Università
Ca' Foscari
Venezia

**Department
of Management**

Working Paper Series

Davide Marchiori, Sibilla Di Guida, and Ido Erev

**Noisy retrievers and the four-
fold reaction to rare events**

**Working Paper n. 3/2013
April 2013**

ISSN: 2239-2734

This Working Paper is published under the auspices of the Department of Management at Università Ca' Foscari Venezia. Opinions expressed herein are those of the authors and not those of the Department or the University. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional nature.

Noisy retrievers and the four-fold reaction to rare events

Davide Marchiori

Department of Management, Ca' Foscari University of Venice
Fondamenta San Giobbe 873 Cannaregio, 30121 Venezia, Italy
E-mail: davide.marchiori@unive.it; marchiorid@yahoo.it

Sibilla Di Guida (Corresponding author)

SBS-EM, ECARES, Université Libre de Bruxelles
Avenue Roosevelt 50, ccp 114, 1050 Brussels, Belgium
E-mail: sibilla.diguida@gmail.com

Ido Erev

Faculty of Industrial Engineering and Management
Technion, 32000 Haifa, Israel
E-mail: erev@tx.technion.ac.il

Acknowledgments

This research was supported by the Max Wertheimer Minerva Center, a grant from the Israel Science Foundation, a fellowship from the Lady Davis Foundation, and a grant from the "Fonds de la Recherche Fondamentale Collective" (research grant n° 2.4614.12). We thank Peter Wakker and Daniel Friedman for useful comments.

Noisy retrievers and the four-fold reaction to rare events

Abstract. Previous research documents two pairs of inconsistent reactions to rare events: 1) Studies of probability judgment reveal conservatism which implies overestimation of rare events, and overconfidence which implies underestimation of rare events. 2) Studies of choice behavior reveal overweighting of rare events in one-shot tasks, and the opposite bias in decisions from experience. The current analysis and experimental results demonstrate that the coexistence and relative importance of the four biases can be captured with simple models that share the assumption that judgments and decisions are made based on the information conveyed by small and noisy samples of past experiences.

Keywords: Black swan; prospect theory; experience-description gap; case-based decision theory; overgeneralization.

JEL Classification Numbers: C79, C91, D81

1. Introduction

Smokers tend to overestimate the probability that they will develop lung cancer (Viscusi and Hakes, 2008). That is, their behavior (smoking) reflects low sensitivity to this risk, although their stated beliefs imply the opposite. Review of previous studies of judgment and decision-making reveals that this apparent inconsistency is not unique to smokers and/or to human reactions to health risks. The results imply a four-fold response pattern to rare events. Studies of probability judgment reveal over- and underestimation, and studies of decision-making reveal over- and underweighting of rare events.

A clear indication of overestimation of rare events is provided by studies, like the “lung cancer” question, that focus on the subjective estimates of events whose objective probability is below 0.5. The typical judgment tends to fall between the objective probability and 0.5. The robustness of this phenomenon is clarified in a comparison of subjective judgments to the prescriptions of Bayes’ rule: When the objective probability (the prescription of Bayes’ rule) is low, the subjective judgments tend to be too high (see Phillips and Edwards, 1966, and Table 1a). Typically, the mean estimate of an event that occurs in 5% of the cases is around 20%. This pattern is referred to as “conservatism”.

An important indication of underestimation of rare events is provided by studies of confidence judgment. People appear to exhibit overconfidence. When the typical subject estimates the probability that a certain statement is accurate by 5%, the accuracy rate tends to be close to 20% (Fischhoff et al., 1977, and Table 1b).

The best-known indication for overweighting of rare events is provided by Kahneman and Tversky’s (1979) analysis of the Allais Paradox (Allais, 1953), and the observation that the same agents buy insurance and lotteries (Friedman and Savage, 1948). Kahneman and Tversky replicated the two phenomena in simple experiments (see Table 1c), and showed that they can be explained (according to prospect theory) by assuming that extreme low probability outcomes are over-weighted (see also Wakker, 2010 for an empirical and axiomatic extension of this analysis to decisions under uncertainty).

Table 1a-d. Summary of experimental studies that demonstrate the distinct effects of rare events.

Typical experimental task	Typical results and interpretation
a. Revision of opinion (Phillips and Edwards, 1966) Urn A includes 30 Red Balls and 70 White balls. Urn B includes 70 Red Balls and 30 White balls. One of the two urns was randomly selected (the prior probability that A will be selected is 0.5). The experimenter sampled with replacement 4 balls from that urn. All 4 balls are Red. What is the probability that the selected urn is A?	Mean estimate is 0.23 Bayes' posterior probability is 0.01 The mean reflects overestimation that can be a product of conservative revision of opinion.
b. Confidence judgment (Fischhoff et al., 1977) At each trial the participants are asked assess the accuracy of a certain assertion (e.g., "NYC is closer to Atlanta than to Chicago"), and then estimate their confidence in their guess.	The proportion of inaccurate guesses given an estimate of 5% (confidence of 95%) is about 20%. This observation is known as overconfidence.
c. Decisions under risk (Kahneman and Tversky, 1979) Choose between the following two options: S: 5 with certainty R: 5000 with $p = 0.001$; 0 otherwise	The choice rate of prospect R is 72%. This possibility effect is captured in prospect theory with the assertion that the low probability outcome is over-weighted.
d. Decisions from experience (Barron and Erev, 2003) At each trial the participants are asked to select between two unmarked keys: one always yielding 3, and the other yielding "32 with probability 0.1, and 0 otherwise". Each selection is followed by immediate feedback concerning the obtained payoff structure.	The risky key is selected only 32% of the times after 100 trials. This tendency to underweight rare event is captured with the assertion that decisions makers rely of small samples of past experiences.

The significance of the tendency to underweight rare events is best illustrated by the tragedy of Dr. Ignaz Semmelweis. In 1847, Dr. Semmelweis first demonstrated that routine hand washing can dramatically reduce the death rate of maternity patients and of their doctors. In one of his studies, the death rate fell from 15% to near 0%. Although his findings were published, there was no apparent increase in hand washing by doctors until the discoveries of Louis Pasteur, years after Dr. Semmelweis' death in a mental asylum (Nuland, 2003). These and similar observations suggest that in many situations people behave as if they believe that "it won't happen to me" (e.g., fail to wash hands). Recent studies of decisions from experience demonstrate the robustness of this phenomenon (see Barron and Erev, 2003; review in Hertwig and Erev, 2009; and Table 1d) and show that it can be a product of a tendency to rely on a small set of past experiences. Experience was found to reduce the weighting of rare events even when the decision makers can rely on a complete description of the payoff rule and the feedback does not add information

(Yechiam et al., 2005; Jessup, Bishara, and Busemeyer, 2008; Lejarraga and Gonzalez, 2011). Taleb (2007) refers to this behavioral pattern as the “Black Swan effect” and used it to predict the 2008 stock market crash.

