# The psychology and rationality of decisions from experience

**Ralph Hertwig**

© Springer Science+Business Media B.V. 2011

**Abstract**    Most investigations into how people make risky choices have employed a simple drosophila: monetary gambles involving stated outcomes and probabilities. People are asked to make *decisions from description*. When people decide whether to back up their computer hard drive, cross a busy street, or go out on a date, however, they do not enjoy the convenience of stated outcomes and probabilities. People make such decisions either in the void of ignorance or in the twilight of their own often limited experience of such real-world options. In the latter case, they make *decisions from experience*. Recent research has consistently documented that decisions from description and decisions from experience can lead to substantially different choices. Key in this description–experience gap is people's treatment of rare events. In this paper, I briefly review studies that have documented the description–experience gap, offer several explanations for this gap, and discuss to what extent people's decisions from experience are in conflict with benchmarks of rationality.

**Keywords**    Choice · Risk · Rationality · Decisions from experience · Rarity

## 1 Introduction

The most famous eruption of Mount Vesuvius occurred in 79 AD, destroying many neighboring towns, among them Pompeii, the luxurious resort of wealthy Romans and now the most renowned still-life of volcanic doom. This eruption, however, was not the most devastating one. As recent volcanological and archaeo-anthropological studies have revealed, the Bronze Age eruption (around 3,780 BC) covered the surrounding area as far as 25 km away, burying land and villages, causing a global climatic

R. Hertwig (✉)
Department of Psychology, University of Basel, Missionsstrasse 60/64, 4055 Basel, Switzerland
e-mail: Ralph.Hertwig@unibas.ch

disturbance and the abandonment of the entire area for centuries. Although the loss of life and property was less extensive in the Bronze Age eruption than in the eruption of AD 79, owing to the sparser settlements, researchers recently discovered evidence of a massive exodus in the form of a huge number of human and animal footprints, pressed into the ash bed and all leading away from the volcano (Mastrolorenzo et al. 2006).

At present, at least three million people live within the area that was destroyed by the Bronze Age eruption. In fact, the periphery of Mount Vesuvius, which includes a significant chunk of the Naples metropolitan area, is among the most populated of any active volcano (Bruni 2003). According to Mastrolorenzo et al.'s (2006) simulations, an eruption comparable in magnitude to the Bronze Age eruption would cause total devastation and mortality within a radius of at least 12 km. In addition, great quantities of fine ash in more distal zones might cause severe respiratory-tract injuries and fatalities due to acute asphyxia. Although it is impossible to predict the exact probability of such a catastrophe happening, volcanologists such as Michael Sheridan have argued that it has been roughly 2,000 years since Pompeii, and "with each year, the statistical probability increases that there will be another violent eruption" (Wilford 2006). In light of these dire forecasts, one might expect that local residents would be keen to move away from the danger zone. On the contrary, relocating residents has proven extremely difficult, despite considerable incentives offered by the regional authorities. "In the shadow of Vesuvius, those residents have cultivated a remarkable optimism, a transcendent fatalism and a form of denial as deep as the earth's molten core" (Bruni 2003).

How can one explain residents' willingness to defy fate? To this end, let us turn to a recent distinction in research on human choice. Choices between risky and uncertain options can be made on the basis of personal experience with the options or on the basis of their symbolic description. This distinction between *decisions from experience* and *decisions from description* matters because these two types of decisions can result in different choices (Hertwig et al. 2004). In order to appreciate this distinction, let us view it in light of the time-honored distinction between *risk* and *uncertainty*.

In the early twentieth century, Knight, one of the founders of the Chicago school of economics, drew a conceptual map that, to this day, has shaped psychologists' and economists' thinking about the world of uncertainty that humans navigate. In *Risk, Uncertainty, and Profit,* Knight (1921, p. 215) distinguished between different types of "probability situations," of which two—risk and uncertainty—have attracted the attention of psychologists and economists (see for example Edwards 1954; Epstein and Wang 1994; Lopes 1983; Luce and Raiffa 1957). According to Luce and Raiffa, we are in the realm of decision making under risk "if each action [of two possible actions] leads to one of a set of possible specific outcomes, each outcome occurring with a known probability. The probabilities are assumed to be known to the decision maker" (p. 13). In contrast, the realm of decision making under uncertainty encompasses situations in which "either action or both has as its consequence a set of possible specific outcomes, but where the probabilities of these outcomes are completely unknown or are not even meaningful" (p. 13). The founder of Bayesian decision theory, Leonard Savage, referred to such situations in which everything—including the probability distributions—is known as "small worlds", and distinguished them from the grand worlds in which a person must take into account many uncertain future possibilities when actually making her choice.

