The coexistence of overestimation and underweighting of rare events and the contingent recency effect

Greg Barron*
Harvard Business School

Eldad Yechiam
Technion — Israel Institute of Technology

Abstract

Previous research demonstrates overestimation of rare events in judgment tasks, and underweighting of rare events in decisions from experience. The current paper presents three laboratory experiments and a field study that explore this pattern. The results suggest that the overestimation and underweighting pattern can emerge in parallel. Part of the difference between the two tendencies can be explained as a product of a contingent recency effect: Although the estimations reflect negative recency, choice behavior reflects positive recency. A similar pattern is observed in the field study: Immediately following an aversive rare-event (i.e., a suicide bombing) people believe the risk decreases (negative recency) but at the same time exhibit more cautious behavior (positive recency). The rest of the difference is consistent with two well established mechanisms: judgment error and the use of small samples in choice. Implications for the two-stage choice model are discussed.

Keywords: decision making, rare events, judgment, choice, learning, terror, probability.

1 Introduction

Studies of human reaction to low probability (rare) events reveal an interesting difference between judgment and decision-making in repeated settings. Judgments (probability estimations) appear to reflect over-sensitivity to rare events. That is, the estimated probability of events that occur with probability below 0.5 tends to be higher than the objective probability (see e.g., Erev, Wallsten & Budescu, 1994; Zacks & Hasher, 2002; Viscusi, 1992). On the other hand, decision-making from experience tends to reflect underweighting of (insensitivity to) rare events (Barron & Erev, 2003; Hertwig, Barron, Weber & Erev, 2004; Weber, Blais, & Shafir, 2004).[1] That is, decision-makers behave as if events that occur with probability below 0.5 occur with smaller probability than their objective probability. The apparent discrepancy is important in light of the two-stage choice model (Fox & Tversky, 1998) which assumes that choice can be predicted from estimated probabilities.[2] The main goal of the current paper is to improve our understanding of this pattern and its implications.[3]

1.1 The contradicting results

Ample experimental and field evidence suggests that subjective probability and frequency estimates reflect over-sensitivity to rare events. In their study on the judged frequency of lethal events, Lichtenstein, Slovic, Fischhoff, and Combs (1978) observed a consistent overestimation of the probabilities related to the rarest causes of death. A similar finding is that teens greatly overestimate the chances of death in the near future; they estimate the probability to be 18.6% when the actual probability is 0.04% (Fischhoff, Parker, Bruine De Bruin, Downs, Palmgren, Dawes, & Manski, 2000). Another remarkable example is that when Americans were asked to estimate the probability that a smoker would develop lung cancer in the future, the mean estimate was 38% whereas the actual probability is between 6% and 13% (Viscusi, 1992).

---

[1]Barron and Erev (2003) and Hertwig et al. (2004) clarify the difference between this behavioral tendency that occurs when people decide based on personal experience, and the important tendency to overweight low probability outcomes in one-shot decisions based on a description of the possible outcomes (see Tversky & Kahneman, 1992; Fox & Tversky, 1998).

[2]The model assumes that people first assess the probability of an uncertain event and then transform the assessment using Prospect Theory's weighting function. Fox & Hadar (2006) make the optimistic assertion that the two-stage model can account for experience-based decisions.

[3]The "Choice-Judgement discrepency" in one-shot decisions is an unrelated phenomona that refers to a different pattern of behavior, a preference to bet on A rather than on B even though B is judged to be at least as probable as A (Heath & Tversky, 1991).

Table 1: Four repeated choice problems and aggregated proportion of R choices reported in Barron and Erev (2003). The notation (v, p) denotes a gamble that provides v with probability p and 0 otherwise.

Problem	S	R	P(R)
1	(3, 1)	(32, 0.1)	0.28
2	(−3, 1)	(−32, 0.1)	0.60
3	(9, 1)	(10, 0.9)	0.56
4	(−9, 1)	(−10, 0.9)	0.37

Interestingly, smokers saw their choice to smoke as being consistent with their risk estimate (i.e., the pleasure is worth the risk), a view consistent with the theory of utility maximization.

Overestimation of rare events in field studies is typically explained by invoking the availability heuristic. Rare events (e.g., unique causes of death) that are more salient are easier to retrieve from memory, hence they are overweighted (see Tversky & Kahneman, 1974). The phenomenon is robust and is observed in controlled laboratory experiments even when long-term memory is not likely to play an important role. In one such study, Erev and Wallsten (1993) had subjects estimate the probability of an icon on the computer screen making its way safely to the other side of a continuously opening and closing sliding door. The amount of time that the door remained open (which determined the probability of success) was varied, and estimates were elicited based on the entire range of objective probabilities. The results indicated a clear overestimation of small success and failure probabilities. Erev et al. (1994) showed that a model assuming that error is added to subjective probabilities can capture the overestimation phenomena. Note that this assumption is consistent with both the "regression effect" (Stevens & Greenbaum, 1966) and the "contraction bias" (Poulton, 1979) that describe shifts in responses towards the middle of a range.

A very different effect of rare events was observed in studies of decisions from experience. These studies (e.g., Barron & Erev, 2003; Hertwig et al., 2004; Weber et al., 2004; Erev & Barron, 2005; Yechiam & Busemeyer, 2006) reflect underweighting of rare events. Table 1 shows four conditions from Barron and Erev (2003) where subjects repeatedly chose between two unmarked buttons that provided outcomes sampled from two distributions, "S" and "R". Let (v, p) denote a distribution where the outcome v occurs with probability p (otherwise zero). The right hand column shows the aggregated proportion of R choices over all trials (400 in Conditions 1 and 2 and 200 in Conditions 3 and 4) with immediate feedback.

Although the results of Conditions 1 and 2 can be accounted for by risk aversion in the gain domain and risk seeking in the loss domain, Conditions 3 and 4 imply the opposite, that decision makers appear to take more risk in the gain than in the loss domain. All four results are consistent with underweighting of small probabilities (of receiving 32 in Conditions 1 and 2 and of receiving 0 in conditions 3 and 4). Further research supports the conjecture that underweighting is, at least in part, the result of a tendency to rely on small samples when making experience based decisions (Hertwig et al., 2004; Erev & Barron, 2005; Erev, Ert, & Yechiam, 2008; Yechiam & Busemeyer, 2006).[4] Hertwig et al. (2004) showed for instance that subjects' choices were significantly associated with their most recent outcomes, suggesting a reliance on only part of the sampled choice outcomes. The tendency to rely on small samples is also consistent with existing research on information search and the perception of variability (Kareev, 2000; Kareev, Arnon, & Horwitz-Zeliger, 2002).