The leading explanations for the four distinct effects of rare events that we have reviewed assume situation-specific psychological processes, and led to the development of elegant models that clarify these processes. Although these models provide convincing explanations of many natural phenomena, they shed limited light on the boundaries of these contradicting biases. Previous attempts to address this problem focused on the co-existence of two of the four phenomena: Fox and Tversky (1998) present an elegant two-stage model for the coexistence of overestimation and overweighting of rare events; Erev et al. (1994) highlight sufficient conditions for over- and underestimation; and Erev, Glozman, and Hertwig (2008) model the coexistence of over- and underweighting of rare events. The current analysis is an extension of these works. Specifically, we build on Erev et al.’s (1994) observation that the co-existence of over- and underestimation can be predicted by “truth-plus-error” models that imply noisy retrieval from memory. The main contribution of the present extension is the identification of a family of “noisy retriever” models that provide sufficient conditions for the coexistence of all four biases, i.e., conservatism (which implies overestimation of low probabilities), overconfidence (which implies underestimation of low probabilities), overweighting of rare events, and underweighting of rare events. In addition, the current analysis sheds light on the conditions that affect the relative importance of the four phenomena.

Our models are examples of case-based decision models (Gilboa and Schmeidler, 1995); they share the assumption that estimations and choices are made based on samples from past experiences in similar situations. Under this assumption, overestimation and overweighting are products of the noisy nature of the sampling process, whereas underestimation and underweighting are products of the small sample size.

The current models include one counterintuitive assertion. They state that people rely on past experiences even in “decisions under risk” i.e., when they receive a complete description of the incentive structure and these experiences do not add any relevant information (see also Cooper and Kagel, 2003). The main goal of the

experimental part of our investigation is to try to understand when and why this assumption is likely to hold. Experiment 1 highlights one condition that gives rise to the assumed behavior. It examines the effect of experience with “decisions with estimated risk” (i.e., decisions made based on subjective estimates of the risk) on subsequent “decisions under risk”. The results reveal that subjective estimates tend to reflect overconfidence, and people learn to respond to this common bias in the “decisions with estimated risk” stage. For example, they learn that the estimate 5% implies an event that occurs in about 20% of the trials. Then, they tend to overgeneralize this learning to subsequent decisions under risk and behave “as if” they doubt the accurate description of payoff distributions.

Experiments 2 and 3 suggest that a similar doubt affects behavior even before subjects face the “decisions with estimated risk” in the laboratory. The results support the hypothesis that subjects learn to doubt descriptions, and the implied tendency to rely on past experiences in natural settings (before they arrive to the laboratory).

A second goal of the experimental part is that of comparing different members of the family of models we consider. The results highlight the value of models that capture individual differences, inertia, and overgeneralization from different decision tasks.

2. A family of Noisy Retriever models

The family of models considered here addresses the two classes of tasks that have been shown to produce the two pairs of biases. The first class involves judgment of the probability of different events, whereas the second involves choices between a safe prospect and a two-outcome gamble with the same expected value. Gambles are of the form “win H if event E occurs, and L otherwise”, where $H > L$. Thus, agents’ responses in both classes reflect their sensitivity to the occurrence rate of events.

The Noisy Retriever models are identified by the following two assumptions, which provide together sufficient conditions for the coexistence of the four biases.

*Assumption 1: Noisy retrieval*¹. When facing the first instance of Task j (j = judgment, decision) associated to an event that occurs with probability p , agents retrieve from their memory m_j previous occurrences (or, simply, *experiences*) that are associated to independent realizations of similar events. We indicate this set of past “similar” experiences with S . More specifically, set S includes “truly similar” experiences, which are associated to realizations of events occurring with the same probability p , and “just look similar” experiences, which are associated to realizations of events occurring with probability different than p . The occurrence rate of the events in the set of “just look similar” experiences is assumed to be identical to the occurrence rate in the whole population of estimated events (also outside set S), and we start with the simplifying assumption that this rate is 0.5². The probability that an element in S is a noisy retrieval (i.e., a “just look similar” experience) is denoted with ϵ_j .

When agents face the same task repeatedly in the same setting, and receive feedback after each trial (as in typical decision from experience tasks), experience increases the number of truly similar experiences in S , thus reducing the level of noise. Specifically, each repeated experience with a choice task with immediate feedback increases the set of truly similar experiences by one. Therefore, after t repetitions of Task j , the expected size of the set of truly similar experiences is $m_j (1 - \epsilon_j) + t$, whereas the expected number of “just look similar” experiences remains constantly equal to $m_j \epsilon_j$.

Assumption 2: Reliance on small samples. Agents’ responses in Task j are based on the information conveyed by a random sample of k_j observations from the set S of past “similar” experiences. Specifically, in judgment tasks, the estimated probability of occurrence corresponds to the proportion of occurrences in the sample

¹ Noisy retriever models can be alternatively formulated as exemplar models that perform approximated Bayesian inference (along the lines suggested by Shi et al., 2010), by assuming uniformly distributed stimuli (event probabilities) and likelihood terms proportional to the proportions of “truly similar” and “just look similar” past experiences.

² This is equivalent to assume that the distribution of event probabilities that agents have encountered in their past history is uniform.

(or the closest feasible response); in choice tasks, the risky prospect is selected if and only if the proportion of H occurrences in the sample is bigger than p .

Notice that the second assumption is also implied by the payoff sampling equilibrium concept proposed by Osborne and Rubinstein (1998). The descriptive potential of this concept in simple games is demonstrated in Selten and Chmura (2008) and in two choice prediction competitions (Erev et al., 2010a; Erev et al., 2010b). We believe that there are at least three different reasons for the tendency to rely on small samples. The first is cognitive limitation. For example, when the relevant information is presented for a short period, the judge will not be able to store, and then recall, more than a small sample of experiences. A second contributor is the cost of sampling. When sampling is costly, reliance on small samples can be optimal. A third, and more interesting, contributor is the assumption that the world can be in one of many states, and this state determines the likelihood of the different events. For example, consider the estimation of the probability that a particular paper will be accepted for publication in a certain target journal. A reasonable judge will not rely on all her past experiences with paper submissions. Rather, she is likely to rely on a small set of past experiences with papers of similar quality in the target of similar journals.

2.1. The basic Noisy Retriever model, its parameters, and predictions

The current analysis starts with a simple member of the family of models described above, which we call “basic noisy retriever” model. We show that this simple model is able to replicate the two pairs of biases observed in judgment and decision tasks under the same parameter set. Specifically, we consider a model with the constraints $m_j = m$ (size of set S of past similar experiences), $\varepsilon_j = \varepsilon$ (probability of a noisy retrieval), and $k_j = k$ (size of the random samples drawn from S), for $j = \text{judgment, decision}$. Numerical examples that clarify the intuition behind the current approach, and simulations that summarize the main predictions are presented below.

Overestimation of rare events (conservatism). According the basic noisy retriever model, the mean initial estimate at $t = 1$ of an event whose objective probability is p is $p_{\text{est}} = p (1 - \varepsilon) + 0.5 \varepsilon$. Thus, probabilities below 0.5 are

overestimated. Figure 1a presents a numerical example of this prediction under several levels of ε .

For example, consider an agent that tries to estimate the probability of an event whose true probability is $p = 0.05$. The agent does not have direct experience with this event ($t = 1$), but she is familiar with a set S of m events that appear to be similar. Assume further that 60% of the events considered by the agent are truly similar to the target event (i.e., their occurrence rate is 0.05), and 40% just look similar (i.e., their occurrence rate is 0.5). Therefore, according to the model, the probability of a noisy retrieval is $\varepsilon = 0.4$. The model implies that the agent draws a random sample of size k from set S , computes the number of relative occurrences in the sample, and uses this rate to formulate her response. In the described settings, the expected response is: $0.05 (1 - 0.4) + 0.50 (0.4) = 0.23$. Notice that this value does not depend on the parameters m and k .

