The description–experience gap in risky choice

Ralph Hertwig¹ and Ido Erev²

¹University of Basel, Department of Psychology, Missionsstrasse 60/62, CH-4055 Basel, Switzerland
²Technion, Max Wertheimer Minerva Center for Cognitive Studies, Faculty of Industrial Engineering and Management, Haifa 32000, Israel

According to a common conception in behavioral decision research, two cognitive processes—overestimation and overweighting—operate to increase the impact of rare events on people’s choices. Supportive findings stem primarily from investigations in which people learn about options via descriptions thereof. Recently, a number of researchers have begun to investigate risky choice in settings in which people learn about options by experiential sampling over time. This article reviews work across three experiential paradigms. Converging findings show that when people make decisions based on experience, rare events tend to have less impact than they deserve according to their objective probabilities. Striking similarities in human and animal experience-based choices, ways of modeling these choices, and their implications for risk and precautionary behavior are discussed.

Beware of black swans
In 2009, the world found itself in the midst of the worst recession since the Great Depression. Events thought of as extremely unlikely, such as the bust of the U.S. housing boom, the meltdown of the financial system, and the bankruptcy of colossal companies, happened in breathtakingly fast succession. Why was the world so badly prepared for these unlikely events? One explanation is that the crisis of the financial industry preceding the economic recession occurred because the industry’s supposedly optimal risk management models failed to reckon with ‘black swans’—unexpected and unpredictable rare events that carry an enormous impact [1]. Of course, not only modern risk management paradigms failed to take the black-swan event into account—so did individual players, such as many homeowners who could no longer afford their mortgages. Can psychological theories and findings account for such blind spots? At first glance, the answer is no. Influential studies in behavioral decision research consistently suggest the opposite propensity: people are oversensitive to rare events. For example, they overestimate the chance of food poisoning or of lung cancer resulting from smoking [2,3]. Moreover, people are depicted as remembering past experiences by how they felt at the peak (rare moment) of their experience and how they felt when the experience came to an end [4]. Such oversensitivity is not only empirically observed but also theoretically suggested. According to the most influential descriptive theory of risky choice, people overweight low-probability events [5]. In fact, cumulative prospect theory explains the puzzling co-occurrence of two behaviors—that the same people who purchase lottery tickets that promise tiny chances of winning also take out insurance against tiny chances of damage [6]—by the assumption that small probabilities receive ‘too much’ weight.

In light of people’s ostensible oversensitivity to rare events, why did so many people, financial experts and private citizens alike, behave as if they were not cognizant of the rare events that triggered what some called a bona-fide depression [7]? Analyses have highlighted a variety of enabling factors, ranging from purportedly rational bankers who acted on strong incentives to take maximum risks in their lending [7] to humans’ ‘animal spirits’ [8]. However, there is another possibly enabling condition. The customary portrayal of people as being oversensitive to rare events neglects that when people recruit their experience sampled across time to make risky decisions, chances are that rare events (such as the burst of housing bubbles) receive less impact than they deserve according to their objective probabilities.

The description–experience gap
Just as biologists use the Drosophila (fruit fly) as one model organism, behavioral decision researchers have used choice between monetary gambles as a model for risky choice, assuming that many real-world options have the same properties as gambles, namely, n outcomes and associated probabilities [9]. Moreover, many researchers have grown accustomed to presenting their respondents with one particular genus of the fruit fly: gambles in which all outcomes and their probabilities are stated, and respondents make a single choice. Figure 1 illustrates a typical description-based decision problem.

In everyday life, however, people can rarely peruse such descriptions of probability distributions (there are a few exceptions such as media weather forecasts stating probabilities of precipitation [10]). When people decide whether to take out a mortgage loan or contemplate the success of a first date there are no tabulated risks to consult. Instead, they need to rely on their experiences—if existent—with these options, thus making decisions from experience rather than decisions from description [11]. Both kinds of decisions can be understood as opposite entries on a continuum of uncertainty about what one is choosing between. Using Knight’s terminology [12], decisions from descriptions involve a priori probabilities, whereas decisions from experience involve statistical probabilities, which must be assessed
IF researchers turned away from the transients of learning (for an exception see, for example [15]). Moreover, with the increasing importance of expected utility theory, the study of anomalies became pertinent; those required the conveying of perfect information about the probabilities of relevant events (Figure 1). Interest in issues of learning and experience-based decisions, however, remained alive in other fields such as operation research (see literature on multi-armed bandit problems [16]).