Recent studies have attempted to measure both judgment and choice at the same time in the context of rare events (e.g., Fox & Hadar, 2006; Hau, Plescak, Kiefer & Hertwig, 2008; Ungemach, Chater, & Stewart, 2009). In contrast to the current paper, these experiments studied one-shot decisions based on repeatedly drawn samples. Overall, they reported evidence for underweighting in choice while subjects remained well calibrated in their estimations, especially for larger samples of outcomes. Although statistically insignificant, a slight tendency towards overestimating small probabilities was also observed. Because these studies lacked individual level analyses, it is difficult to know if the subjects who overestimated rare events were also those who underweighted the events in choice.

The current paper's main contributions are as follows. First, we demonstrate simultaneous overestimation of probabilities and underweighting in choice at the individual subject level. As noted earlier, Hau et al. (2008) and Ungemach et al. (2009) did not demonstrate that individual subjects displayed both biases at the same time and these could in fact be two separate groups of individuals within the sample. Secondly, we provide evidence for specific underlying mechanisms that can explain, at least in part, the coexistence of overestimation and underweighting within individual subjects. Finally we extend the findings of Hau et al. (2008) and Ungemach et al. (2009) to a different experience-based paradigm, namely

---

[4]It is still debated whether such reliance constitutes completely ignoring events that are not in the sample (e.g., Hertwig et al., 2004) or just decreasing their relative weight in the decision (Yechiam & Busemeyer, 2006). The latter view is more consistent with traditional learning models such as those of Bush and Mosteller (1955) and March (1996).

repeated decisions with immediate feedback. This is different from the sampling paradigm used in these earlier studies, where a single decision is made based on a sample observed over time (with no monetary implications).

1.2 The coexistence hypothesis

Our interpretation of the results is referred to as the *coexistence* hypothesis. It assumes that there are qualitative, yet simultaneous, differences in the effect of rare events on judgment and decision processes. As noted above, these differences have been well studied: Rare events are overestimated due to their increased availability in memory but are underweighted in choice due to the tendency to rely on small samples in experience based decisions. Although coexistence seems a reasonable hypothesis given the prior research, it has not been shown within subjects in previous research and is inconsistent with the two-stage model of choice that predicts a certain consistency between estimates and choices. Specifically, in assuming Prospect Theory's weighting function, the two-sage model predicts that choice will reflect overweighting of estimated small probabilities.

Another reason to predict coexistence pertains to recency effects. Barron and Erev (2003) note that the tendency to underweight rare events in choice can be a product of a positive recency effect: Oversensitivity to recent outcomes (i.e., one type of small sample as suggested by Hertwig et al., 2004). This explanation implies that, since rare outcomes are less likely to occur recently or in any cognitively limited small sample, on average they will be underweighted in choice. In contrast, judgment tasks typically produce evidence for negative recency (or "gamblers fallacy") in prediction tasks. (See the review in Lee, 1971, and recent research by Sundali & Croson, 2006.) The above logic implies that overestimation can be a product of a negative recency effect in estimation tasks. Negative recency (the expectation that the state of the world will change between sequential trials) implies overestimation of the probability of the event that did not occur in the last trial.

This prediction is supported by Ayton and Fischer's (2004) study. In their experiment subjects repeatedly predicted the outcome of a roulette spin (red or blue with equal probability) and indicated their confidence level (from "no confidence" to "strong confidence") in the prediction. Although the results demonstrated negative recency in the prediction task, simultaneous positive recency was observed for subjects' confidence in their predictions. That paper concluded that sequences of outcomes reflecting human performance yield anticipations of positive recency, whereas outcomes due to inanimate chance mechanisms, such as coins, dice and roulette wheels, yield anticipations of negative recency. The *con-tingent recency effect* implies that these results will be robust to simultaneous choice and judgment in contexts where, outside of Las Vegas, people do not have precise information regarding the dependency of outcomes.

2 Study 1: The Coemergence of Overestimation and Underweighting

To evaluate the three alternative explanations, Study 1 examined both judgments and choices in the same context. Subjects performed a repeated choice task in which one of the alternatives included a rare (low probability) event of a negative payoff (loss of points). During the second half of the task they were asked to estimate the probability of this event.

2.1 Method

2.1.1 Design

In a within-subject design, each subject performed a binary choice task and a probability assessment task. The binary choice task was performed under uncertainty for 100 rounds, with immediate feedback. The probability assessment task following each choice in rounds 51–100. Upon completion, subjects performed a one-time retrospective probability assessment task.

In the binary choice task, subjects chose between two unmarked buttons presented on the screen (see Appendix). Each button was associated with one of two distributions referred to here as S (for safe) and R (for risky). The S distribution provided a certain loss of 3 points while the R distribution provided a loss of 20 points with probability 0.15 and zero otherwise. Thus, the two distributions had equal expected value. To reduce noise and sampling error, random sequences of 100 outcomes were produced repeatedly and the first sequence with an observed probability of 0.15 for the −20 outcome was used for all subjects. The sequence provided the −20 outcome on rounds 12, 15, 19, 20, 21, 23, 25, 35, 40, 41, 60, 73, 80, 87, and 96. The position of the S and R buttons (right vs. left) was randomly determined for each subject. At the conclusion of the study, points were converted to monetary payoffs according to the exchange rate: 100 points = 1 Shekel (about 18 US cents), and were subtracted from the show up fee.

In the probability assessment task, performed after each binary choice in trials 51–100, subjects were prompted to estimate the chances (in terms of a percentage between 0 and 100) of −20 appearing (on the R button) on the next round.

After completing 100 rounds subjects were asked to estimate ("end-of-game estimates"), to the best of their recollection, two conditional probabilities: (1) the chances of -20 appearing after a previous round with a -20 outcome $[SP(-20|-20)]$ and (2) the chances of -20 appearing after a previous round with a 0 outcome $[SP(-20|0)]$.