Fig. 1a (left) Predicted average judgment as a function of the objective probabilities under several levels of ε . The results reveal overestimation (conservatism). For example, the average estimate of the 0.05 probability event is bigger than 0.15.

Fig. 1b (right) Predicted occurrence rate (objective probability) as a function of the judged probabilities under several levels of ε and k . The results reveal underestimation (overconfidence). For example, when the estimated occurrence rate is 0.05, the actual occurrence rate is bigger than 0.2.

Underestimation of rare events (overconfidence). The current models (following Erev et al., 1994) assume that the difference between over- and underestimation of rare events is a result of distinct ways to analyze the data. That is, this difference does not reflect different behaviors by the agents, but it rather reflects different approaches adopted by researchers to analyze empirical data.

To clarify this point, consider the agent discussed above, with the simplifying (although unrealistic) assumption that she relies on a sample of size one (i.e., $k = 1$). This additional assumption does not change the predicted average estimate. As explained above, this prediction is not affected by k . Yet, the sample size k does affect the accuracy rate given the best estimate of the true probability. With $k = 1$, the estimate of the occurrence rate of an event can be either “1”, if in the sampled experience the event had occurred, or “0”, if in the sampled experience the event had not. Thus, when the true probability is between 0 and 1, the estimates will reflect overconfidence.

Notice that the current model predicts overconfidence even when the retrieval process is not noisy ($\varepsilon = 0$). An increase in k reduces, but does not eliminate, overconfidence.

Another manipulation that reduces, but that does not eliminate, overconfidence is the addition of the constraint of a finite set of possible responses strictly larger than zero, and strictly smaller than one. Figure 1b summarizes the results of a virtual experiment in which agents, behaving in accordance with the current model with a finite set of possible responses, estimate the probability of multiple events. Our simulations consider the case in which the objective probabilities are uniformly distributed in the five categories 0.05, 0.25, 0.50, 0.75, and 0.95, and agents are forced to indicate one of these values as their estimate. The figure presents the average objective occurrence rate as a function of the estimate, for distinct combinations of ε and k . Results reveal that the occurrence rates of events that were estimated with low probabilities are higher than the estimates. This bias increases when k decreases. For example, when $\varepsilon = 0.33$ and the estimated probability is 0.05, the actual average occurrence rate is 0.10 when $k = 10$, and 0.13 when $k = 5$. Notice that the predicted overconfidence effect does not depend on the parameter m (the size of the set S of past similar experiences).

Over- and underweighting of rare events. Consider a virtual agent that has to choose between “sure gain of 5” and a risky prospect that pays “100 with probability 0.05; 0 otherwise”. The current model states that she will draw a sample of size k from set S (including experiences that are either “truly similar” or “just look similar”

to the “0.05 event”), and that she will prefer the gamble if the occurrence rate in the sample is larger than 0.05—so that the implied estimated expected value is larger than 5. With $\varepsilon = 0.4$, the mean occurrence rate in set S is $0.05 (1 - 0.4) + 0.5 (0.4) = 0.23$. When $k > 2$ most of the samples will favor the gamble, thus reflecting overweighting of rare events.

Assume now that the agent faces repeatedly the above described problem, and that she receives feedback after each trial. The current model implies that feedback will reduce this bias by incrementally changing the proportion of “truly similar” experiences relative to the “just look similar” ones in set S. The occurrence rate will move toward the true rate of 0.05. The tendency to rely on small samples implies that this decrease will reverse the typical choice; that is, it will lead to underweighting of rare events. For example, when the occurrence rate is 0.10, and $k = 4$, 65.61% of the samples will favor the safe prospect (as they do not include the 0.1 event).

Figure 2a and 2b present computer simulations that clarify the robustness of these predictions. Fig 2a presents an experiment in which the probability of the high payoff H is 0.05, whereas Fig. 2b presents an experiment in which the occurrence rate of H is 0.95. Each curve presents the proportion of risky choices as a function of the trial number under distinct parameter sets. When k is sufficiently large, the results reveal overweighting of rare events at the very first trial $t = 1$. The typical agent behaves in accordance with the prediction of prospect theory: They prefer the risky prospect when $P(H) = 0.05$, and the safe one when $P(H) = 0.95$. Experience reverses this pattern, as it increases the proportion of “truly similar” past experiences relative to the proportion of “just look similar” ones: After t repetitions with feedback of the choice task, the expected number of “truly similar” experiences in S is $m (1 - \varepsilon) + t$, whereas the expected number of “just look similar” experiences is constantly equal to $m \varepsilon_j$. After a sufficient number of trials and for sufficiently small values of k , agents exhibit underweighting of rare events: They learn to prefer the safe prospect when $P(H) = 0.05$, and the risky one when $P(H) = 0.95$. In addition, a decrease in the size of set S has the effect of speeding up the process of learning, thus anticipating the switch from over- to underweighting.

Fig. 2a (left) and 2b (right) Predicted proportion of risky choices under different levels of ϵ and k , when $m = 12$ (results averaged over ten thousand simulations). The value of $P(H)$ is 0.05 (left graph), and 0.95 (right graph). Results show initial overweighting of rare events, and a reversal of this bias with experience for opportune combinations of ϵ and k . An increase in ϵ increases overweighting, whereas a decrease in k increases the tendency to underweight rare events.

2.2. Overconfidence Triggers Overweighting

Table 2 compares the predictions of our basic model, presented above (and in first row of Table 2), with the predictions of several other interesting members of the family of noisy retriever models. The second row in Table 2 presents our favorite variant of the basic model. This model replaces the assumed symmetry between estimation and choice, with the assumption that the magnitude of the initial bias toward overweighting of rare events in decisions under risk (ϵ_{dec}) is a product of prior experience with overconfident estimates. The motivation behind this “Overconfidence Triggers Overweighting” (OTO) model is the assertion that many decisions that look like decisions under risk involve subjective estimates of the relevant probabilities. Therefore, it is natural to assume that past experiences with decisions with estimated risk affect subsequent decisions under risk. The model assumes that the probability of overgeneralization from past experiences with estimated risk (in the first trial of a decision under risk task) is given by the parameter GE_{dec} . Denoting the occurrence rate of events estimated to occur with probability p ($0 < p < 0.5$) as p' , the current model can be described as a noisy retriever model with $\epsilon_{\text{dec}} = GE_{\text{dec}} (p' - p) / (0.5 - p)$.

Table 2. Predictions of different Noisy Retriever models. Average results over ten thousand simulations of the models. The results reveal robustness of the main predictions (the fourfold effect of rare events) to distinct assumptions. The analysis of the different variants of the OTO model shows that a decrease in k_{jud} (that implies underestimation) can trigger overweighting of rare events in subsequent decisions under risk.