This dichotomy between risk and uncertainty has become deeply ingrained in economic and psychological theorizing about human choice. Knight's (1921) nomenclature, however, was subtler than the simple risk–uncertainty dichotomy (see also Runde 1998). He did not propose a dichotomy, but a trichotomy of probability situations involving (a) *a priori probabilities*, (b) *statistical probabilities*, and (c) *estimates* (pp. 224 and 225). A priori probabilities refer to situations where events are identical, finite, and equally probable, so that the probability of an event occurring can easily be assigned via mathematical calculation, as in deducing the probability of getting a "3" when throwing a fair die. Statistical probabilities refer to situations where events are not necessarily identical, but can be treated as such in an empirical manner. Doing so permits probabilities to be gauged empirically. The European Space Agency (ESA), for instance, may wish to know how successful the European launcher technology Ariane has proved to be. A reasonable method is for ESA to treat this risk actuarially and approximate the probability of an accident from an empirical sample of past events. The sample could, for instance, include all the 116 launches of the Ariane 4 rockets. Out of those launches, 113 were successful and ESA could conclude that the probability of an accident using this launcher technology is .97.[1]

Finally, Knight's (1921) term "estimates" refers to situations of utter uncertainty, in which "there is *no valid basis of any kind* for classifying instances" (Knight 1921, p. 225) or, in other words, where events are truly unique and probability cannot be meaningfully applied or empirically derived. Take, for instance, two defining events of the dramatic economic downturn in 2008. In the autumn of that year, the U.S. government had no reasonable set of *similar* past events via which to assess the consequences and likelihoods of their decision to allow the investment bank Lehman Brothers to collapse. In hindsight, it is widely believed that the collapse was the catalyst triggering the worst global economic slowdown since the 1930s. Past experience was also scant in the second defining event: the U.S. government's announcement of an unprecedented $700 billion economic bailout package to prevent a global economic meltdown—"the most expensive government intervention in history" (Herszenhorn 2008).

Knight (1921) thought the situations involving statistical probabilities to be "extremely common", and he emphasized that—a key point to which I will return throughout this article—"the statistical treatment never gives closely accurate quantitative results" (pp. 215–216). In other words, the probabilities with which we are typically forced to be content invariably fall short of the standards of accuracy set by a priori probabilities. In fact, in the case of a perfect die, Knight thought it "ridiculous to undertake to throw it a few hundred thousand times to ascertain the probability of its resting on one face or another" (Knight 1921, p. 215). Conversely, it would be extremely difficult and perhaps impossible to calculate from a priori principles the chance that a building will be accidentally destroyed by, say, the eruption of a volcano in a given region and at a given time. Rather, this chance must be assessed "if at

---

[1] The statistics were retrieved in October 2009 from the ESA website (http://www.esa.int/esaCP/index.html). Interestingly, the industries producing the Ariane launcher may arrive at quite different estimates of risk by computing the security factor from the design features of the individual parts of the rocket (see Gigerenzer 2002).

all, by tabulating the results of experience" (Knight 1921, p. 215), and our (limited) experience can be fraught with inaccuracies.

How does Knight's trichotomy of probability situations relate to the recent distinction between decisions from description and decisions from experience (Hertwig et al. 2004)? Decisions from experience refer to probability situations in which people cannot help but rely on their experience of previous events and their likelihoods. Decisions from description, in contrast, refer to the choice setting as it has been implemented countless times in the laboratories of psychologists and economists: People are presented with gambles that are characterized in terms of their monetary outcomes and stated a priori probabilities—probabilities akin to those involved in the roll of a die or the spin of a roulette wheel.

Take, for example, Allais' (1953) classic article on "The foundations of a positive theory of choice involving risk". Herein, he described a person who tenders successive choices between four independent prospects, which are described as follows (p. 41):

$$(P_0) \begin{cases} \frac{1}{2} & \$100 \\ \frac{1}{2} & 0 \end{cases} \quad (P_0') \begin{cases} 1 \ \$100 \end{cases} \quad (P_1) \begin{cases} \frac{1}{2} & \$1000 \\ \frac{1}{2} & 0 \end{cases} \quad (P_1') \begin{cases} 1 \ \$300 \end{cases}$$

That is, the person encounters four options, each involving explicitly stated outcomes and a priori probabilities. Outside the laboratory, rare are situations that involve explicitly stated probabilities (there are a few exceptions involving explicitly stated, statistical probabilities such as probabilistic weather forecasts; see Gigerenzer et al. 2005). Most of the time, people can only call upon some sense of events' statistical probability, garnered from their or others' experience. Notwithstanding its ubiquity, this situation has received scant attention in recent research on risky choice (although see Busemeyer 1985[2]; Katz 1962; Myers and Katz 1962; Myers and Sadler 1960). Only recently has this blind spot begun to be eliminated with the advent of a new, simple experimental tool that has enabled psychologists to study how people make choices when probabilities are not explicitly stated.

## 2 A new experimental tool and the description–experience gap

The simplest version of this tool presents respondents with two gambles on a computer screen. A box represents each gamble, which contains a set of outcomes that occur with some probability. People are not told anything about the properties of the gambles, but are allowed to explore each by sampling from them. Specifically, clicking on a box triggers a random draw of an outcome from the associated set of outcomes. People are encouraged to sample until they feel confident enough to decide which box

---

[2] In the 1950s and early 1960s, prior to the advent of modern behavioral decision research, decision scientists investigated decisions from experiences and focused on questions such as whether people manage to learn the probability structure over outcomes (for a review see Luce and Suppes 1965). Hertwig and Erev (2009) speculated that the impracticability of the designs (purportedly hundreds of trials are needed before behavior will stabilize) and the then newly discovered focus on violations of expected utility (with the requirement of conveying perfect information about the probabilities of rare events) may explain why researchers in psychology began to study, nearly exclusively, decisions from description.

is "better", in the sense that they would prefer to draw from it during a final trial involving real monetary payoffs. This paradigm represents a situation in which respondents successively find out which outcomes occur and can—if they choose to—approximate probabilities from their experienced samples.[3] In Knight's (1921) terminology, they would thus extract a sense of the statistical probabilities, and make *decisions from experience*. Note, however, that they can arrive at their choices without forming an explicit representation of their encountered statistical probabilities—an issue to which I will return later.