Modern decision scientists' interest in decisions from experience has been rekindled by the recent observation of systematic and robust differences between decisions based on experience and description. Research on decisions from experience has come with a simple experimental tool, a 'computerized money machine.' Respondents see two buttons on a computer screen, each one representing an initially unknown payoff distribution (Figure 1). Clicking a button results in a random draw from those distributions. Three variations of this experimental tool have been employed. In the **sampling paradigm**, people first sample as many outcomes as they wish and only then decide from which distribution to make a single draw for real [11,17]. In the **full-feedback paradigm**, each draw contributes to people's earnings and they receive draw-by-draw feedback on the obtained and the forgone payoffs (i.e. payoff received had the other option been selected) [18]. The **partial-feedback paradigm** is identical to the full-feedback paradigm, except that people only learn about the obtained payoffs [19,20]. Here, unlike in the first two paradigms, respondents face an exploitation–exploration tradeoff. Exploitation and exploration represent two goals associated with every choice, namely, to obtain a desired outcome (exploitation) or to gather new information about other, perhaps better, actions (exploration) [21].

Across all three experiential paradigms, a robust and systematic description–experience gap has emerged in numerous studies. Figure 2 illustrates this gap in six decision problems [22]. Each one offers a choice between a risky option with two outcomes and a safe option. In the risky options, either the desirable outcome or the less desirable outcome occurs with low probability (probability of 0.1 or less). In all three experiential paradigms, respondents tend to select the risky option when the desirable outcome occurs with high probability, and select the safe option when the desirable outcome occurs with low probability. This tendency is reversed in decisions from description, and the general pattern can be summarized as follows: in decisions from experience, people behave as if the rare events have less impact than they deserve according to their objective probabilities, whereas in decisions from description people behave as if the rare events have more impact than they deserve (consistent with cumulative prospect theory).

**What causes the description–experience gap?**

Several causes have been proposed as contributing to the description–experience gap. These are reviewed in the following section.

**Small samples**

Based on people's search in the sampling paradigm, reliance on small samples has been proposed as one factor...
that contributes to the attenuated impact of rare events [11]. Across numerous studies, respondents typically proved restrained in their information search, with a median number of samples per choice problem typically ranging between 11 and 19 across studies (reviewed in [13]). For such samples, chances are that a person does not even experience the rare event. More generally, one is more likely to undersample than to oversample the rare event. This is because the binomial distribution for the number of times a particular outcome will be observed in \( n \) independent trials is markedly skewed when \( p \) is small (i.e. the event is rare) and \( n \) is small (i.e. few outcomes are sampled). Reliance on small samples has also been discussed as a potential explanation for bumblebees’ underweighting of rare events (Box 1).

This explanation in terms of small samples has prompted a critical response [26] and has led to an ongoing debate. What appears to be underweighting of rare events in decisions from experience could be consistent with overweighting of low probabilities as assumed in cumulative prospect theory. When the probability experienced in a sample is smaller than the event’s objective probability, people might still overweight this sample probability. Despite this overweighting, the erroneous impression of underweighting would emerge if the overweighting did not fully compensate for the underestimation because of the skew in small samples. In this view, the description–experience gap is statistical (sampling error) rather than psychological in nature.

Several approaches have been taken to examine whether the gap observed in the sampling paradigm can indeed be reduced to sampling error (why people appear to rely on small samples has received scant attention but see Ref. [15,27]). If sampling error were the sole culprit, then reducing the error by extending the sample should attenuate and eventually eliminate the gap. Increasing sample sizes substantially (up to 50 and 100 draws per choice problem) could reduce but did not eliminate the gap [13,28]. If sampling error caused the gap, then removing the error by aligning experienced probabilities in the sample to the objective probabilities should eliminate it; it did not [29]. If sampling error were the sole culprit, then removing the error by aligning experienced probabilities in the sample to the objective probabilities should eliminate it; it did not [29]. If sampling error were the sole culprit, then removing the error by aligning experienced probabilities in the sample to the objective probabilities should eliminate it; it did not [29]. If sampling error were the sole culprit, then removing the error by aligning experienced probabilities in the sample to the objective probabilities should eliminate it; it did not [29].