Model	Judged probability		R-rate in under risk decisions when $P(H) = 0.05$	
	Mean judgment when $p = 0.05$	Avg. objective probability given the 5% judgment (p')	First 5 trials	Last 5 trials
Basic $\epsilon_{dec} = \epsilon_{jud} = .5$, $k_{dec} = k_{jud} = 5$, $m_{dec} = m_{jud} = 10$	0.27	0.16	0.51	0.34
Overconfidence Triggers Overweighting (OTO) $\epsilon_{dec} = GE_{dec}(p' - p) / (0.5 - p)$, $\epsilon_{jud} = 0.5$, $m_{jud} = m_{dec} = 10$, $k_{dec} = 5$, and:				
• $GE_{dec} = 0.5$, $k_{jud} = 5$	0.27	0.17	0.38	0.26
• $GE_{dec} = 0.5$, $k_{jud} = 2$	0.27	0.31	0.54	0.30
• $GE_{dec} = 1$, $k_{jud} = 5$	0.27	0.17	0.53	0.29
I-OTO (OTO with Individual heterogeneous parameters) k_{jud}^i and $k_{dec}^i \sim U(1, \dots, 10)$ m_{jud}^i and $m_{dec}^i \sim U(1, \dots, 10)$ $\epsilon_{jud}^i \sim U(0, 1)$, $GE_{dec} = 1$	0.27	0.32	0.55	0.28
I2-OTO (I-OTO with inertia) $P_i(\text{Inertia}) \sim U(0, 1)$	0.27	0.32	0.57	0.31
I2-OTO-FP (I2-OTO with sensitivity to the Grand Mean) $w \sim U(0, 0.5)$	0.27	0.32	0.55	0.27

To derive the predictions of the OTO model, we ran simulations that included three stages. At the first stage, the virtual agent estimates probabilities; the second stage involves decisions with estimated risk (using the estimates from the first stage); finally, the third stage involves decisions under risk (with some reliance on the experiences from the second stage)³. The results show that the changes do not impair the fit of the four basic phenomena. In addition, the OTO model implies that the overweighting of rare events increases with a decrease in k_{jud} (the sample size in the estimation stage), and with an increase in GE_{dec} . Moreover, it implies that initial

³ This analysis clarifies the difference between the current model and Fox and Tversky's (1998) two-stage model. In that article, the authors focused on decisions under uncertainty, and assumed a two-stage cognitive process (estimation and choice). Differently, we focus on decisions under risk (or with estimated risk), and use different stages to capture the environment (and experimental manipulations).

decisions “under risk” are better correlated with previous decisions “with estimated risk” than with subsequent decisions “under risk”.

2.3. Other members of the family

The lower rows of in Table 2 presents variants of the OTO model that relax the different assumptions. The model in the third row relaxes the assumption that all the individuals behave in accordance to the same parameters. Specifically, this model assumes that the values of the different parameters (traits) of agent i (the values m_{jud}^i , ϵ_{jud}^i , k_{jud}^i , m_{dec}^i , and k_{dec}^i) are drawn from uniform distributions between zero and a parameter that captures the highest bound of the relevant trait (see Erev et al., 2010a). The right hand column in Table 2 shows that this relaxation of the OTO model, referred to as I-OTO, has very little effect on the predictions of the aggregate pattern.

The fourth row of Table 2 presents a variant of I-OTO, referred to as I2-OTO, which allows for the possibility of inertia (see Erev et al., 2010a). Under inertia, agents simply repeat their last choice. The results reveal that the addition of inertia does not affect substantially the main predictions.

The fifth row of Table 2 presents a joint generalization of I2-OTO and the fictitious play model (see Fudenberg and Levine, 1998; and a related idea in Camerer and Ho, 1999; Erev and Roth, 1998). This I2-OTO-FP model assumes that the Estimated Subjective Value (ESV) of the risky option is computed as the linear combination of its average outcome in a small sample of past experiences and its average outcome over all past experiences (Grand Mean). The weight w_i given to the Grand Mean is a trait of agent i . Agents select the risky option if its ESV is bigger than the average return from the safe alternative. Simulation results show that introducing choice dependence upon the Grand Mean has little effect on the predictions of the model.

2.4 Summary of the assumed relationship between judgment and decision making

The analysis presented above shows that the two pairs of contradicting reactions to rare events can be the product of a single process that underlies both judgment and decision making. Yet, the analysis also suggests two differences between the

judgment and decision making phenomena. One difference pertains to the nature of the two pairs of apparent inconsistent reactions to rare events. The current family of models assumes that the coexistence of over- and under-estimation of rare events is a result of distinct analyses, and that it does not reflect differences in decision makers' behavior. The same estimates can reflect overestimation of low probability events, when the mean estimation is compared to the objective probability, and underestimation of low probability events (overconfidence), when a given estimate is compared to the best estimate of the occurrence rate. In contrast, the coexistence of over- and under-weighting of rare events reflects the effect of experience on human decisions. The current models imply that people behave as if they overweight rare events when they first see a choice problem, but that this bias is reversed by experience.

A second difference between judgment and decision making regards the reasons for the assumed process of noisy retrieval. The current models imply that the reliance on “just look similar” events during judgment is a result of a general discrimination error, and allow for the possibility of a more interesting motivation for the biases affecting decisions under risk. Specifically, the “overconfidence-triggers-overweighting” assumption implies that overweighting of rare events in initial decisions under risk stems from the tendency to overgeneralize prior experience with overconfident probability estimates.

3. Experiments

The main goal of the experiments described below is to improve our understanding of the descriptive value of the assumption that people rely on past experiences even in decisions under risk, when these experiences cannot provide useful information. Experiment 1 examines the role of one possible contributor to this hypothetical tendency: prior experience with “decisions with estimated risk”. Experiments 2 and 3 examine the role of prior experience in decisions under risk before presenting subjects with decisions with estimated risk.

3.1. Experiment 1

In this experiment, subjects faced, in sequence, the following three distinct stages.

Stage 1: Probability judgment. This stage included a sequence of 40 judgment trials, and each trial included three phases. In the first phase, subjects were given the description of their goal as follows: “In the next screenshot you will see a matrix of 100 numbers. You are asked to estimate the proportion of numbers between *min* and *max*, extremes included”. In each trial, the integer *min* was picked up randomly from $[10, 90]$, and *max* set equal to $min + 9$. In the second phase, subjects were presented for 0.75 seconds with a 10x10 matrix of integers. Of these 100 integers, n were randomly sampled (with replacement) from interval $I = [min, max]$, and the remainders sampled from those in $[10, 99]$ and outside I , with n belonging to the set $\{5, 25, 50, 75, 95\}$. The 100 numbers so obtained were shuffled and used to fill the matrix. Finally, in the third phase, subjects were asked to indicate their judgment (estimated proportion) by selecting one of the five possible answers: 0.05, 0.25, 0.5, 0.75, and 0.95.

In this stage, subjects saw in random order eight matrices for each of the five proportions 0.05, 0.25, 0.5, 0.75, and 0.95, for a total of 40 judgment problems. Subjects were informed that they would have used their judgments in the following stage of the experiment.

Stage 2: Decision with estimated risk. At this stage, subjects had to perform, in random sequence, 60 choice tasks. Specifically, at each trial, subjects had to choose between two prospects of the form:

Risky option: You win H if an event E, whose probability was estimated to be equal to 0.05, occurs; you win L otherwise.

Safe option: You win M for sure.