2.1.2 Subjects

Twenty-four Technion students served as paid subjects in the study. Most of the subjects in this and the other studies described in this paper were second and third year industrial engineering and economics majors who had taken at least one probability or economics course. In addition to the performance contingent payoff, described above, subjects received 28 Shekels for showing up. The final payoff was approximately 25 Shekels (about $5 US).

2.1.3 Apparatus and procedure

Subjects were informed that they were operating a "computerized money machine" (see a translation of the instructions in the Appendix) but received no prior information as to the game's payoff structure. Their task was to select one of the "machine's" two unmarked buttons (see the figure in the Appendix) in each of the 100 trials. In addition, they were told that they would be asked, at times, to estimate the likelihood of a particular outcome appearing the following round. As noted above, this occurred in trials 51–100.

Subjects were aware of the expected length of the study (10–30 minutes), so they knew that it included many rounds. To avoid an "end of task" effect (e.g., a change in risk attitude), they were not informed that the study included exactly 100 trials.[5] Payoffs were contingent upon the button chosen; they were produced from the predetermined sequence drawn from the distribution associated with the selected button, described above. Three types of feedback immediately followed each choice: (1) the payoff for the choice, which appeared on the selected button for the duration of 1 second, (2) payoff for the forgone option, which appeared on the button not selected for the duration of 1 second and (3) an update of an accumulating payoff counter, which was constantly displayed.

---

[5]Not knowing the length of the study also prevents subjects from using probability-based reasoning (the focus on the likelihood of achieving a particular aspiration level) (Lopes, 1996). This type of reasoning bases choice on the probability of coming out ahead, which is a function of the number of choices to be made. A second reason for not telling subjects the game's length is that this better approximates the real-world small decisions that interest us. In such situations, the number of future choices to be made is often unknown and it is difficult to prescribe optimal behavior.
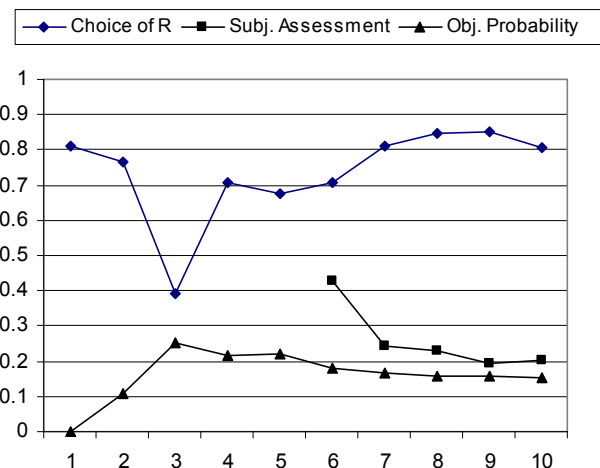

Figure 1: Mean proportion of R choices in Study 1, mean subjective assessments in trials 51–100, and observed objective probability of the rare outcome in 10 blocks of 10 trials. Objective probability is the observed proportion of trials where the rare outcome was observed, recalculated after each trial.

2.2 Results

2.2.1 Judgment and choice in the same context

The aggregated assessments and proportion of R (risky) choices are shown in Figure 1. The mean probability assessment from trials 51–100, aggregated over trials and over subjects, was 0.27. This value is significantly larger than 0.163, the mean running average of the observed probability of the -20 outcome (t[23] = 3.11, p < 0.01).[6] Thus, the results reflect overestimation of the rare event.

As shown in Table 2, over all 100 trials, subjects' aggregate proportion of R choices was 0.74 (significantly larger than 0.5, t[23] = 7.47, p < 0.001). This result is consistent with the assertion of underweighting of rare events in choice. The rate of R choice over trials 51–100 was 0.80 (significantly larger than 0.5, t[23] = 6.78, p < 0.001).

The comparison of the judgment and choice data for trials 51–100 supports the "coexistence" hypothesis. The results demonstrated different reactions to rare events in judgment and in choice within the same context.

We next asked whether the different reactions occur at the level of the individual subject. For 63% (15/24) of the subjects, assessment and choice results were not consistent in terms of the implied weighting of the -20 outcome aggregated over trials 51–100. Overestimation

---

[6]Although the rare events' probability was 0.15, the mean running average of its observed probability can be considerably higher if it occurs more often early on in the sequence. To see this, consider the simple sequence [1, 0]. While the mean is 0.5, the mean running average is $(1 + 0.5)/2 = 0.75$.

Table 2: A summary of the aggregate results of Study 1. The number in parenthesis denotes the null hypothesis for the test reported in the text. All t-tests are one-sample tests unless otherwise noted.

Statistic	Trials 1–100	Trials 51–100	Retrospective
P(R): proportion of R choices	0.74(0.5**)	0.80(0.5**)	
SP(−20): Mean subjective assessment of the probability of a −20 outcome	–	0.27(0.163**)	
P(R|−20): Prop. of R choices after a trial with a −20 outcome	0.56 (paired 0.77**)	0.74 (paired 0.81$^\diamond$)	
P(R|0): Prop. of R choices after a trial with a 0 outcome	0.77	0.81	
SP(−20|−20): Mean assessment of the probability of a −20 outcome after a trial with a −20 outcome		0.18 (paired 0.28*)	0.08 (paired 0.26**)
SP(−20|0): Mean assessment of the probability of a −20 outcome after a trial with a 0 outcome		0.28	0.26

$^\diamond$p<0.1, *p<0.05, **p<0.01, ***p<0.001.

and underweighting of rare events was found to occur in 100% of these 15 cases.

2.2.2 The contingent recency effect

The central column in Table 2 presents the mean judgment and choice over trials 51–100 and presents the results conditional on the most recent outcome. Although the proportion of R choices in trials after an outcome of 0, aggregated for each subject over all 100 trials, was 0.77, it dropped significantly to 0.56 for trials after an outcome of −20 appeared (paired t-test, t[23] = 5.66, p < 0.01). Similar, but slightly weaker evidence of positive recency was observed in trials 51–100 (the same trials analyzed above) with [P(R|0)] = 0.81 and [P(R|−20)] = 0.74, (paired t-test, t[23] = 1.83, p = 0.08).