In studies using this experimental and related experience-based tools, a systematic and robust "description–experience gap" has emerged.[4] The gap describes the observation that respondents who make decisions from experience have different preferences concerning options involving rare events than respondents who make *decisions from description* (involving stated outcomes and probabilities, as in the case of Allais' (1953) gambles). For example, across six problems, Hertwig et al. (2004) found that the average (absolute) difference between the choice proportions of an experience and a description group was 36 percentage points. This gap has been replicated in a number of other studies that used the same or slightly modified experimental procedures (and it also has been observed in the partial-feedback and full-feedback paradigm; see Footnote 1). Table 1 shows the magnitude of the description–experience gap in those studies. In the majority of problems, the same key result emerged (right-most column): Although rare events appear to be receive more weight than they deserve in decisions from description (Kahneman and Tversky 1979; Tversky and Kahneman 1992), they receive less impact than they deserve according to their objective probabilities in decisions from experience.

One immediate consequence of this reversed impact of rare events is that the *fourfold pattern* (the common pattern of risk seeking *and* risk aversion observed in choices between simple, *described* options) is reversed. Let us consider the fourfold pattern in more detail. The classic model of decisions under risk, expected utility theory, assumes that individuals are generally risk-averse. However, Tversky and Kahneman (1992) showed that people are both: risk averse and risk seeking. The fourfold pattern summarizes these conflicting risk attitudes as follows: In decisions from description, people are generally risk averse when the probability of winning is high but risk seeking when

---

[3] Arló-Costa and Helzner (2009) showed how this paradigm can be modified to study decisions from experience in the case of uncertainty.

[4] Within research on decisions from experience, three types of experimental paradigms have been used (see Hertwig and Erev 2009). In the *sampling paradigm*, people merely make a single consequential choice subsequent to having terminated information search. Thus, their decision is one-shot in nature, and they do not face a tradeoff between exploring and exploiting options (see Berry and Fristedt 1985; Daw et al. 2006). If people make repeated consequential choices, they have to strike a balance between exploring alternatives and exploiting them. The *partial-feedback paradigm* investigates this tradeoff, which adds a different dynamic to decision making, and also makes such repeated choices distinct from decisions from description. Specifically, the diminished impact of rare events found in one-shot experience-based decisions, relative to description-based decisions, has also been observed in the partial-feedback design (e.g., Barron and Erev 2003; Erev and Barron 2005). The *full-feedback paradigm* is identical to the previous design, except that is also provides feedback concerning the forgone payoffs (i.e., the payoffs that the person would have received, had she chosen the other option). Choices similar to those in the other two experienced-based paradigms have been observed.

**Table 1** The reality of the description–experience gap (adapted from Hau et al. 2010)

| Authors | Study/condition | No. of problems | Sample size (median) | Description– experience gap[a] (in %) | Correct predictions[b] |
|---|---|---|---|---|---|
| Hertwig et al. 2004 | | 6 | 15 | 36 | 6/6 |
| Hau et al. 2008 | 1 | 6 | 11 | 27 | 6/6 |
| | 2 | 6 | 33 | 13 | 6/6 |
| | 3 | 6 | Set to 100 | 17 | 5/6 |
| Weber et al. 2004 | Experiment 1 vs. experiment 2 | 5 | 17 | 23 | 2/2 |
| Rakow et al. 2008 | Objective description | 12 | 15 | 27 | 11/12 |
| | Percentage sample description[c] | 12 | 15 | 7 | 9/12 |
| | Frequency sample description[c] | 12 | 15 | 7 | 7/12 |
| Ungemach et al. 2009 | Free sampling | 6 | 19 | 31 | 6/6 |
| | Matched sampling[d] | 6 | Set to 80[e] | 15 | 4/6 |
| Gottlieb et al. 2007 | One-by-one vs. percentage[d] | 16 | Set to 20[e] | 11 | 2/4 |

[a] The description–experience gap is measured as the mean absolute difference between choice proportions in the description and experience conditions

[b] This column indicates how many of the differences between the description and experience conditions are consistent with the thesis that rare events have less impact on decisions from experience than on decisions from description (Hertwig et al. 2004); only problems for which a clear prediction can be derived are considered

[c] Description and decisions from experience groups were yoked, thus eliminating sampling error as an explanation for the gap

[d] Samples were chosen such as to reflect the objective probabilities, thus eliminating sampling error

[e] Termination of search was not determined by respondents but by experimenters

it is low. In addition, people are risk averse when the probability of losing is low but risk seeking when it is high. Table 2 exemplifies the fourfold pattern (see Tversky and Fox 1995). Choices consistent with this pattern have been observed in studies investigating decisions from description (e.g., Hershey and Schoemaker 1980; Wehrung 1989). Instead of exclusively explaining this pattern in terms of the shape of the utility function (e.g., Friedman and Savage 1948; Markowitz 1952), some authors have shifted the burden of explanation from the utility function to the assumption of nonlinear transformation of probabilities (Edwards 1962). Nonlinear treatment of probabilities is also one of the main pillars of prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992), currently the most influential descriptive account of risky choice. Prospect theory assumes an inverse S-shaped weighting function, thus postulating overweighting of small probabilities and underweighting of moderate and high probabilities. The theory, for instance, can thus explain simultaneous insurance purchasing and gambling—two behaviors that imply opposite risk attitudes and are hard to explain in terms of expected utility theory—by overweighting of small probabilities (of losses in the case of insurance purchasing and of gains in the case of gambling).