When presented to the subjects, the two options in each problem were labeled as “Option A” and “Option B”. The three parameters of the choice problems (H, L, and M) were modified from trial to trial to insure that the subjects would not

encounter the same problem twice. We used four levels of L (0, 1, 2, and 3), set $M = L + x$ with five levels of x (1, 2, 3, 4, and 5), and then set $H = L + 20x + u$ with three levels of u (-1, 0, and 1). Supplementary Materials 1 presents the 60 problems.

Subjects received feedback about the outcome of their choices and foregone payoffs starting from the fifth trial (thus, the first five decisions were made without feedback).

The probability $P(H)$ of winning H depended upon subjects' own accuracy in the previous judgment tasks. More specifically, $P(H)$ was set equal to the average actual proportion of numbers in an interval given the judgment (estimated proportion) of 5%. For example, if a subject used the estimate 5% only twice, and the true proportions in these cases were 0.05 and 0.25, the probability of H would have been set equal to 0.15⁴.

Notice that our experimental design simulates a common relationship between choice and judgment. Indeed, there are many situations in which judgment is made based on a perceptual task, and the feedback is about success rate. One set of examples comes from medicine: Radiologists' judgment is based on perception, and the feedback is the success of the treatment.

Stage 3: Decision under risk. Subjects faced, in a different random order, 60 pairs of gambles similar, but not identical, to the gambles they faced in Stage 2 (same values of H , M , and L). However, in the current stage, the description of the probability of winning H (0.05) was *objective* (i.e., H did occur in 5% of the cases). At each trial, subjects were presented with two prospects of the form:

Risky option: You win H if an event E , whose probability is 0.05, occurs; you win L otherwise.

Safe option: You win M for sure.

⁴ Since 5% is the lowest possible response during the estimation stage, noise during the estimation implies overconfidence. That is, the events estimated with 5% must occur with equal or higher probability. Our model, and the previous results that it summarizes, imply overconfidence even when the subjects can use lower estimates. Yet, the current experiment was not designed to test this well-known pattern; rather, it focuses on the effect of overconfident estimates on subsequent decisions under risk.

As in Stage 2, subjects were informed about the outcome of their choices and foregone payoffs starting from the fifth trial.

Participants. Twenty Technion students participated in this experiment. They were paid based on the outcome of their decision in one randomly selected trial (payoffs in Shekels) plus a show-up fee of 25 Shekels (about \$7).

Results. Figures 3a and 3b summarize the results of the judgment phase. The results reveal the over- and underestimation pattern discussed above. Figure 3a plots the average subjective probability as a function of objective probability (i.e., the true proportion of numbers in a given interval), showing conservatism (in particular, overestimation of rare events). For example, when the objective probability was 0.05, the average estimate was 0.23. Figure 3b reports the average occurrence rate as a function of subjective probability, this time showing overconfidence (in particular, underestimation of rare events). For example, the mean occurrence rate given the estimate 0.05 was 0.26.

Fig. 3a (left) Experiment 1: Average probability judgment as a function of the objective probability.
Fig. 3b (right) Experiment 1: Occurrence rate (objective probability) as a function of the probability judgment.

Figure 4 reports the proportion of risky choices (R-rates) averaged over blocks of five trials (and Supplementary Materials 1 shows the main results by problem). In the “under risk” stage, we observe initial overweighting of rare events and a subsequent reversal of this pattern. For example, in the first block of decisions under

risk, when subjects are not given any feedback, the proportion of risky choices is 63%. However, as subjects start receiving feedback, they learn to avoid the risky prospect. Experience reduced the rate of the risky choice to 32% in the last five trials ($t_{19} = 3.04$, $p = 0.01$).

In the “with estimated risk” stage, average behavior shows the opposite trend of that in the “under risk” one. For example, subjects select the risky option 44% of the times in the first five trials, and 69% of the times in the last five trials. This increase in the R-rate is significant ($t_{19} = 2.80$, $p = 0.01$).

Fig. 4 Experiment 1: rate of risky choices averaged over blocks of five trials. The figure reports the correlations between risky rate in the first block of decisions under risk with that in the previous and subsequent 55 trials.

Correlation analysis reveals that the R-rate in the first five trials “under risk” is better correlated (Spearman’s $\rho^5 = 0.92$, $p < 0.01$) with the R-rate in the 55 previous decisions (made with estimated risk), than with the subsequent 55 decisions (Spearman’s $\rho = 0.48$, $p = 0.03$) made under risk. The difference between the two correlations is significant (Fisher t transform for paired correlations, $t_{18} = 3.52$, $p < 0.01$; see also Supplementary Materials 3 and 4). This pattern is consistent with the hypothesis that initial decisions under risk reflect overgeneralization from prior experience with estimated risk. The relative low correlation with subsequent

⁵ Spearman’s rho statistic is used to estimate a rank-based measure of association, and this test may be used if the data do not necessarily come from a bivariate normal distribution. However, we also tested for association between paired samples using the Pearson’s product moment correlation coefficient, obtaining substantially the same results.

decisions under risk challenges the common assumption that decisions under risk are reflections of a static weighting probability function and/or reflections of a stable risk attitude⁶.

3.2. Experiment 2

The design of Experiment 2 is identical to that of Experiment 1, except for the order of the three stages. Subjects (20 Technion students that did not participate in Experiment 1) first faced the sequence of decisions under risk, then the judgment stage, and, finally, the sequence of decisions with estimated risk.

Results. Figure 5 and Supplementary Materials 1 present the observed R-rate in Experiment 2⁷. The first five trials of the “under risk” stage (the decisions made without feedback) reveal the overweighting pattern predicted by prospect theory. The R-rate is larger than 50%. In addition, the results reveal that experience reverses this pattern. The R-rate in the 55 decisions under risk with feedback is 30%. This value is significantly lower than the R-rate in the first five trials ($t_{19} = 3.52$, $p < 0.01$), and significantly lower than 0.5 ($t_{19} = 4.02$, $p < 0.01$). The results in the “with estimated risk” phase reveal a gradual increase in the R-rate.

⁶ Under one alternative explanation, the relatively low correlation between the initial (trials 1-5) and all subsequent decisions under risk might reflect a floor effect. In theory, it is possible that the strong tendency to reject the gamble following experience masks the indications for individual differences in risk attitudes. In order to test this hypothesis we also computed the correlation of the R-rate under risk in trials 1-5 with the last 30 trials with estimated risk (R-rate = 65%), and with the R-rate in the first 30 decision under risk with feedback (trials 6-36, R-rate= 56%). The results reveal that the initial R-rate under risk is still better correlated with the previous decisions with estimated risk (Spearman's $\rho = 0.94$, $p < 0.01$), than with the subsequent decisions under risk (Spearman's $\rho = 0.69$, $p < 0.01$). This pattern cannot be explained as a reflection of a floor effect.

⁷ The results of the judgment phase are very similar to the results of the estimation phase in Experiment 1, and for that reason we do not present them.

Fig. 5 Experiment 2: rate of risky choices averaged over blocks of five trials. The figure reports the correlations between the risky rate in the first block of decisions under risk with that in the subsequent 55 trials and that in the last 55 trials made with estimated risk.