In order to evaluate the recency effect on judgment we first computed mean conditional subjective probability assessments, SP(−20|−20) and SP(−20|0), for each subject by aggregating separately estimates from rounds after a −20 outcome, and the estimates from rounds after a 0 outcome. The results (see Table 2) revealed a negative recency effect, with subjects judging the −20 outcome less likely after a previous outcome of −20 (SP(−20|−20) = 0.18 and SP(−20|0) = 0.28, paired t-test, t[23] = 1.99, p < 0.05). This result is interesting considering that the conditional objective probability OP(−20|−20) was *larger* than OP(−20|0). Note also that even in trials that occur after an appearance of the rare outcome, the subjective assessment (0.18) is still overestimated.

Examination of the retrospective estimation of SP(−20|−20) and SP(−20|0) at the end of 100 rounds

show a similar negative recency pattern. The estimated probabilities are 0.08 and 0.26 respectively. Thus, subjects judged the −20 outcome to be less likely after a previous −20 outcome (paired t-test, t[23] = 3.26, p < 0.01).

The contingent recency effect described above cannot by itself explain the observed overestimation and underweighting in choice. As noted earlier, the mean estimation immediately following a rare event (SP(−20|−20) = 0.18) was lower than the mean estimation following a frequent event, but still reflected overestimation (of the objective probability). And the proportion of R choices was higher than 0.50 (0.74) even after the −20 outcome. A second relevant observation is the correlation across subjects between judgment, SP(−20), and choice of R for each trial for which estimations were given (trials 52 to 100).[7] Computation of this correlation by experimental trial reveals negative correlations in 36 of the 49 trials (p< 0.001 in a sign-test). Thus, while the results supported the coexistence hypothesis, there remained a consistency between judgments and choices, as subjects tended to avoid option R when they judged the probability of a loss to be high.

3 Study 2: Generality over payoff domain and payoff rule

Although Study 1's results are consistent with the contingent recency effect, an alternative explanation remains

---

[7] Assessments were elicited starting from trial 51 so that the first choice following an estimate is on trial 52.

for the finding of positive recency for choice. In particular, subjects may place a different value on a certain loss immediately following a preceding loss from the risky option (a de-sensitization effect). In Study 2 we examined this possibility by paying subjects according to the outcome of a single trial drawn at random at the end of the game. By replicating Study 1 in both the gain and loss domains, we also tested the hypothesis that the preference for option R in Study 1 might reflect a tendency to avoid alternatives with a larger proportion of losses (as was observed in Erev & Barron, 2005).[8] Additionally, outcomes in Study 2 were randomly drawn from the distributions described next (without the pre-selection of a single series that was employed in Study 1) and the study was conducted for 400 trials. The distributions were also changed so as not to include zero, as several studies have demonstrated unique behavior related to zero outcomes or costs (Ariely, Gneezy, & Haruvy 2005; Festinger & Carlsmith, 1959).

3.1 Method

3.1.1 Design

The design was the same as for Study 1 with the exception that outcomes were randomly drawn in real-time, the study was run for 400 trials (with assessments elicited on trials 201–400 and not at the end) and subjects were paid according to one randomly chosen trial. In the Loss condition the S distribution provided a certain loss of 1.3 points while the R distribution provided a loss of 3 points with probability 0.15 and a loss of 1 point otherwise. Thus, the two distributions had equal expected value. For the Gain condition, a constant of 4 was added to all payoffs so that S provided (2.7, 1) and R provided (3, 0.85; 1).

3.1.2 Subjects

Forty Technion students served as paid subjects in the study. In addition to the performance contingent payoff, subjects in the Gain and Loss conditions received 25 Shekels or 29 Shekels for showing up. The conversion rate for the one randomly chosen trial was 1 point = 1 Shekel. The final average payoff was approximately 27 Shekels (about 5 US dollars).

3.1.3 Apparatus and procedure

The task and instructions were as in Study 1 except that the subjects were told that they would be paid according to one randomly sampled trial at the end of the experiment.

[8]Loss Aversion, as quantified by Prospect Theory, would not imply a preference for R since both S and R are in the loss domain.
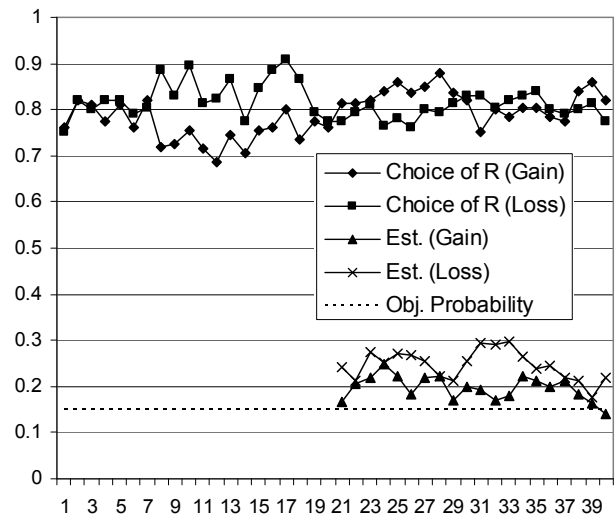
Figure 2: Mean proportion of R choices in Study 2, and subjective assessments of the rare outcome's probability in 40 blocks of 10 trials.

3.2 Results

3.2.1 Judgment and choice in the same context

The results reveal the same pattern observed in Study 1. Figure 2 presents the subjects' aggregate proportion of R choices and probability assessments in 40 blocks of 10 trials. Across all 400 trials and two conditions, subjects' mean proportion of R choices was 0.80 (significantly larger than 0.5, $t[39] = 8.92$, $p < 0.001$), consistent with the underweighting of rare events in choice behavior. In trials 201–400 (see Table 3), when probability assessments were also elicited, the mean proportion of R choices was 0.81 (greater than 0.5, $t[39] = 7.17$, $p < 0.001$) again consistent with the underweighting of rare events in choice. Consistent with the visual impression in Figure 2, there was no significant difference between the Gain and Loss conditions ($t[38] = 0.21$, ns).