In decisions from experience, the fourfold pattern is reversed, as illustrated in Table 3. In Hertwig et al. (2004), the majority of respondents behaved risk seeking when the probability of winning was high but risk averse when it was low. At the same time, they behaved risk seeking when the probability of losing was low but risk averse

**Table 2** The fourfold pattern in decisions from description

| Probability | Gain | Loss |
|---|---|---|
| Low | $C(100, .05) = 14$ | $C(-100, .05) = -8$ |
| | Risk seeking | Risk aversion |
| High | $C(100, .95) = 78$ | $C(-100, .95) = -84$ |
| | Risk aversion | Risk seeking |

*Note* $C(100, .05)$ represents the median certainty equivalent for the gamble to win \$100 with probability $p = .05$, otherwise nothing (based on Tversky and Fox 1995)

**Table 3** The reversed fourfold pattern in decisions from experience

| Probability | Gain | Loss |
|---|---|---|
| Low | 32, .1 vs. 3, 1.0 | −32, .1 vs. −3, 1.0 |
| | Risk aversion: 20% | Risk seeking: 72% |
| | Rare event: 32, .1 | Rare event: −32, .1 |
| High | 4, .8 vs. 3, 1.0 | −4, .8 vs. −3, 1.0 |
| | Risk seeking: 88% | Risk aversion: 44% |
| | Rare event: 0, .2 | Rare event: 0, .2 |

*Note* Choice proportions refer to the percentage of choices of the risky option in each pair of gambles (based on Hertwig et al. 2004)

when it was high. Assuming for the moment nonlinear transformation of probability, it appears as though the S-shaped weighting function, assumed by prospect theory, is reversed. In other words, in choices involving statistical probabilities, people behave as if they underweight small probabilities and overweight moderate and high probabilities. Reversed nonlinear probability weighting, however, is only one of several possible explanations of the description–experience gap. In what follows, I review and discuss the following non-exclusive explanations for the description–experience gap: (a) reliance on small samples, (b) weighting of probabilities, and (c) format-dependent cognitive heuristics (see also Hertwig and Erev 2009).

## 3 Why do decisions from experience and description diverge?

Humans are the only animals that have the ability to process abstract, symbolic representations of risky prospects. Other animals' decisions (e.g., about where to forage) are by necessity experience-based. Weber et al. (2004) reported some striking similarities in the choices of humans and lower animals when humans are placed in situations in which they, like all other animals, must make decisions from experience. Take, for example, the foraging decisions of bumblebees in the context of different floral reward distributions. Tacitly assuming that bumblebees operate on some representation of experienced (statistical) probabilities, Real (1991) concluded from their foraging behavior that "bumblebees underperceive rare events and overperceive common events" (p. 985), and explained this pattern as a consequence of the fact that "bees frame their decisions on the basis of only a few visits" (Real 1992, p. S133). As

we see shortly, people, like bumblebees, also frame their choice between two payoff distributions on the basis of frugal information foraging. In fact, reliance on small samples has been a robust finding in investigations of decisions from experience. In what follows, I will discuss several possible explanations for the description–experience gap; reliance on the small samples is the first.

## 3.1 Reliance on small samples and sampling error in decisions from experience

How small is small? Hertwig et al. (2004) observed that the typical number of draws that respondents made was approximately 7 from each payoff distribution (henceforth, gamble) and a median of 15 across both distributions. As Table 1 shows, subsequent studies observed similarly small samples, with the exception of Study 2 in Hau et al. (2008)—a study to which I return shortly. Sample sizes in the range between 11 and 19 draws (across both distributions) are fairly small given that each draw takes only a few seconds, and given that small samples have two obvious consequences: First, the smaller the number of draws, the larger the probability that a decision maker will never come across rare events, remaining ignorant of their existence. In fact, across the six problems investigated by Hertwig et al., rare events were not encountered in 44% of all samples. Second, small samples also make it more probable that one will encounter the rare event less frequently than expected given its objective probability. Averaged across problems, Hertwig et al. observed that 78% of respondents encountered the rare event less frequently than expected (i.e., fewer than $np$ times) whereas merely 22% of respondents encountered the rare event as frequently or more frequently than expected.

The reason for this underrepresentation of rare events in small samples is that the binomial distribution for the number of times a particular outcome will be observed in $n$ independent trials is skewed when $p$ is small (i.e., the event is rare) and $n$ is small (i.e., few outcomes are sampled). For such distributions, one is more likely to encounter the rare event less frequently than expected ($np$) than more frequently than expected (for details, see Hertwig et al. 2006). Figure 1 illustrates this systematic distortion in the sample probabilities. It plots the binomial sample probabilities as a function of three sets of sample sizes. As shown in the graph, the sample probabilities underrepresent the rare events and overrepresent common events. The smaller the sample size, the stronger this symmetrical effect.