Correlation analysis reveals a surprising pattern. The R-rate in the first five trials “under risk” is better correlated (Spearman’s $\rho = 0.52$, $p = 0.02$) with the R-rate in the final 55 trials with estimated risk, than with the final 55 trials (Spearman’s $\rho = 0.24$, $p = 0.32$) made under risk. Like the correlations in Experiment 1 this, this pattern suggests a large difference between decisions under risk that are made with and without feedback. Under the current explanation of this pattern, decisions under risk without feedback reflect overgeneralization from past experiences with estimated risk. The current design (the fact that the “under risk” stage was played first) suggests that here the overgeneralization is from experiences that occurred before the beginning of the experiment.

3.3. Experiment 3

According to the Noisy Retriever models, the decrease in the observed R-rates in the “under risk” conditions of Experiment 1 and 2 is a reflection of two factors: reliance on small samples, and a decrease in the overgeneralization from decisions with estimated risk. Under an alternative explanation, the results reflect learning to become risk averse (see March, 1996). To compare the two explanations, Experiment 3 focuses on a situation in which the current model predicts an increase in risk seeking with experience in decisions under risk.

The experimental design in Experiment 3 is identical to the design in Experiment 2 with three exceptions. First, the probability $P(H)$ to win the high payoff

H was set equal to 0.95. Second, the values of the three payoff parameters were modified to insure that (under risk) the gamble and the safe prospect had the same expected values. We used four values of H (104, 103, 102, and 101), set $M = H - x$ with five values of x (1, 2, 3, 4, and 5), and then set $L = H - 20x + u$ with three levels of u (-1, 0, and 1). Supplementary Materials 2 presents the implied problems. A final difference was a lower conversion rate (to insure the same expected earnings as in Experiments 1 and 2). The conversion rate here was 0.1 Shekels for each unit.

Results. Figure 6 and Supplementary Materials 2 present the observed R-rate in Experiment 3⁸. The first five trials “under risk” (the decisions made without feedback) reveal the overweighting pattern (of the low 0.05 payoff) predicted by prospect theory. The observed R-rate is below 50%. Experience reverses this pattern. The R-rate in the 55 decisions under risk with feedback is 59%. This value is significantly higher than the R-rate in the first five trials ($t_{19} = 2.20$, $p = 0.04$), although not significantly higher than 50% ($t_{19} = 1.13$, $p = 0.27$). Thus, the results reject the hypothesis that experience increases risk aversion. The observed learning curve is consistent with the basic assumption of the current model. The availability of feedback in decisions under risk leads to underweighting of rare events, even when this tendency implies risk seeking. The pattern observed in the “with estimated risk” phase reveals a gradual decrease in the R-rate.

Correlation analysis reveals the robustness of the pattern observed in Experiments 1 and 2: The R-rate in the first five trials under risk is better correlated (Spearman’s $\rho = 0.38$, $p = 0.10$) with the R-rate in the final 55 trials with estimated risk, than with the final 55 trials (Spearman’s $\rho = 0.31$, $p = 0.18$) made under risk. As noted above, this pattern is consistent with the assertion that decisions under risk without feedback reflect overgeneralization from past experiences with estimated risk.

⁸ The results of the estimation phase are very similar to the results of the estimation phase in Experiment 1, and for that reason we do not present them.

Fig. 6 Experiment 3: rate of risky choices averaged over blocks of five trials. The figure reports the correlations between the average risky rate in the first block of decisions under risk, with that in the subsequent 55 trials and that in the last 55 trials made with estimated risk.

4. Individual differences and the value of different quantifications

4.1. Analysis of the judgment task

Figure 7 presents the distribution of errors in the estimation of the eight 0.5 probability events (the eight trials in which the objective probability was 0.5), by all the 60 participants in the three experiments. The horizontal axis presents the proportion of smaller errors (the responses “0.25” or “0.75” that imply an absolute error of 0.25), whereas the vertical axis presents the proportion of large errors (the responses “0.05” or “0.95” that imply an absolute error of 0.45). The family of noisy retriever models implies that these error rates are solely determined by the value of k_{jud} . Thus, Figure 7’s results allow a direct estimation of the sample size (i.e., k_{jud}^i) that best describes the behavior of each subject. The estimated k_{jud}^i value, for each participant, is the number at the right of its dot.

The results reveal that the vast majority of the estimated k_{jud}^i values (52 out of 60) are between 3 and 8, the mode is 3, and the median is 4. In addition, the results show large individual differences. Allowing for agent heterogeneity in the parameter k_{jud}^i significantly improves the fit of the data. A likelihood ratio test rejects the hypothesis of fixed k_{jud} across agents (likelihood ratio test, $\chi^2_{59} = -2 * \log L = 149.52$, $p < 0.01$).

Figure 8 presents the distribution (over the 60 subjects) of the estimated bias toward 0.5 (noise) based on the assessment of the 0.25 and 0.75 events. The estimated bias was computed for each subject as $0.5 + M25 - M75$ where M25 and

M75 are, respectively, the mean assessments of the 0.25 and 0.75 events. Under the current models, the estimated bias for subject i reflects the value of ε_{jud}^i . Thus, Figure 8 implies that the mode and the median of the estimated ε_{jud}^i are 0.25. As in the case of k_{jud}^i , the results reveal large individual differences.

Fig. 7 Proportion of small errors against proportion of large errors in the estimates of the 50% probability events. The size of each dot is proportional to the number of subjects with the same values. At the right of each dot is the corresponding estimated value of k_{jud} .

Fig. 8 Estimated bias toward 0.5 (ε_{jud}^i), computed for each of the 60 subjects as $0.5 + M25 - M75$; M25 and M75 are, respectively, the mean assessments of the 0.25 and 0.75 events.

To evaluate the predictive value of the model, we used the individual parameters estimated above to predict the proportion of accurate estimates of the 0.05 and 0.95 probability events. Each point in Figure 9 represents one of the participants. The correlation (Spearman's rho) between the predicted and the observed rates is 0.31

($p = 0.01$). One analysis that clarifies the “usefulness” of these predictions focuses on the equivalent number of observations (ENO, see Erev et al., 2007) of the model. The ENO of the current model is 6.15⁹.

Fig. 9 Predicted against observed proportions of accurate estimates in 0.05 probability events (60 subjects). The correlation (Spearman’s rho) between the two variables is 0.31 ($p = 0.01$).

4.2. Analysis of the choice task

Recall that the current experiments involve choices between prospects under five distinct payoff magnitudes. The payoff magnitude is determined by the value of the payoff magnitude parameter (labeled with x) in our problem-generating algorithm. For example, in Experiments 1 and 2 when $x = 1$, the difference between H and L is 20, and when $x = 5$ the difference between H and L is 100. Table 3 presents the mean R-rates in the decisions under risk conditions as a function of x over the three experiments, and the correlation between choices under the five different values of x (using subjects as a unit of analysis). The results reveal two interesting patterns. First, the mean rates reveal a U shape: The highest R-rate is in correspondence of with $x = 1$, and the second highest rate corresponds to $x = 5$. The difference between the R-rates given $x = 4$ and $x = 5$ is significant ($t_{59} = 3.28$, $p < 0.01$), as well as between

⁹ To clarify this measure, consider the task of predicting the probability that the next estimate of the 0.05 probability event, provided by a certain subject, will be accurate. Assume that you can base your estimate on the observed accuracy rate in the first n estimates made by this subject, and on a point prediction made by the current model. It is easy to see that the value of the observed rate increases with n . ENO of 6.15 means that prediction of the model is slightly more accurate than the prediction based on the observed rate with $n = 6$. Since the parameters of the models were estimated based on only 8 observations per event type, this value is close to the maximal possible level.

the R-rates given $x = 1$ and $x = 2$ ($t_{59} = 4.08$, $p < 0.01$). Second, the correlation between the R-rates with $x = 1$ and $x = 5$ is low (0.21) and not significant. The correlations between closer x -values are positive and significant. The highest value (0.89) is the correlation between $x = 4$ and $x = 5$.

