small samples of past experiences. This restricted version of the model (with $C_i = 0$ and only one free parameter—S_i) can fit the main results on Study 1, but it fails in Study 2. For example, it predicts exploration rates below 50% in the first block of the Explore−1 and the Explore−10 conditions, and an increase in exploration rates over time in the Explore+1 condition (therefore cannot account for aforementioned observation 4).

In summary, both components of the two-parameter explorative sampler model are necessary to capture the current data. Only the full model captures all seven qualitative phenomena documented earlier. In addition, the two-parameter model provides much better quantitative fit for the mean exploration rates. The average mean square deviation (MSD) between the observed and reproduced exploration rates across all conditions for the two-parameter model is .003. The MSD scores of the restricted one-parameter versions are much higher (.020 for the mere noise model and .030 for the only sampling model).

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). The next study was designed to address this possibility. Study 3 examines the *ex ante* predictive value of the modified explorative sampler model in a broader set of payoff structures, in which the congruence between the frequent experience with exploration of new keys and the average experience is varied.

STUDY 3: EXAMINING *EX ANTE* PREDICTIONS IN VARIOUS ENVIRONMENTS

Method

Participants

Forty Technion students (23 women and 17 men, with an average age of 25 years) 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 the Appendix). Figure 7 presents the 10 conditions randomly chosen according to this algorithm:

In the condition names, "C" and "B" reflect whether exploration of new keys is "Costly" or "Beneficial" in the long term according to expected values, whereas the number represents the probability of obtaining a higher payoff from exploration of new keys compared with the constant payoff obtained from selection of familiar keys. Where this number is higher than 50 (the probability to receive a higher payoff from exploration >.5), the frequent outcome from selecting a new key (shown in bold in the figure) is better than from selecting a familiar key; where it is lower than 50, the frequent outcome from exploration is worse.

Notice that in conditions C95 and C75, the frequent outcome from exploration of new keys is better than that from selection of familiar keys, but exploration is costly in the long run. Therefore, these conditions are different versions of the Rare Disasters condition from Study 1. Similarly, conditions B25 and B5 are different versions of the Rare Treasures condition from the first study, as the frequent outcome from exploration is worse, but exploration is beneficial in the long run.

Each condition consisted of 60 alternatives and 50 trials. Each participant experienced all 10 conditions, with the order randomly counterbalanced across participants. Between conditions, participants were informed that the instructions for the following game (condition) are unchanged, but the payoff structure would be different. Unlike the first studies, 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. 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 or she has earned enough money and so no longer needs to pay attention to the task. As 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).

DOI: 10.1002/bdm

Condition	Selection of familiar alternatives	Exploration of new alternatives	Expected value from exploration
C95: Costly .95 (Rare Disasters)	9.5	(**10**,.95; -10)	9
C75: Costly .75 (Rare Disasters)	11.5	(**15**, .75; -1)	11
C50: Costly .50	4.5	(11, .50; -3)	4
C25: Costly .25	8.5	(50, .25; **-6**)	8
C5: Costly .05	5.5	(100, .05; **0**)	5
B95: Beneficial.95	11.5	(**13**, .95; -7)	12
B75: Beneficial.75	0.5	(**4**, .75; -8)	1
B50: Beneficial.50	2.5	(8, .50; -2)	3
B25: Beneficial.25 (Rare Treasures)	5.5	(27, .25; **-1**)	6
B5: Beneficial.05 (Rare Treasures)	1.5	(116, .05; **-4**)	2

Figure 7. The 10 randomly selected conditions. In each condition, exploration is either costly or beneficial in the long run (represented by C or B in the condition name), and there is a .95 to .05 probability of obtaining a higher payoff from exploration of new keys than selection of familiar ones (represented by the digits in the condition name). The columns show the payoffs and probabilities for selecting new or familiar keys and the expected value from exploration of new keys; in the middle column, the frequent outcome from exploration is shown in bold. For example, in the last condition—B5—pressing a familiar key always results in a payoff of 1.5, and exploring a new key results in a payoff of +116 with probability .05 and −4 otherwise. The expected value from exploration in this condition is 2. Thus, exploration is beneficial in the long run, although the frequent experience with exploration is disappointing (−4 compared with 1.5)

Predictions

The left-hand columns in Figure 8 present the predictions of the three-stage explorative sampler model for the current study. The upper bar graph show the prediction of the model to the five games, in which exploration of new alternatives is costly in the long run and the lower graph present the games in which exploration was beneficial. 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 Studies 1 and 2) participate in the new 10 conditions of Study 3.