The mean probability assessment from trials 201–400, aggregated over trials and conditions, was 0.22 (see the third row of Table 3 and Figure 2). This is significantly larger than 0.15, the objective probability of the rare outcome ($t[39] = 3.35$, $p < 0.01$). This result is consistent with an overestimation of rare events in probability assessments. There was no significant difference in the probability assessments between the Gain and Loss conditions ($t[38] = 1.15$, ns).

At the individual level, for 55% (22/40) of the subjects assessment and choice results were inconsistent in terms of the implied weighting of the rare outcome (1 in the Gain condition and −3 in the Loss condition) aggregated over trials 201–400. As can be seen in Table 4, overestimation and underweighting of rare events was found to

Table 3: A summary of the results of Study 2 aggregated over Gain and Loss conditions. The number in parenthesis denotes the null hypothesis for the test reported in the text. All t-tests are one-sample tests unless otherwise noted.

Statistic	Trials 1–200	Trials 201–400
P(R): proportion of R choices	0.79(0.5***)	0.81(0.5***)
SP(LowProb): Mean subjective assessment of the probability of the rare outcome	–	−0.22(0.15***)
P(R|LowProb): Prop. of R choices after a trial with a rare outcome	0.71	0.77 (paired 0.82**)
P(R|HighProb): Prop. of R choices after a trial with the high probability outcome	0.81	0.82
SP(LowProb|LowProb): Mean assessment of the probability of the rare outcome after a trial with a rare outcome		0.19 (paired 0.23**)
SP(LowProb|HighProb): Mean assessment of the probability of a rare outcome after a trial with a high probability outcome		0.23

*p<0.10, **p<0.05, ***p<0.01.

occur for 91% of these 22 subjects (p < 0.001, McNemar's test)

3.2.2 The contingent recency effect

The second column of Table 3 (rows 3–6) presents the mean judgment and choice over trials 201–400 conditional on the most recent outcome. For each subject we calculated two proportions, the proportion of R choices following an observation of the rare event (aggregated over trials 201–400) and the proportion of R choices following observations of the more common outcome. Aggregating over both conditions, a significant amount of positive recency for choice was observed; subjects were 5% less likely to choose R on trials that immediately followed an observation of the rare event (i.e., the bad outcome) (t[39] = 2.49, p < 0.05). On those same trials, the mean assessment of the rare event was 4% lower (i.e., they estimated them as *less* likely) than on trials not following an observation of the rare event (t[39] = 2.29, p < 0.05), which is consistent with negative recency. No significant difference was found between the Gain and Loss conditions (t[38] = 1.40, n.s., for positive recency and (t[38] = 0.97, n.s., for negative recency). As was the case in Experiment 1, the pattern of positive recency for choices and negative recency for probability assessments was consistent with the contingent recency hypothesis.

The contingent recency effect contributes to, but cannot explain by itself, the main results, since overestimation and underweighting were observed even immediately after observing the rare event. The average estimation in these trials was 0.19, and the proportion of

Table 4: Counts of individuals aggregated choice of R relative to 0.5 in trials 201–400, and subjective probability estimation (SP), relative to 0.15 in Study 2.

	P(R) > 0.5	P(R) < 0.5
SP(R) < 0.15	13	2
SP(R) > 0.15	20	5

R choices was 0.77. Additionally, while both overestimation and underweighting were concurrently observed there was also an overall consistency between judgments and choices. An examination of the association between the mean choice rate of R and mean estimation (trials 201 to 400) over the 40 subjects reveals a correlation of $r(38)$ = -0.48, p < 0.01.

A within-person contingent recency effect (positive recency in choice and negative recency in estimations) was found to occur for 11 of the 40 subjects. While only 5 subjects displayed the opposite tendency, negative recency in choice and positive in estimations, the difference in counts was not significant. Finally, a within-person correlation between judgment and choices in trials 201 to 400 showed negative correlations for most of the subjects (19 of 33)[9], again reflecting consistency between judgment and choice.

[9]Correlations could not be computed for seven of the 40 subjects who chose R in every trial between 201–400.

3.2.3 Framing as an alternative mechanism for over-estimation

In both Experiments 1 and 2 the rare event provided a worse outcome than the more common result from the risky distribution. These, comparatively bad, outcomes may have been framed as losses by subjects. If "losses loom larger than gains" (Kahneman & Tversky, 1979) then these outcomes may have been more salient in memory than the relative gains, and subjects may have overestimated their probability for this reason. It is desirable to differentiate between this mechanism, the increased availability of losses, and the mechanism we assumed based on previous research: the increased availability of all rare events for probability assessments (and the addition of error to subjective judgments). Study 3 was designed as a test of these two mechanisms.

4 Study 3: Generality over framing of rare events

If loss aversion (relative to a reference point) is the prime driver of the observed discrepancies in Studies 1 and 2, the effect should diminish when the rare event is framed as a good outcome.

4.1 Method

4.1.1 Design

The design was the same as for Study 2 with the exception that there was only one condition where the S distribution provided a certain gain of 2.7 points while the R distribution provided a gain of 18 points with probability 0.15 and 0 points otherwise. Thus, the expected values and the S distribution were identical to those used in the Gain Condition of Study 2. The change is that the rare event (18 points) was a relatively *good* outcome.

4.1.2 Subjects

Twenty Technion students served as paid subjects in the study. In addition to the performance contingent payoff, subjects received 25 Shekels for showing up. The conversion rate for the one randomly chosen trial was 1 point = 1 Shekel. The final average payoff was approximately 27 Shekels (about 5 US dollars).

4.1.3 Apparatus and procedure

As in Study 2.
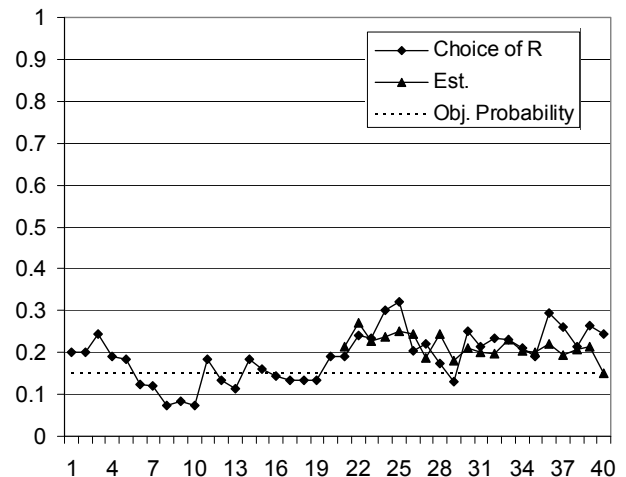
Figure 3: Mean proportion of R choices in Study 3, and subjective assessments of the rare outcome's probability in 40 blocks of 10 trials.