Table 3. Mean R-rates and correlations (Spearman's rho, p-values reported in parentheses) between R-rates for the five levels of parameter x (payoff magnitude).

Average R-rate	Levels of x	Correlation (p value)			
		2	3	4	5
0.58	1	0.64	0.41	0.24	0.21
		(0.00)	(0.00)	(0.07)	(0.11)
0.44	2		0.73	0.63	0.61
			(0.00)	(0.00)	(0.00)
0.42	3			0.88	0.85
				(0.00)	(0.00)
0.39	4				0.89
					(0.00)
0.46	5				

The models discussed above cannot capture Table 3's "magnitude effects". Indeed, these models ignore the effect of payoff magnitude, and imply high correlation between the different magnitude conditions¹⁰.

Nevertheless, it is possible to design members of the family of noisy retriever models that can capture these magnitude effects. The correlation pattern can be captured with two possible additions to the I2-OTO model. The first assumes that the payoff magnitude affects sampling (people tend to sample from similar x -levels), whereas the second assumes that payoff magnitude affects the nature of inertia (people tend to repeat the last choice made in a similar x level). The data show clear support for the "magnitude-based inertia" hypothesis: The overall proportion of "repeating the last choice in the same x level" is 0.77, and the proportion of "repeating the very last choice" is only 0.72 (see Table 4). Evaluation of the "magnitude based-sampling" hypothesis does not provide clear results. For example, the proportion of choices in trials 51 to 60 (over the 60 subjects) that reflects best

¹⁰ This shortcoming, of course, is not unique to the noisy retrieval models discussed above. For example, like these models, prospect theory predicts insensitivity to payoff magnitude.

replay to all previous trials is 0.59, and the proportion that reflects best replay to previous trials with the same x level is 0.55.

The U-shape effect of payoff magnitude can be captured with the assertion that it reflects experiences that occurred before the current experiment. For example, it is possible that people had better experiences with the risky prospect with the smallest variance. However, we do not think that the current results can be used to evaluate this hypothesis.

Table 4. Inertia in subjects' choice behavior.

	Experiment 1	Experiment 2	Experiment 3	Mean
P(Repeat very last choice)	0.68	0.71	0.78	0.72
P(Repeat last choice with the same x level)	0.70	0.75	0.84	0.77
P(Repeat last choice when $x = 1$)	0.75	0.73	0.82	0.77
P(Repeat last choice when $x = 5$)	0.71	0.74	0.86	0.77

5. Summary and Discussion

Previous research documents two pairs of contradicting reactions to rare events. One pair emerges in studies of probability judgment. These studies reveal conservatism that implies overestimation of rare events, and overconfidence that implies the opposite bias. A second pair emerges in studies of choice behavior: The results reveal overweighting of rare events in decisions from description, and the opposite bias in decisions from experience. The current analysis makes three steps ahead in clarifying this pattern. The first step shows that all four phenomena can be captured by the assumption that people tend to rely on a small set of experiences in similar situations. This assumption implies a family of noisy retriever models that highlights sufficient conditions for the different biases.

The second step is the exploration of when and why people tend to rely on past experiences during decisions under risk, when these experiences do not add useful information. The results suggest that subjects behave as if they doubt the description of the incentive structure. One likely contributor to this doubt is overgeneralization from “decisions with estimated risk” to “decisions under risk”. The coexistence of adaptive learning and this overgeneralization suggests that

decisions under risk are the product of two related biases: overconfidence (e.g., events estimated with 5% tend to occur in 20% of the times), and initial “over-debiasing” (e.g., an initial tendency to treat the objective probability of “5%” as an indication of higher probability). Thus, overconfidence in probability judgments implies underestimation, but leads to overweighting of rare events in decisions under risk.

The last step in our analysis focuses on individual differences and compares different members of the family of noisy retriever models. It shows the value of models that imply large individual differences, strong inertia, and allow for non-monotonic sensitivity to payoff magnitude.

5.1. Relationship to prospect theory

Kahneman and Tversky (1979) capture the tendency to overweight rare events in decisions under risk with the inverse S-shape of the subjective probability weighting function assumed by prospect theory. Their analysis focuses on the fact that this function explains important puzzles (including the Allais paradox, and the tendency to buy lotteries and insurance), but does not try to explain it.

Subsequent research provides three interesting explanations for the inverse S-shape of the probability weighting function. One explanation assumes a bias toward equal weighting of all the listed outcomes (Viscusi, 1989; Birnbaum, 2008; Erev et al. 2008). A second explanation (Stewart et al., 2006) assumes that target (probability) values are evaluated by their relative rank in a sample (of past stated probabilities) drawn from memory. The inverse S-shape is explained by the fact that large and small probability values are more frequent than those about 50%. A third explanation (Gayer, 2010) assumes a similarity-based process for probability estimation. The stated probability of an event is the weighted average of past stated event probabilities and realizations, a process commonly known in statistics as kernel estimation.

The current analysis implies a fourth explanation to the observation that an inverse S-shape weighting function best describes decisions under risk. Our explanation differs from the explanations provided by Stewart et al. and Gayer with respect to the nature of the experiences that affect behavior. Whereas Stewart et al. and Gayer focus on the effect of past “descriptions”, the current explanation focuses on the effect of past “outcomes”. The main advantage of the current explanation is its

ability to capture the observation that experience with immediate feedback leads to underweighting of rare events (that is, reverses the implied shape of the weighting function).

One advantage of the explanations proposed by Stewart et al. and Gayer is their ability to capture the fact the inverse S-shape of the weighting function crosses the 45° line when the objective probability is close to 0.3 (c.f. Figure 10). The low cutting point is a product of the assumption that most (experienced) objective probabilities are lower than 0.5 (because the risk implies at least two mutually exclusive events). This insightful assumption can be used to refine the Noisy Retriever models presented here. The 45° crossing point of the implied weighting function, under the current models, is determined by the distribution of the objective probabilities of the estimated events. This point is clarified in Figure 10. Each curve in this figure presents the weighting function implied by the basic Noisy Retriever model under different distributions of the objective probabilities. The results reveal that the crossing point of the 45° line decreases with an increase in the proportion of rare events.

Fig. 10 The implied certainty equivalent of a gamble that provides “1 with probability p ; 0 otherwise” under the noisy retrieval model (I-OTO model, with the parameters specified in Table 2), given two distinct distributions of objective probabilities. Distribution “Uniform” is the focus of the previous discussion; it implies equal relative frequency (20%) of the five values 0.05, 0.25, 0.5, 0.75 and 0.95. Under the distribution “More low” there are more experiences with low objective probabilities. Specifically, the relative frequencies were 36%, 28%, 20%, 12% and 4% for the objective probabilities 0.05, 0.25, 0.5, 0.75 and 0.95, respectively. The implied certainty equivalence function under the “More low” distribution is similar to the function implied by prospect theory.