These observations suggest that it is not necessary to assume that people *weight* probabilities differently in decisions from description and experience, respectively. The grounds for the description–experience gap may already be laid when the mind tries to estimate the statistical probabilities based on small samples. Samples of experience introduce error via the sampling process, and therefore statistical probabilities will not be as accurate as a priori probabilities. Sampling error is, of course, larger the smaller the sample, and the less common the event in question. The different impact of rare events in decisions from experience, relative to decisions from description, might thus stem exclusively from their systematic underrepresentation in small samples, as has been suggested by, for instance, Fox and Hadar (2006) and Rakow et al. (2008), rather than their *underweighting*.
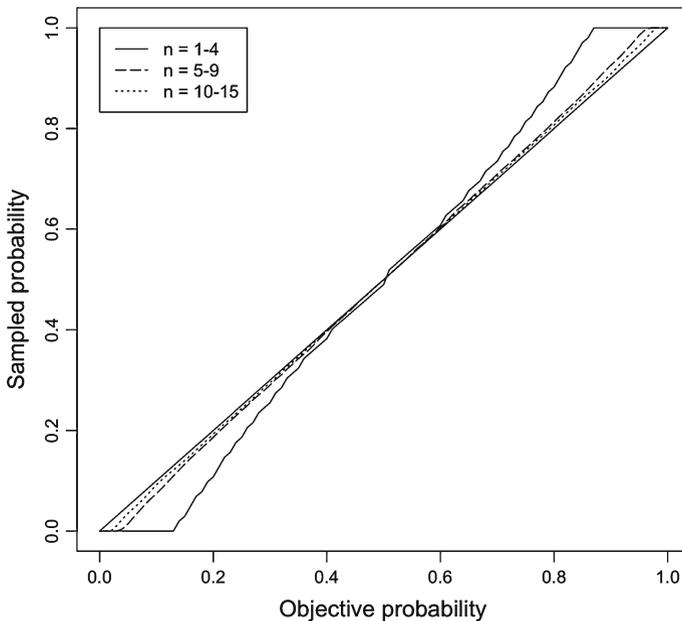
**Fig. 1** Binomial sample probabilities as a function of three sample sizes ($n = 1$–$4$, $n = 5$–$9$, $n = 10$–$15$). The values represent the sample probabilities averaged across the respective sample sizes. For instance, the binomial sample probability for sample size 1–4 and objective probability .2 amounts to .108. This value represents the 25%-trimmed mean, averaging all values from the 25th to the 75th quantile

If, indeed, the only culprit behind the description–experience gap were sampling error, then there would be two straightforward strategies for reducing and even eliminating the gap. One would be to have people sample more, and to see whether the gap gets smaller. The other would be to make decisions from experience and decisions from description strictly equivalent. Both tests of this sampling-error explanation of the description–experience gap have been conducted.

### 3.1.1 Does more experience reduce the gap?

Hau et al. (2008) tested the impact of extensive experience in two studies. In one, they increased the payoffs by an order of magnitude, relative to Hertwig et al. (2004). With more at stake, respondents were eager to sample more: a median of 33 times across both payoff distributions (Table 1, Study 2). More experience decreased the description–experience gap, from 27 (Study 1) to 13 percentage points in Study 2.

Could one reduce the gap even more by having people acquire even larger samples of experience? Hau et al. (2008) further boosted respondents' search efforts by instructing them to sample a total of 100 times from both payoff distributions. However, tripling the sample effort—from 33 draws to 100 draws—did not further shrink the gap (17 percentage points; Table 1, Study 3). This set of results suggests that sampling error contributes substantially to the gap but is not solely responsible for it.

Why does extensive experience not do away with the description–experience gap? One possible explanation is that, paradoxically, people continue to rely on small

samples even if they garner more experience. How so? There is evidence in research on memory and belief-updating (see Hogarth and Einhorn 1992, for a review) that observations encountered early or late in a sequence of information receive more weight than they deserve (i.e., more than $1/n$). Consistent with such order effects, Hertwig et al. (2004) found that observations that people encountered in the second half of the sequence of $n$ observations had more impact on the ultimate choice than draws encountered earlier. This recency effect makes the *functional* sample size of observations from which people derive choices small, regardless of how much information they initially sampled. Despite the elegant relationship between small samples and recency effect, one should not overstate the latter's importance. Unlike Hertwig et al., Hau et al. (2008) and Rakow et al. (2008) found its impact on the description–experience gap to be quite limited.

### 3.1.2 Does strictly equivalent probability information eliminate the gap?

Another test of the impact of sampling error on the description–experience gap is to describe one person's *experienced* sample probabilities to another person (that is, the very probabilities that occurred in the sample). These "twins" are thus yoked, facing exactly the same probabilities and outcomes (see Rakow et al. 2008). The only difference is the format of information: one person experiences probabilities in terms of sequentially sampling events, one by one; the other will learn about probabilities in terms of summary descriptions. With the sampling error thus included into stated (described) probabilities, the twins' choices should no longer differ. Indeed, in the first study that implemented a yoked design the gap disappeared (Rakow et al. 2008). A later study, however, found a different result. In Hau et al.'s (2010) implementation, each participant in the experience group made five choices per problem, having sampled 5, 10, 20, 35, and 50 outcomes, respectively. Twins in the description group made the same five choices. The two key results were as follows: With very little experience (5 draws), choices in both the groups closely tracked each other. With extensive experience (50 draws), however, choices of the yoked description group became increasingly dissimilar to those of the experience group. To put it differently, when sampling error was pronounced (small sample), the description–experience gap disappeared, but when sampling error was small (large sample), the gap reappeared.