4.2 Results

4.2.1 Judgment and choice in the same context

The results revealed the same pattern observed in Studies 1 and 2, namely, a robust underweighting in choice along with overestimation. Figure 3 presents subjects' aggregate proportion of R choices and probability assessments in 40 blocks of 10 trials. Over all 400 trials, subjects' mean proportion of R choices was 0.19 (significantly smaller than 0.5, t[19] = 32.32, p < 0.001), consistent with the underweighting of rare events in choice behavior. In trials 201–400 (see Table 5), when probability assessments were also elicited, the mean proportion of R choices in these trials was 0.23 (less than 0.5, t[19] = 26.91, p < 0.001) again consistent with the underweighting of rare events in choice.

The mean probability assessment from trials 201–400 aggregated over trials and conditions was 0.21 (see the second row of Table 5 and Figure 3). This is significantly larger than 0.15, the objective probability of the rare outcome (t[19] = 10.64, p < 0.001). This result is consistent with an overestimation of rare events in probability assessments.

At the individual level, for all 20 subjects, assessment and choice results were not consistent in terms of the implied weighting of the 18 outcome, aggregated over trials 201–400. Overestimation and underweighting of rare events was found to occur for every subject.

In summary, even when the rare event is a relatively good outcome, we found robust overestimation and underweighting of rare events, as predicted by the coexistence hypothesis. The result is consistent with the assumption that overestimation reflects the greater saliency of rare events rather than the salience of negative events.

Table 5: A summary of the results of Study 3. The number in parenthesis denotes the null hypothesis for the test reported in the text. All t-tests are one-sample tests unless otherwise noted.

Statistic	Trials 1–200	Trials 201–400
P(R): proportion of R choices	0.15(0.5***)	0.23(0.5***)
SP(LowProb): Mean subjective assessment of the probability of the rare outcome	–	0.21(0.15***)
P(R|LowProb): Prop. of R choices after a trial with a rare outcome	0.27 (paired 0.13***)	0.29 (paired 0.22*)
P(R|HighProb): Prop. of R choices after a trial with the high probability outcome	0.13	0.22
SP(LowProb|LowProb): Mean assessment of the probability of the rare outcome after a trial with a rare outcome		0.22
SP(LowProb|HighProb): Mean assessment of the probability of a rare outcome after a trial with a high probability outcome		0.21

*p<0.10, **p<0.05, ***p<0.01.

4.2.2 The contingent recency effect

The second column of Table 5 (rows 3–6) presents the mean judgment and choice over trials 201–400 conditional on the most recent outcome. A significant amount of positive recency for choice was observed; subjects were 7% more likely to choose R on trials that immediately followed an observation of the rare event (i.e., the good outcome) (t[19] = 1.65, p = 0.057). No significant tendency towards negative recency was observed in this condition, and the effect appears to be weaker when the rare outcome is relatively favorable.

5 Study 4: The effect of rare terrorist suicide attacks

Studies 1 through 3 focused on abstract low-stake decisions. They demonstrate that the well established mechanisms of judgment error and reliance on small samples can lead to the coexistence of overestimation and underweighting of rare events. The contingent recency effect contributes to this coexistence and was found in three out of four conditions tested, when the rare event was a relatively bad outcome. Study 4 was designed to evaluate the generality of this effect to events outside the laboratory in natural settings. It examines natural high-stake decisions where the rare event is clearly disastrous: Human reaction to suicide bombings in Israel.

During the al-aqsa intifada there was a period of 700 days (September 30, 2000 to August 31, 2002) in which suicide-bombing attacks were carried out on 71 different days (Associated Press, 2002). Immediately following this period, Israeli students were asked about their behavior and their probability assessments regarding the threat of suicide bombings during the intifada. The hypothesis was that, while students would assess the probability of an attack on the day after a previous attack to be lower than after an attack-free day (negative recency), they would choose to behave as if the probability had increased (positive recency).

5.1 Method

5.1.1 Design

Subjects were randomly assigned to one of two conditions: Choice (43 subjects) or Probability (42 subjects). The between subject design was chosen to eliminate the possibility that questions regarding choice behavior would affect probability assessments and vice-versa.

5.1.2 Subjects

In the summer of 2002, following the intifada, Eighty-five (46 males and 39 females) Technion students served as paid volunteers who came to fill out a number of unrelated questionnaires. Subjects were paid 40 Shekels (about 8 US Dollars) for their time.

5.1.3 Apparatus and procedure

In both conditions subjects answered three questions on 5-point scales. Subjects were instructed that the questions pertained to the events of the (then) recent intifada.

Figure 4: Mean response (on a 1 to 5 scale) to the three questions (on the x-axis) in conditions Choice and Probability. In Condition Choice the responses reflect the relative extent to which the responders felt cautious; in Condition Probability the responses reflect relative magnitude of estimated chance of attack.

The first question asked about days on which there was *no* attack on the previous day, the second question asked about days on which there *was* an attack on the previous day, but *without* fatalities. The third question asked about days on which there was an attack *with* fatalities on the previous day.[10] In the Choice condition subjects were asked about their behavior while in the Probability condition they were asked about their estimate. For example, in the Choice condition, the third question was:

"The day after a suicide bombing with fatalities, *I am cautious* about another suicide attack:"

Much **less** than usual   Much **more** than usual

| 1 | 2 | 3 | 4 | 5 |

The same question in the Probability condition was: "The day after a suicide bombing with fatalities, the chance of another suicide attack is:". The same five-point scale accompanied all three questions in both conditions.

5.2   Results

Figure 4 presents the mean response to the three questions in conditions Choice and Probability. As can be seen,

[10] The distinction between attacks with and without fatalities was introduced to capture the intuition that the media's differential coverage of the two events might have a different effect on subjects.