As Figure 8 shows, the model predicts that the coexistence of over- and under-exploration of new alternatives, 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 Disasters environments) than in conditions B25 and B5 (Rare Treasures environments). As noted before, in these conditions, greater sensitivity to the frequent outcome (compared with the average outcome) results in suboptimal exploration rates: that is, under-exploration in conditions B5 and B25 and over-exploration in conditions C75 and C95.

As seen in Figure 8, the model implies that the contingent decrease in exploration over time, discussed earlier, 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 10 conditions. In addition, the model predicts a greater decrease when the frequent 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 ($F(9, 351) = 15.85$, $\eta_p^2 = .29$; $F(1, 39) = 115.64$, $\eta_p^2 = .75$; $F(9, 351) = 6.42$, $\eta_p^2 = .14$, respectively, with $p < .001$ for all effects). The right-hand side of Figure 8 presents the observed exploration rates for each of the 10 conditions over the two blocks. The results reveal high correspondence with 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 .94.[4]

As predicted, higher exploration rates were observed when the frequent experience with exploration was rewarding than in conditions where the frequent 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 frequent outcome from exploration was reinforcing (an extreme version of the Rare Disasters condition), and explored new keys in only 36.9% of the trials in condition B5, in which exploration was optimal but the frequent 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 ($p < .001$). In addition, the gap between exploration rates in conditions C75 and B25 (moderate versions of the Rare Disasters and Rare Treasures conditions, respectively) reached 22% and was also highly significant ($p < .0001$).

An examination of the observed exploration rates over the two blocks (25 trials per block, for each 50-trial game) reveals that the general 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 B25 and B5), we observed a dramatic decrease in exploration rates even though the optimal strategy in these conditions is to explore new

[4]Notice that the correlation of the observed results with the probability to receive a higher payoff from exploration is even higher (.96). Yet a simplified model that predicts a matching of exploration rate to the probability of success (e.g. Ci = 0 and Si = 1) fails to capture other features of the data such as the high initial exploration rates.