One possible explanation for this puzzling effect is a phenomenon analyzed by Hertwig and Pleskac (2008). They showed that when people rely on small samples, they experience a difference between options' average rewards (i.e., the sample means) that can be substantially larger than the difference in the expected values of the described options. Take, for example, the first yoked-design study by Rakow et al. (2008). Based on a median sample of 15 outcomes, the median absolute difference between the sample means in their study was 2, more than 15 times the size of the median difference between the objective expected values of 0.125 (presented in their decisions from description condition; see Hau et al. 2010 for this reanalysis). This amplification effect eases the difficulty of choosing between options (Hertwig and Pleskac 2010), and appears to be one factor that makes people content with relatively small samples in decisions from experience investigations. For small samples, however, the amplification effect can become so large that when one introduces sampling error

and, by extension, amplification into yoked decisions from description, then those choices can become trivial and the description–experience gap disappears (as Rakow et al. observed); if, however, the sampling error and the amplification effect become smaller (as in the case of the large samples studied by Hau et al.) the gap reappears.

After this short excursion into the amplification effect, let us return to the attempt to make the probability information in decisions from experience and description strictly equivalent. Beyond the yoked design, one can achieve such equivalence by fixing people's sampling experience. Ungemach et al. (2009) had people sample 80 draws and, in addition, they devised the experienced probabilities such that they were identical to the stated objective probabilities seen by other respondents who were asked to make decisions from descriptions. A substantial gap occurred.

Let us recap what we know thus far. Unless people have a strong incentive to sample extensively or are instructed to do so, they, like Real's (1991) bumblebees, consistently rely on relatively small samples. Small samples exact the price of sampling error (Fig. 1). Systematic disparities between the experienced sample probabilities and the described objective probabilities account for a substantial portion of the description–experience gap. However, the underrepresentation of rare events in small samples is not the whole story, as Hau et al. (2008, 2010) and Ungemach et al. (2009) demonstrated. What other factors contribute to the gap?

### 3.2 Are experienced probabilities weighted and weighted differently than stated probabilities?

In a classic review, written merely seven years after Von Neumann and Morgenstern (1947) formalized expected utility theory in their *Games and Economic Behavior*, Edwards (1954) discussed empirical findings indicating that expected utility theory was descriptively inadequate. One set of conflicting findings dealt with how people appear to treat probabilities. Preston and Baratta (1948), for instance, found preliminary evidence that people overweight small probabilities and underweight large probabilities. In 1962, Edwards proposed the replacement of probabilities with weights, and in 1979 Kahneman and Tversky proposed a comprehensive model of decision making under risk (later extended to decision making under uncertainty; Tversky and Fox 1995). One major component of their prospect theory is Edwards' notion of decision weights, which "measure the impact of events on the desirability of prospects, and not merely the perceived likelihood of these events" (Kahneman and Tversky 1979, p. 280). The probability-weighting function, $\pi(.)$ describes how different levels of *stated* (described) probability contribute to the evaluation of a gamble's desirability. Tversky and Kahneman (1992) assumed the following one-parameter form of the weighting function:

$$
w(p_j) = \begin{cases} \dfrac{p_j^{\gamma}}{\left(p_j^{\gamma}+(1-p_j)^{\gamma}\right)^{1/\gamma}}, & \text{if } x \geq 0 \\[2ex] \dfrac{p_j^{\delta}}{\left(p_j^{\delta}+(1-p_j)^{\delta}\right)^{1/\delta}}, & \text{if } x < 0. \end{cases}
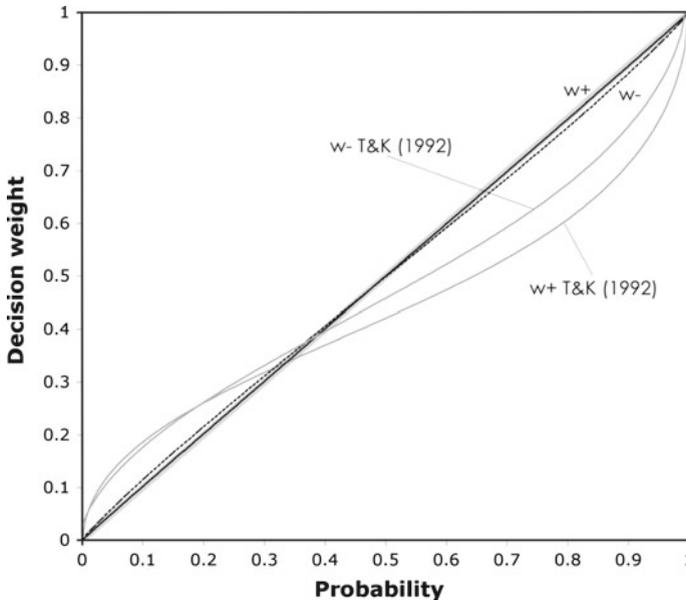$$

**Fig. 2** Weighting functions for decisions from description and decisions from experience. The weighting functions for gains (w+) and losses (w−), respectively, with the parameters estimated from Hertwig et al.'s (2004) data and Tversky and Kahneman (1992), respectively. Note that w+ is visually indistinguishable from the *light grey diagonal* that indicates linear weighting; w− is slightly more S-shaped. From Hau et al. (2008, p. 512). Copyright Wiley-Blackwell, UK. Reproduced with permission of the copyright holder

The $\gamma$ and $\delta$ parameters are adjustable and characterize the shape of the function for gains and losses, respectively. Both parameters are typically assumed to range between 0 and 1, resulting in an inverse S-shape. Tversky and Kahneman fitted this form to cash-equivalent data and derived the median estimates of $\gamma = 0.61$, and $\delta = 0.69$, respectively. The resulting weighting function is plotted in Fig. 2.