Figure 5: Percentage of days where a suicide bombing occurred according to the previous day for the period of September 30, 2000 - August 31, 2002, during the al-aqsa intifada.

subjects in the Choice condition reported more cautiousness after an attack with fatalities than after a day without an attack (3.56 and 2.58 respectively, t[42] = 4.35, p < 0.001). Yet, subjects in the Probability condition reported that they believe the chances of another suicide attack to be smaller in the day after an attack with fatalities than after a day without an attack (2.21 and 3.52 respectively, t[41] = 6.36, p < 0.001). In addition, these conflicting positive and negative sequential dependencies were significantly different (0.98 and −1.3 respectively, t[83] = 7.5, p < 0.001). While seemingly paradoxical, these results are consistent with the results from Studies 1 and 2, with subjects exhibiting negative recency in their probability assessments while exhibiting positive recency in choices.

The previous result is sufficient to provide a demonstration of inconsistent choice and judgment in the context of small probabilities. Nonetheless, we completed a brief analysis of the *objective* sequential dependencies in the bombing data. Figure 5 presents the percentage of days where a suicide bombing occurred according to what happened the previous day for the period of September 30, 2000 to August 31, 2002, the period of al-aqsa intifada (Associated Press, 2002). While an attack was almost twice as likely the day after a previous attack (with or without casualties) than after a normal day, this difference is marginally significant only after combining days after attacks with and without casualties (chi-squared(1)=3.54, p=0.06). This result suggests positive recency in the series of suicide bombings for this period.

Assuming an objective positive sequential-dependency in the data above it is interesting to note that, in the current context, people's reported choice behavior (the decision to be more cautious) was more consistent with the objective sequential dependencies than was their judgments (the belief in negative recency). Still, the more important finding is the concurrent positive and negative recency effects.

6 Discussion

The current research demonstrates the coexistence of overestimation and underweighting of rare events in a within-subject design. The subjects in our studies overestimated low probability events, but chose as if they underweighted these events. The results suggest that judgments and choices reflect two separate processes and that the well known behavioral tendencies that are associated with judgment and choice can coexist. While estimates are sensitive to the larger saliency (and therefore availability) of rare events and are overestimated, choice reflects reliance on small samples and the subsequent underweighting of rare events. Useful descriptive models of both these processes already exist and predict the pattern observed in Studies 1–3. Erev, Wallsten, and Budescu's (1994) model describes the addition of error to estimates, producing overestimation in judgments; while learning models that assume reliance on small samples (for example, Erev & Barron, 2005; Camerer & Ho, 1999; to name just two) predict underweighting of rare events in choice.[11]

The main contribution of the current paper is in demonstrating that these phenomena are observed concurrently. The finding is important because it points out a limitation of the two-stage choice model (Fox & Tversky, 1998) for experience-based decisions that involve rare events. That model, in applying Prospect Theory's probability weighting function to people's estimates, predicts that events associated with small subjective probabilities will be overweighted in choice. In fact, we observe the opposite, namely, that people make choices as if they are underweighting the rare event.

---

[11]The data from Experiments 1–3 can also be described with the probability matching assumption (see Estes, 1950). Under this assumption, the proportion of time an alternative is selected is identical with the proportion of time in which this alternative provides the best outcome. This is identical to the assumption that subjects are choosing the best reply to the most recent set of outcomes or to a randomly selected pair of outcomes (in the current context). Thus, a tendency to rely on small samples (the underlying cognitive mechanism) can give rise to what is often desribed as probability matching. Beyond the current results, Erev & Barron (2005) compared models that quantify the small samples assumption with models of probability matching. They report that the former provide a better fit of demonstrations of underweighting and of other deviations from maximization.

Yet, we do observe an overall consistency between judgment and choice, such that, subjects who judged the rare event to be more probable chose the distribution associated with the event more often if the event was relatively good, and less often if the event was relatively bad. This is consistent with previous reports of simultaneous underweighting in choice and good calibration of estimations. However, note that good calibration does not imply linear weighting. For example, the calibration of subjects whose estimates coincided perfectly with Prospect Theories weighting function would still be r = 0.98. Thus, the current results do not violate the two-stage model's assumption of consistency, but rather, its assumption of Prospect Theory's weighting function that overweighs small probabilities for decisions under experience. It is worthy to note that Prospect Theory's weighting function was both formulated and parametrized using data from a description based decision task where objective probabilities were known and therefore overweighted (Tversky & Kahneman, 1992). In contrast, in experience based tasks such as those in the current studies, where probabilities are not known, underweighting is the typical finding (Barron & Erev, 2003; Weber, Blais, & Shafir, 2004; Hau et al., 2008; Hertwig et al, 2004; Yechiam & Busemeyer, 2006). We conclude that the two-stage model, as currently defined, is of limited use in predicting repeated experience-based decisions involving rare events.

This paper's second contribution is in demonstrating the contingent recency effect of judgment and choice. While probability estimates reflected negative recency, positive recency was observed for choices. The results extend Ayton and Fischer's (2004) work that demonstrated simultaneous negative and positive recency for individual subjects performing a binary prediction task. While subjects' predictions showed negative recency, their beliefs in the sequence of success and failure of their predictions showed positive recency. That paper concluded that sequences of outcomes reflecting human performance yield anticipations of positive recency, whereas outcomes due to inanimate chance mechanisms yield anticipations of negative recency. The current paper supports this interpretation of their results and clarifies them. Whereas beliefs were associated with positive recency in Ayton and Fischer (2004) they were associated with negative recency in the current Studies 1 and 2 since, in our studies, beliefs were being elicited about a chance mechanism and not regarding human performance. Similarly, it was the choice task in Studies 1–3 that required "human performance" and was subsequently associated with positive recency. Finally, Study 4 demonstrated the generality of these findings to a real world context with non-trivial outcomes. As the event of a suicide attack cannot be predicted, probability estimates concerning it reflected negative recency. Alternatively, cautious behavior, arguably

a performance measure in this context, reflected positive recency. This is also consistent with the results of Newell and Rakow (2007) who showed that the underweighting phenomenon in one-shot decisions from experience is facilitated by active sampling of the choice alternatives.