Conclusions

At first glance, the assertion that “people tend to rely on past experiences in similar situations” does not seem to be very interesting and/or important. However, a more careful analysis of this assertion reveals that it can be used to predict the conditions under which interesting phenomena are likely to occur. These phenomena include: “human judgment reflects overconfidence” (Lichtenstein et al., 1982), “revision of opinion is biased by conservatism” (Slovic and Lichtenstein, 1971), “accidents are driven by the belief that it won’t happen to me” (Barron and Erev, 2003), and “an event has greater impact when it turns impossibility into possibility” (Fox and Tversky, 1995). Although these results might suggest that the implied effect of rare events is situation specific, they can be predicted with simple, general models.

References

- Allais, M. (1953): “Le Comportement de l’Homme Rationnel Devant le Risque: Critique des Postulats et Axiomes de l’Ecole Americaine,” *Econometrica*, 21 (4), 503–546.
- Barron, G., and I. Erev (2003): “Small Feedback-Based Decisions and Their Limited Correspondence to Description-Based Decisions,” *Journal of Behavioral Decision Making*, 16 (3), 215–233.
- Birnbaum, M. H. (2008): “New Paradoxes of Risky Decision Making,” *Psychological Review*, 115 (2), 463–501.
- Camerer, C. F., and T. H. Ho (1999): “Experience-Weighted Attraction Learning in Normal Form Games,” *Econometrica*, 67 (4), 827–874.
- Cooper, D. J., and J. H. Kagel (2003): “Lessons Learned: Generalizing Learning Across Games,” *American Economic Review*, 93 (2), 202–207.
- Erev, I., E. Ert, and A. E. Roth (2010a): “A Choice Prediction Competition for Market Entry Games: An Introduction,” *Games*, 1 (2), 117–136.
- Erev, I., E. Ert, A. E. Roth, H. Haruvy, S. M. Herzog, R. Hau, R. Hertwig, T. Stewart, T., R. West, and C. Lebiere (2010b): “A choice Prediction Competition: Choices from Experience and from Description,” *Journal of Behavioral Decision Making*, 23 (1), 15–47.
- Erev, I., I. Glozman, and R. Hertwig (2008): “What Impacts the Impact of Rare Events,” *Journal of Risk and Uncertainty*, 36 (2), 153–177.
- Erev, I., and A. E. Roth (1998): “Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed-Strategy Equilibria,” *American Economic Review*, 88 (4), 848–881.
- Erev, I., A. E. Roth, R. L. Slonim, and G. Barron (2007): “Learning and Equilibrium as Useful Approximations: Accuracy of Prediction on Randomly Selected Constant Sum Games,” *Economic Theory*, 33 (1), 29–51.

- Erev, I., T. S. Wallsten, and D. V. Budescu (1994): "Simultaneous Over- and Underconfidence: The Role of Error in Judgment Process," *Psychological Review*, 101 (3), 519–527.
- Fischhoff, B., P. Slovic, and S. Lichtenstein (1977): "Knowing with Certainty: The Appropriateness of Extreme Confidence," *Journal of Experimental Psychology: Human Perception and Performance*, 3 (4), 552–564.
- Fox, C. R., and A. Tversky (1995): "Ambiguity Aversion and Comparative Ignorance," *Quarterly Journal of Economics*, 110 (3), 585–603.
- Fox, C. R., and A. Tversky (1998): "A Belief-Based Account of Decision under Uncertainty," *Management Science*, 44 (7), 879–895.
- Friedman, M., and L. J. Savage (1948): "Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56 (4), 279–304.
- Fudenberg, D., and D. K. Levine (1998): *The Theory of Learning in Games*. Cambridge, MA: The MIT Press.
- Gayer, G. (2010): "Perception of Probabilities in Situations of Risk: A Case Based Approach," *Games and Economic Behavior*, 68 (1), 130–143.
- Gilboa, I., and D. Schmeidler (1995): "Case-Based Decision Theory," *Quarterly Journal of Economics*, 110 (3), 605–639.
- Hertwig, R., and I. Erev (2009): "The Description–Experience Gap in Risky Choice," *Trends in Cognitive Sciences*, 13 (12), 517–523.
- Jessup, R. K., A. J. Bishara, and J. R. Busemeyer (2008): "Feedback produces divergence from prospect theory in descriptive choice," *Psychological Science*, 19 (10), 1015–1022.
- Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47 (2), 263–291.

- Lejarraga, T., and C. Gonzalez (2011): "Effects of Feedback and Complexity on Repeated Decisions from Description," *Organizational Behavior and Human Decision Processes*, 116 (2), 286–295.
- Lichtenstein, S., B. Fischhoff, and L. D. Phillips (1982): "Calibration of Probabilities: The State of the Art to 1980," in D. Kahneman, P. Slovic, A. Tversky (Ed.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 306–334). Cambridge, UK: Cambridge University Press.
- March, J. G. (1996): "Continuity and Change in Theories of Organizational Action," *Administrative Science Quarterly*, 41 (2), 278–287.
- Nuland, S. B. (2003): *The Doctors' Plague: Germs, Childbed Fever, and the Strange Story of Ignac Semmelweis*. New York: W. W. Norton & Co Ltd.
- Osborne, M. J., and A. Rubinstein (1998): "Games with Procedurally Rational Players," *American Economic Review*, 88 (4), 834–847.
- Phillips, L. D., and W. Edwards (1966): "Conservatism in a Simple Probability Inference Task," *Journal of Experimental Psychology*, 72 (3), 346–354.
- Selten, R., and T. Chmura (2008): "Stationary Concepts for Experimental 2x2 Games," *American Economic Review*, 98 (3), 938–966.
- Shi, L., T. L. Griffiths, N. H. Feldman, and A. N. Sanborn (2010): "Exemplar Models as a Mechanism for Performing Bayesian Inference," *Psychonomic Bulletin & Review*, 17 (4), 443–464.
- Slovic, P., and S. Lichtenstein (1971): "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *Organizational Behavior and Human Performance*, 6 (6), 649–743.
- Stewart, N., N. Chater, and G. D. A. Brown (2006): "Decision by Sampling," *Cognitive Psychology*, 53 (1), 1–26.
- Taleb, N. N. (2007): *The Black Swan: The Impact of the Highly Improbable*. New York: Random House and Penguin.

- Viscusi, K. W. (1989): "Prospective Reference Theory: Toward an Explanation of the Paradoxes," *Journal of Risk and Uncertainty*, 2 (3), 235–263.
- Viscusi, K. W., and J. K. Hakes (2008): "Risk Beliefs and Smoking Behavior," *Economic Inquiry*, 46 (1), 45–59.
- Wakker, P. P. (2010): *Prospect Theory for Risk and Ambiguity*. Cambridge, UK: Cambridge University Press.
- Yechiam, E., G. Barron, and I. Erev (2005): "The role of Personal Experience in Contributing to Different Patterns of Response to Rare Terrorist Attacks," *Journal of Conflict Resolution*, 49 (3), 430–439.