DOI: 10.1002/bdm

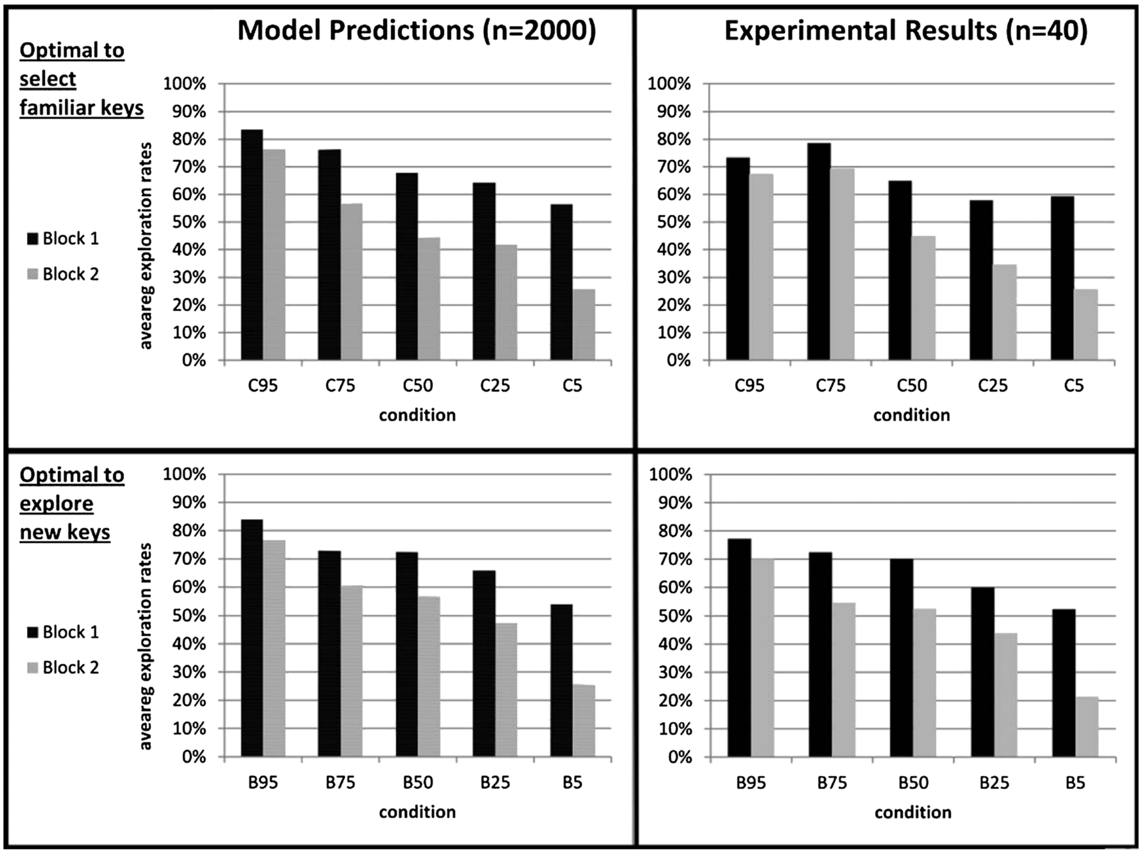

Figure 8. The left side displays the model's *ex ante* predictions on the basis of 2000 simulations with the same value for the trait parameters as in the first two studies (without any fitting procedure). The right side displays the average exploration rates over the two blocks (of 25 trials each) across all subjects ($n = 40$) in each of the 10 conditions. The upper bar graphs present the conditions in which exploration of new alternatives is costly, and the lower graphs present the conditions in which exploration is beneficial. Black bars indicate exploration rates in the first block, and gray bars represent exploration rates in the second (last) block

keys, and in the Rare Disasters environments (conditions C95 and C75) participants continued to explore new keys (with a very small decrease in exploration rates) even though selecting 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 frequent outcome from exploration and the average outcome. For the congruent conditions (C25, C5, B95, and B75), the average maximization rate was 66.15%, whereas in the incongruent conditions (C95, C75, B25, and B5), the average maximization rate reached only 32.22% (the model simulation results were about the same, with maximization rates of 67% and 35%, respectively). Therefore, it seems that strong deviation from optimal exploration is evident in cases where the frequent outcome from exploration is misleading with respect to the optimum.

Equivalent number of observations

In order to clarify the predictive value of the current model, we computed its equivalent number of observations (ENO; Erev, Roth, Slonim, & Barron, 2007). The ENO of a model is an estimate of the size of the experiment that has to be run to obtain estimations that are more accurate (in terms of MSD) than the model's prediction. 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. The ENOs of the current model are 47.1 and 32.3 for the first and second blocks, respectively. These values are comparable with the ENO of the best models in recent choice prediction competitions (Erev et al., 2010b). The ENOs of the one-parameter simplifications of the model are 6.5 and 3.0 (in the first and the second blocks) for the mere noise variant ($S_i > 1000$), and 4.4 to 8.3 for the "only small samples" variant ($C_i = 0$). Thus, both components are necessary to derive useful *ex ante* predictions.

GENERAL DISCUSSION

Previous research suggests that many behavioral problems can be described as the product of deviations from optimal exploration of new alternatives. 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 elemental conditions that lead to over- and under-exploration of new alternatives. Study 1 shows that the coexistence of these contradictory biases can be the product of a tendency to underweight rare events and overweight the frequent outcomes: Under-exploration was observed when the frequent outcome from exploration was disappointing (but exploration was beneficial on average), and over-exploration was observed when the frequent outcome from exploration was reinforcing (but exploration was costly on average). Study 2 shows a decrease in exploration rates 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 a three-stage model, which quantifies two basic ideas. First is the assumption that agents first decide whether to collect data for future choices or whether to rely on previous outcomes from exploration. The second assumption is that when relying on past experiences, agents first decide whether to explore new alternatives or to select familiar ones, and only then decide between the specific options themselves. Although data collection behaviors can be approximated as random choices, outcome-driven choices involve reliance on small samples. Study 3 shows that the model provides useful *ex ante* predictions of behavior in a wide set of payoff structures.