It is interesting to compare the current results to the literature on earthquakes and judgment and decision making. Specifically, Beron, Murdoch, Thayer and Vijverberg (1997) found that after a quake,[12] people were *less* willing to pay for a reduction in the probability of property damage, suggesting that they decreased their estimate of another quake. On the other hand, studies in the US and Japan show that land prices are generally lower for areas with high risk of natural disasters such as earthquakes and floods (e.g., Nakagawa, Saito, & Yamaga, 2007; Carbone, Hallstrom, & Smith, 2006; Bin & Polasky, 2003; see also Beron et al., 1997 although the trend there was not significant), suggesting that potential buyers are more cautious about purchasing in these areas. While these findings appear to reflect negative recency for estimations and positive recency for choice (the choice to buy a house in the same area) they should be evaluated with caution. Most importantly, people have clear priors about their estimates of earthquake risks and their damage, and one of the explanations for Beron et al.'s (1997) finding of decreased risk evaluations following an earthquake is that people's priors were initially too high. A similar finding is that the online availability of the Colorado Springs Fire Department rating of wildfire risk in 35,000 housing parcels has eliminated the association between the presence of fires and home price in the entire county (Donovan, Champ, & Butry, 2007). Apparently, an event that is highly localized also has an information value for those areas that it did not occur in, or which had sustained lower damage from it. Further work is necessary to evaluate the boundaries of the current findings and to extend them to contexts where there are clear priors concerning the relevant risks. While their limitations are not yet clear, the ease with which they are applied to real-world situations, such as terrorist attacks as demonstrated in Study 4, suggests that they may be robust.

References

Ariely, D., Gneezy, U., & Haruvy, E. (2004). Social norms and the price of zero. Unpublished manuscript, MIT.

Associated Press. (2002). 70 Palestinian suicide bombing attacks against Israel in 21 months of violence. Jun 19, 2002.

Ayton, P., & Fischer, I. (2004). The Gambler's Fallacy and the Hot-Hand Fallacy: Two Faces of Subjective Randomness? *Memory and Cognition*, *32*, 1369–1378.

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

Beron, K., Murdoch, J., Thayer, M., & Vijverberg, W. (1997). An analysis of the housing market before and after the 1989 Loma Prieta earthquake. *Land Economics*, *73*, 101–113.

Bin, O., & Polasky, S. (2004). Effects of flood hazards on property values. Evidence before and after Hurricane Floyd. *Land Economics*, *80*, 490–500

Bush, R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.

Camerer, C., & Ho, T. (1999). Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*. *67*, 827–874.

Carbone, J. C., Hallstrom, D. G., & Smith, V. K. (2006). Can natural experiments measure behavioral responses to environmental risks? *Environmental and Resource Economics*, *33*, 273–292.

Donovan, G. H., Champ, P. A., & Butry, D. T. (2007). Wildfire risk and housing prices: a case study from Colorado Springs. *Land Economics*, *83*, 217–233

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

Erev, I., Bornstein, G., & Wallsten, T. S. (1993). The negative effect of probability assessments on decision quality. *Organizational Behavior and Human Decision Processes, 55*, 78–94.

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. & Wallsten, T. S. (1993), The effect of explicit probabilities on the decision weights and the reflection effect. *Journal of Behavioral Decision Making*, *6*, 221–241.

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.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*, 94–107.

Festinger, L., & Carlsmith, J M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203–210.

Fischhoff, B., Parker, A. M., Bruine De Bruin, W., Downs, J., Palmgren, C., Dawes, R., & Manski, C. F. (2000). Teen expectations for significant life events. *Public Opinion Quarterly*, *64*, 189–205.

---

[12]This study focused on the 1989 Loma Prieta earthquake in northern California, the most significant earthquake in the United States since 1906.

Fox, C., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber, & Erev (2004). *Judgment and Decision Making*, *1*, 159–161.

Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science, 44*, 879–895.

Hau, R., Plescak, 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.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4*, 5–28.

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

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.

Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General*, *131*, 287–297.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*, 311–328.

Lee, W. (1971). *Decision Theory Human Behavior*. New York: Wiley.

Lichtenstein, S., Slovic, P., Fischhoff, B., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 551–578.

Lopes, L. L. (1996). When time is of the essence: Averaging, aspiration, and the short run. *Organizational Behavior and Human Decision Processes*, *65*, 179–189.

March, J. G. (1996). Learning to be risk averse. *Psychological Review*, *103*, 309–319.

Nakagawa, M., Saito, M., & Yamaga, H. (2007). Earthquake risks and housing rents: Evidence from the tokyo metropolitan area. *Regional Science and Urban Economics*, *37*, 87–99.

Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin and Review*, *14*, 1133–1139.

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, *86*, 777–803.

Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, *1*, 439–446.

Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making*, *1*, 1–12.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1130.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *9*, 195–230.

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

Viscusi, W. K. (1992). *Smoking: Making the risky decision*. New York: Oxford University Press.

Wallsten, T. S., & Gu, H. B. (2003). Distinguishing choice and subjective probability estimation processes: Implications for theories of judgment and for cross-cultural comparisons. *Organizational Behavior and Human Decision Processes*, *90*, 111–123.

Weber, E. U., Blais, A. R., & Shafir, S. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, *111*, 430–445.

Yates J. F., Lee J. W., Shinotsuka H., Patalano A. L., & Sieck W. R. (1998). Cross-cultural variations in probability judgment accuracy: Beyond general knowledge overconfidence? *Organizational Behavior and Human Decision Processes*, *74*, 89–117.

Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, *19*, 1–16.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier, & B. Tilmann (Eds.), *ETC. Frequency processing and cognition* (pp. 21–36). New York, NY, US: Oxford University Press.

Appendix: Instructions to subjects

In this experiment you are operating a money machine. Upon pressing a button, you will win or lose a number of points. Your goal is to complete the experiment with as many points as possible.

It is given that there is a difference between the buttons.

Upon pressing a button you will receive the following information:

- The number of points you received from the chosen button.

- The number of points you would have received had you chosen the other button.

- Your total earnings.

Sometimes, you will be asked to estimate the chances that a certain outcome will appear in the next round. Your answer must be in percentages. For example, if you estimate that there is a 50–50 (0.5) chance that the outcome will appear then you should enter 50%.

The basic payment is 28 shekels. Your final payment is comprised of the points you earn (1 points = 1 agora) and the basic payment.

For your information, the exact "machine" is likely to differ between subjects.

Good luck.

Experimental screen