Figure 2. The description–experience gap. Proportion of choices of the risky option (i.e. risky choice) as a function of the probability of the more desirable outcome in 6 of 120 problems studied in [22]. Each presents a choice between a risky option and a safe option. The decision problems and the expected values (EV) of the risky options are displayed below. Each problem was studied using the four paradigms listed in Figure 1 [22; the data from the full-feedback are reported in Nevo, I, and Erev, I, unpublished]. Participants (20 per paradigm) were paid (in Shekels) for one of their choices, randomly selected. The partial and full-feedback paradigms involved 100 choices per problem, and the reported proportions are the means over these choices and participants.
Humans and animals alike face uncertain environments. By studying decisions from experience, researchers enjoy the surplus benefit of being able to better understand human choice by discovering how it resembles or differs from choices made by other animals. There is some indication that an attenuated impact of rare events characterizes both humans’ and animals’ decisions from experience. Investigating the foraging behavior of bees across floral rewards distributions, Real concluded that ‘bumblebees under-perceive rare events and over-perceive common events’ (p. 985) [23]. Moreover, he observed that ‘bees frame their decisions on the basis of only a few visits’ (p. S133), and suggested that such reliance on small samples might be adaptive when there is a high degree of spatial autocorrelation in the distributions of floral rewards [24]. Placing people into situations in which they render decisions from experience also led to the discovery that risk preference in humans, birds and insects is better predicted in terms of the coefficient of variation (CV), a measure of risk per unit of return, than outcome variance, the measure typically used in normative models such as expected utility theory [17].

A striking difference between humans’ and other animals’ decisions from experience could be resolved by taking properties of reward into account [25]. In studies of animal choice, ‘participants’ such as rats receive rewards in the form of cups of water and thus discern differences in rewards less accurately than humans who receive the rewards in the form of digital numbers. When the perceptual noise in the presentation of rewards is increased, however, humans exhibit the same pattern of choices as animals. Similarly, when perceptual noise is decreased and discrimination becomes easier, honeybees display the same choices as humans. In general, in experience-based choices humans and honeybees alike appear to deviate from maximization in favor of the alternative perceived to lead to better outcomes most of the time (thereby underweighting rare events).

Finally, the gap persisted even when people were presented both descriptions and experience, relative to just descriptions [31]. At this point, the reality of the description–experience gap across the three experiential paradigms is unchallenged—its cause, however, is debated. Some researchers have argued that the gap in the sampling paradigm is statistical in nature [26,30,32]; others have proposed that the sampling error is not the sole cause [11,13,28,29]. Regardless of how this debate will advance, it is informative to go beyond the sampling paradigm. Reliance on small samples, for example, cannot be the reason behind the description–experience gap in the full-feedback paradigm, in which rare events’ impact is attenuated even after a hundred of trials with perfect feedback. Beyond sampling error, what psychological factors might be in play?

Recency
A psychological factor proposed to contribute to the description–experience gap is recency [11]. Ubiquitously observed in memory, belief updating and judgments [33], recency refers to the phenomenon that observations made late in a sequence receive more weight than they deserve (i.e. more than 1/n). Recency is closely related to reliance on small samples: the small sample of recent events can reintroduce the aforementioned skew into large samples of experience. Although the original finding was that people give more weight to outcomes that occurred recently in the flow of their experience relative to previous outcomes [11], no or little impact of recency was observed in later studies [13,29,30].

Estimation error
In theory, the description–experience gap could also be the consequence of a systematic estimation error [26], with people systematically underestimating the frequencies of the rare event experienced in the sample. Studies of frequency and probability assessments, however, commonly report overestimation of rare events [2,34]. Moreover, studies recording people’s estimates of rare events in the sampling paradigm found them to be well calibrated or a little too high, relative to the experienced frequency [28,29]. That is, people do not systematically estimate rare things to be even rarer.

Contingent sampling
Still another factor that could underlie the description–experience gap, especially in the feedback paradigm, is the notion that people inform their decisions by recruiting recent and past experiences garnered in similar situations ([35,36] for related notions). Such contingent sampling is likely to be ubiquitous ‘in the wild’ [37]. For example, when firefighters need to predict the behavior of a fire they appear to retrieve from memory similar instances from the past. Contingent sampling implies recency and reliance on small sampling to the extent that similarity decreases with time. Furthermore, in dynamic environments (e.g. restless bandit problem [38]), reliance on similar experiences is an efficient heuristic [39]. Admittedly, one weakness of the notion of contingent sampling is the fuzzy term ‘similarity.’ However, defining similarity within a formal model is not impossible, as the result of a recent model competition demonstrated (Box 2).

Information format and cognitive algorithms
Finally, the description–experience gap might be partly caused by different formats of statistical information triggering different cognitive algorithms [13,28]. Most investigations of decisions from description represent the likelihood of outcomes in terms of single-event probabilities (or percentages). In experiential paradigms, in contrast, people come across sequences of events, without an explicit synopsis of the events’ probabilities. This difference echoes the distinction between single-event probabilities and natural frequencies in Bayesian reasoning, with the different format giving rise to different cognitive algorithms and degrees of Bayesian reasoning [44,45]. By analogy, the sequential experience of events might trigger different cognitive algorithms than those triggered by described single-event probabilities.