Organizational strategy and implicit exploration decisions by individuals

The basic properties of the decision to explore by individuals, as suggested here, are surprisingly similar to the basic properties of the decision to explore by firms (Levinthal & March, 1993; Gavetti & Levinthal, 2000). In accordance with our results, Levinthal and his colleagues suggest that firms try to explore in two modes, forward-looking (equivalent to the current "data collection" choices) and backward-looking (equivalent to the current "outcome-driven" choices) and rely on small samples.

We believe that this similarity reflects two features of typical exploration problems, which are common to both individuals and firms. The first is the fact that performance tends to improve when the explorer (an individual or a firm) takes the future into consideration and learns from previous experiences with exploration. Thus, the attempt to improve performance implies the coexistence of backward- and forward-looking exploration. The second is the fact that there are many reasons for reliance on small samples (Hertwig & 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 from similar situations can enhance performance).

The main difference between the current results and the assumed properties of exploration by organizations 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 & 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.

In our experiments, the hot stove effect was not very strong as we observed similar sensitivity to positive and negative rare events. This pattern is captured here with the assumption of reliance on small samples of past experiences, which could result from cognitive limitations and/or beliefs that the environment is dynamic. As such, 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 Disasters 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, such as (i) maximization of expected value; (ii) risk aversion; (iii) loss aversion and the status quo bias (Kahneman, Knetsch, & Thaler, 1991); (iv) inertia (Cooper & Kagel, 2008); (v) familiarity (Huberman, 2001); and (vi) the possibility effect (Kahneman & Tversky, 1979). Our findings suggest that these concepts do not provide a good prediction of behavior in the current context. Rather, the assumption that the decision to explore reflects reliance on small samples appears to provide more accurate predictions.

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 (Hertwig & Erev, 2009; Ungemach, Chater, & Stewart, 2009; Rakow & Newell, 2010). 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, 2010a; Erev, Ert, Roth, et al., 2010b).

The current analysis extends this research to address choice in multi-alternative settings. The results demonstrate that the models that best predict binary decisions in the choice prediction competitions do not provide good predictions of behavior in the current multi-alternative setting. Yet the addition of one assumption, an outcome-driven choice among strategies, is sufficient to eliminate this gap.

DOI: 10.1002/bdm

Practical implications

At first glance, the current results appear to be inconsistent with empirical analyses of exploration by individuals. Although our results suggest that excessive exploration is not necessarily less common than insufficient exploration, it is much easier to find empirical demonstrations of the latter. One simple explanation might be that Rare Treasures-like environments are more common in real life than Rare Disasters-like environments (because in many cases, exploration of new alternatives demands effort, which is not followed by an immediate reinforcement). 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 process responsible for the decision whether to explore new alternatives 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 Disasters problem: The frequent outcome is that a new product will benefit the buyer in some way, but in rare occasions, a new purchase might lead to a negative result (because of the waste of time, money, and/or space in our home[5]).

Summary

The current analysis suggests that the implicit decision between "exploring new alternatives" and "selecting familiar alternatives," in multi-alternative choice tasks, is similar to explicit decisions from experience in binary choice tasks. The main deviation from optimal exploration can be described as the product of reliance on small samples of past experiences that lead to underweighting of rare events from exploration. A bias toward insufficient exploration is observed when the frequent outcome from exploration is disappointing, and a bias toward excessive exploration is observed when the frequent outcome from exploration is reinforcing. We believe that this observation can shed light on the processes that underlie decisions from experience in natural multi-alternative settings.