Could it be that another contributory factor to the description–experience gap is that the weighting function takes on a different shape in experience-based choices? To the best of my knowledge, there are currently only two published attempts to fit Tversky and Kahneman's one-parameter form of the weighting function to decisions from experience. Hau et al. (2008) varied the adjustable parameters (including those of the value function) in steps of 0.01 between 0 and 1 and calculated the number of correct "predictions" for each value or combination of values. Then, they selected the set of parameters that produced the maximum number of correctly predicted choices, namely, $\gamma = .99, \delta = .93$. Figure 2 shows the resulting weighting function. It no longer has the characteristic S-shape. Instead, the optimal parameters imply nearly linear weighting of probabilities. That is, the objective probability scale and the decision weight scale almost perfectly map onto each other. Using a similar procedure (Ungemach et al. (2009); Fig. 1) found that the best fit to the observed experience-based choices was reached with parameter values, implying underweighting of rare events.

Both analyses suggest that the experienced probabilities—to the extent that they play a role at all in how people make choices—are treated differently than stated probabilities. Hau et al.'s (2008) analysis suggests that experienced probabilities may contribute linearly to the evaluation of a gamble, and therefore, there may be no need to replace experienced probabilities with weights. Ungemach et al.'s (2009) analysis suggests that rare events may—on top of being underrepresented in small samples—also be underweighted in experience-based choices. One final remark is in order: With the impact of prospect theory, the assumption of an inverse S-shaped probability weighting has attained the status of a cut-and-dried fact in many researchers' minds. Yet, there are theories of risky choice that successfully predict people's description-based risky choices without assuming any nonlinear probability weighting function (e.g., Brandstätter et al. 2006).

## 3.3 Are people's estimate of their experienced probabilities systematically biased?

In theory, the description–experience gap could also be the consequence of a systematic estimation error, with people systematically *underestimating* the frequencies of the rare event experienced in the sample. Two sets of results fail to support this possibility. First, across a wide range of studies of frequency and probability assessments, people have consistently been found to overestimate rare events (see Hertwig et al. 2005 and the many references therein). Second, studies that have explicitly asked people to report their estimates of the likelihood of rare events in the samples they experienced found the estimates to be well calibrated or a little too high, relative to the actually experienced frequency (Hau et al. 2008; Ungemach et al. 2009). Thus there is no evidence for a systematic discrepancy between the actually experienced frequencies in the sample and people's estimates thereof that could explain the description–experience gap. The fact that people's estimates are quite accurate is consistent with the notion that encoding of frequency information is an inevitable consequence of attending to events, and is thus "obligatory" (Zacks and Hasher 2002, p. 34). The availability of frequency knowledge, however, does not mean that its use in making a choice is obligatory—a point to which I now turn.

## 3.4 Format-dependent cognitive strategies

The previous discussions on probability estimation and weighting rest on a premise: In the process of sampling outcomes, respondents form a mental representation of the relative frequency (probability) with which events occur. Combined with outcome information, these probabilities then enter the evaluation of the gambles' desirability.[5] Do decisions from experience, however, inevitably give rise to an explicit representa-

---

[5] The assumption that people form an explicit representation of the probability distributions suggests the possibility that in light of insufficient information to identify a unique probability distribution, people's representation of uncertainty could better be modeled in terms of imprecise probabilities (e.g., Walley 1991). To the best of my knowledge, this link between decisions from experience and imprecise probabilities and decision rules operating on them (e.g., Gärdenfors and Sahlin 1982; Levi 1974) has not been elaborated.

tion of probabilities? Like descriptive theories of risky choice that retain the expected utility scaffolding such as prospect theory (Kahneman and Tversky 1979), cumulative prospect theory (Tversky and Kahneman 1992), disappointment theory (Bell 1985; Loomes and Sugden 1986), and decision affect theory (Mellers 2000), expected utility theory postulates that human choice can or should be modeled by assuming that people behave as if they multiplied some function of probability and value, and then maximized. Applied to decisions from experience, these models demand explicit representation of probabilities. There are, however, at least two classes of descriptive theories that assume that the decision maker can and will do without probabilities: associative learning models and outcome heuristics.

### 3.4.1 Associative learning models

In this theory class, human choice is conceptualized as a learning process (see e.g., Busemeyer and Myung 1992; Bush and Mosteller 1955). Learning consists in changing the propensity to select a gamble based on the experienced outcomes. Good experiences boost the propensity of choosing the gamble associated with them, and bad experiences diminish it (e.g., Barron and Erev 2003; Denrell 2007; Erev and Barron 2005; March 1996). Two associative-learning models have been proposed to capture decisions from experience.

The *value-updating model* (Hertwig et al. 2006) assumes that learners update their estimates of the value of the gamble after each new draw from it. Specifically, the model computes the weighted average of the previously estimated value and the value of the most recently experienced outcome. The model consists of two steps:

*Step 1.* Set your initial expectation about the values $A_j(0)$ of gamble $j$ to zero. After the $t$th draw from it, update its value according to the following term:

$$A_j(t) = (1 - \omega_t) A_j(t-1) + (\omega_t)(x_t), \tag{2}$$

where $A_j(t)$ represents the value of the gamble $j$ after $t$ samples are drawn; $A_j(t-1)$ represents the value based on the $t-1$ previously sampled outcomes; and $x_t$ is the value of the most recent outcome. The weight accorded to the newly drawn outcome is $\omega = (1/t)^{\phi}$, where $\phi$ is a recency parameter. If $\phi = 1$, all outcomes are equally weighted; if $\phi < 1$, more recent outcomes receive more weight, implying a recency effect; and if $\phi > 1$, earlier outcomes receive more weight, implying a primacy effect.[6]

*Step 2.* Once sampling is terminated, select the gamble with the highest value.