In sum, modern behavioral decision research was strongly focused on people’s responses to descriptions of events. In recent years, three experiential paradigms have been used to study how experience affects risky choice. A consistent picture has emerged. Where rare events are implicated, description-based and experience-based decisions can dramatically diverge. Next to sampling error, a number of psychological factors have been proposed. One important question for the future is to find which of these
Decisions from description can differ systematically from decisions from experience. Experience-based decisions, for example, appear to reverse the fourfold pattern of risk attitudes that has been found in decisions from description [40], provoking a call for two different theories of risky choice [11,17,41]. Others, however, have responded that cumulative prospect theory, initially designed for decisions from description, could also account for decisions from experience [26]. To evaluate how well a model predicts people’s choices, its performance needs to be compared to those of other modes. Several model competitions involving decisions from experience have been conducted [28,42]; the most extensive one was the choice prediction competition held at the Technion [22].

The Technion competition consisted of three contests involving the description, the sampling and the partial-feedback paradigms (Figure 1). The objective in all three contests was to predict the proportion of people who prefer a safe option with a medium payoff to a risky option yielding two payoffs, one high, the other one low. Each contest was based on two sets of choice data, a fitting and a competition set. In order to generate both datasets, the same method and participant pool was used, and each dataset consisted of 60 randomly selected problems with the aforementioned properties. The fitting set, together with the fit of several baseline modes, was posted on a website, and enabled researchers to fit the model that they thought would perform best in predicting the competition set. Fourteen teams responded to the challenge, representing a wide range of theoretical approaches (e.g. heuristics, algebraic models, logit-regression model and a model based on the ACT-framework).

Again, the empirical data revealed a substantial description–experience gap, thus replicating previous results. Notwithstanding the limited psychological plausibility of some of the leading models, the description–experience gap is also reflected in the models that performed well in the three contests. In decisions from description, the best-performing model was a stochastic version of cumulative prospect theory with decision weight estimates reflecting over-weighting of rare events. All leading models in the two experience contests share the assumption that prior to choice the decision maker recalls a small set of about five experiences with each option, and tends to select that option resulting in better outcomes in this sample. In addition, the winning model in the partial feedback paradigm implemented the notion of contingent sampling (see text). Note that this model competition is featured in a special edition of the Journal of Behavioral Decision Making devoted to the topic of ‘decisions from experience’ [43].

A final issue concerns the potential redundancy of the three experiential paradigms (Figure 1). They are not redundant. The partial-feedback paradigm, for example yields more safe choices than the full-feedback paradigm [18,48]. This difference might be a product of the hot-stove effect. As Mark Twain observed, if a cat happens to jump on a stove that is hot, it will never jump on one again, thus never learning to distinguish between hot and cold stoves. When feedback is limited to the obtained payoff, as in the partial-feedback paradigm, people are more likely to behave risk averse, thus reducing the probability of sampling alternatives with poor past financial and social outcomes [49,50]. This brings the discussion full circle to where it began. Generations growing up in a period of low stock returns appear to take an unusually cautious approach to investing, even decades later (although recent experience matters, too) [Malmendier, U. and Nagel, S., unpublished]. In other words, today’s teenagers who have experienced this dramatic economic slump might years later enter the stock and housing market much more cautiously than their parents did. Perhaps, it is also they who will experience other global changes first-hand and behave differently than their parents did (Box 4).
Concluding Remarks

Modern behavioral decision research has commonly focused on decisions from description. The observations stemming from this research suggest that humans overestimate and overweight rare events. Recent research stemming from this research suggest that humans overfocused on decisions from description. The observations of human behavior—their contrast is enlightening. However, to better understand how people make decisions with incomplete and uncertain information ‘in the wild’, there is a need to study experiential choices that are often representative of people’s actual choices [57], and to learn from the rich repertoire of previous research into experience-based decisions and behaviors [14,58].

Acknowledgements

We are grateful to Robin Hau and Ben Newell for their helpful comments. We also thank Laura Wiles for editing the manuscript, and the Swiss National Science Foundation for a grant to the first author (100014-118283).

References

29 Ungemach, C. et al. (2009) Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychol. Sci.* 20, 473–479

Trends in Cognitive Sciences
Vol.13 No.12

523