APPENDIX: THE ALGORITHM USED TO DETERMINE THE PAYOFFS IN STUDY 3

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 Familiar (the constant payoff obtained from selection 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 many of the Familiar and Low parameters (Familiar = 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 .05 and .95. Then the High parameter was determined such that the expected value from exploration would be equal to the Familiar parameter ($H = round\ [(Familiar - Low * Plow)/(1 - Plow)]$). Ten games in which the aforementioned constraints were met were randomly chosen. Then, to ensure an optimal strategy, we added .5 to the Familiar value in the first five games and subtracted .5 from the Familiar value in the other five games. This way, for each Plow value (.05, .25, .5, .75, and .95), there was one game in which the optimal strategy was to select familiar keys and one game in which the optimal strategy was to explore new keys.

Accordingly, game no. 1, in which *Plow* = .05 and the optimal strategy is to select familiar keys, was an extreme version of the Rare Disasters condition in Study 1, and game no. 10, in which *Plow* = .95 and the optimal strategy is to explore new keys, was an extreme version of the Rare Treasures condition in Study 1.

Game number	1	2	3	4	5	6	7	8	9	10
Condition's name	C95	C75	C50	C25	C5	B95	B75	B50	B25	B5
Plow	.05	.25	.5	.75	.95	.05	.25	.5	.75	.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 (=Familiar)	9	11	4	8	5	12	1	3	6	2
Noise	+.5	+.5	+.5	+.5	+.5	−.5	−.5	−.5	−.5	−.5
Familiar final (Familiar + Noise)	9.5	11.5	4.5	8.5	5.5	11.5	.5	2.5	5.5	1.5
Optimal strategy	Select familiar keys	Select familiar keys	Select familiar keys	Select familiar keys	Select familiar keys	Explore new keys	Explore new keys	Explore new keys	Explore new keys	Explore new keys

ACKNOWLEDGEMENTS

This paper was supported by grants from the Israel Science Foundation and Technion–Microsoft Electronic-Commerce Research Center.

[5]For example, it is estimated that Australians alone spend on average ~$US9.99 billion every year on goods they do not use (i.e., that never even make it out of the box). That is an average of $US1156 for each household (Botsman & Rogers, 2010).

REFERENCES

Barron, G., & Erev, I. (2003). Small feedback based decisions and their limited correspondence to description based decisions. *Journal of Behavioral Decision Making*, *16*, 215–233.

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.

Busemeyer, J. R., & Myung, I. J. (1987). Resource allocation decision making in an uncertain environment. *Acta Psychologica*, *66*, 1–19.

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

Erev, I., & Grainer, B. (in press). The 1-800 critique, counter-examples, and the future of behavioral economics. In: G. Frechette, & A. Schotter (Eds.), *The Methods of Modern Experimental Economics*. Oxford University Press, forthcoming.

Erev, I., Ert, E., & Roth, A. E. (2010a). 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., et al. (2010b). 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.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*, 493–518.

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., et al. (1996). A component analysis of cognitive-behavioral treatment for depression. *Consulting and Clinical Psychology*, *62*, 295–304.

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.

Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition*, *38*, 667–676.

Landmine Monitor reports (2010). International Campaign to Ban Landmines. http://www.the-monitor.org/index.php/publications/display?url=lm/2010/es/Casualties_and_Victim_Assistance.html

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.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, *23*, 1–14.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology. General*, *135*, 207–236.

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.

DOI: 10.1002/bdm

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.

Authors' biographies:

Kinneret Teodorescu received her BA in Psychology and Economics from Tel-Aviv University in 2008 and her MSc in Cognitive Psychology from the Technion, in 2010. Mrs. Teodorescu is currently studying toward her PhD in Behavioral Economics at the Technion. Her research interests include learning, decision making, and quantitative models of human behavior.

Ido Erev is the "ATS' Women's Division Professor" of Industrial Engineering and Management at the Technion and a Visiting Professor of Behavioral Economics at Erasmus University. His research focuses on decisions from experience and the economics of small decisions.

Authors' addresses:

Kinneret Teodorescu and Ido Erev, Max Wertheimer Minerva Center for Cognitive Studies, Technion-Israel Institute of Technology, Israel

DOI: 10.1002/bdm