The model includes two parameters, namely, the number of draws and the recency parameter. The former parameter is determined empirically, and the second is adjustable (i.e., fitted to the data). Importantly, the model does not require representation of probabilities.

---

[6] Hertwig et al. (2006) also assumed that the experienced values are transformed using cumulative prospect theory's value function. However, this assumption does not appear to boost its predictive power (Hau et al. 2008); I drop it here.

Weber et al. (2004) used another associative learning model, the *fractional adjust-ment model* (March 1996) to describe the learning process in decisions from experi-ence. It consists of the following two steps:

*Step 1.* Set the initial propensity $p_{j,0}$ to choose each gamble to 0.5. Increment the propensity of choosing gamble $j$ after sampling trial $t$ according to the obser-vation of the payoff, $x_j$, using the following rule,

$$p_{j,t} = \begin{cases} 1 - \left[ (1-\alpha)^{x_j} \left(1 - p_{j,t-1}\right) \right] & \text{if } x_j \geq 0 \\ p_{j,t-1} \left(1-\alpha\right)^{x_j} & \text{if } x_j < 0. \end{cases}$$

Parameter $\alpha$ is adjustable and controls the degree of learning. It is bound between 0 and 1. If $\alpha = 0$, then there will be no learning.

*Step 2*. Choose gamble $j$ with probability $p(\text{j})$.

This model also assumes number of draws and the recency as parameters. The former is determined empirically and the second fitted to the data. As with the value-updating model, this model does not require an explicit representation of probabilities. Both associative-learning models assume that initial impressions are continuously updated in a way that gives recent events more weight than temporally remote events (the value-updating model also permits primacy). Because rare events have a small(er) probability of having occurred recently, they tend (on average) to have a smaller impact on the decision than their objective likelihood of occurrence would warrant.

### 3.4.2 Cognitive heuristics

A class of model that aims to describe both process and outcome of choice is cogni-tive choice heuristics (see Brandstätter et al. 2006). Heuristics can be separated into two classes: those that use solely outcome information and ignore probabilities (out-come heuristics), and those that use at least rudimentary probability information (dual heuristics). Outcome heuristics such as *maximax* and *minimax* (Luce and Raiffa 1957; Savage 1954/1972) were originally proposed as models for *decisions under ignorance* in which people have no information whatsoever about probabilities. Another cogni-tive heuristic that focuses on outcomes is the *natural-mean heuristic* (Hertwig and Pleskac 2008). It works as follows:

Step 1. Calculate the natural mean of outcomes for both gambles by summing, sep-arately for each gamble, all $n$ experienced outcomes and then dividing by $n$.

Step 2. Choose the gamble with the larger natural mean (i.e., the gamble that had the best average outcome in the sampling phase).

The natural-mean heuristic was originally proposed in the context of $n$-armed bandit problems (Sutton and Barto 1998) as a simple method for estimating the values of actions (e.g., the play of one of a slot machine's levers), and for using the estimates to select between actions: "the true value of an action is the mean reward received when the action is selected. One natural way to estimate this is by averaging the

rewards actually received when the action was selected" (Sutton and Barto 1998, p. 27). The natural-mean heuristic has two interesting properties: (a) It is well tailored to sequentially encountered outcomes. (b) It arrives at the same choice prediction as the expected-value calculus assuming the latter is applied to sampled probabilities and outcomes. Notwithstanding the equivalent outcome, the process, however, is radically different (Hoffman 1960). The heuristic, instead of keeping track of all outcomes, tallying their relative frequencies, multiplying outcomes and probabilities, and summing the values across each gamble, tots up all experienced rewards (or losses) per gamble and then divides this sum by the sample size per gamble to arrive at the "natural mean." One interpretation of the natural-mean heuristic is that in decisions from experience it is a simple and psychologically plausible—in particular, in continuous outcome distributions—instantiation of the expected-value calculus.

In light of these models that do not demand explicit representations of probabilities, let us return to the question of what the possible co-determinants of the description–experience gap are. The two associative-learning models and the natural-mean heuristic are format-dependent. That is, they cannot capture decisions from description, because the input into these models consists of a sequence of outcomes that are being integrated into one summary measure. They have no conceptual parameters to take probability information into account, and, in fact, probabilities are not directly put on view in decisions from experience. In decisions from description, in contrast, probabilities are explicitly displayed in front of decisions makers. As a consequence, differences in description- and experience-based choices could partly arise because different formats of mathematically equivalent information trigger different cognitive strategies (see Gigerenzer and Hoffrage 1995, for a related argument in Bayesian reasoning).

Consistent with this possibility, Hau et al. (2008, Fig. 6) observed in a model competition involving 16 choice models for decisions from experience that the format-dependent natural-mean heuristic emerged as one of the four most competitive models. Among these best models were also two other heuristics, the maximax and the lexicographic heuristics. They are not format-dependent. In theory, they can be applied to both kinds of decisions. In reality however, people appear to recruit them in experience- but not in description-based choices. In another model competition, which involved description-based choices (Brandstätter et al. 2006, Fig. 7), both heuristics performed dismally. Thus, the maximax heuristic, which requires no probability information, and the lexicographic heuristic, which requires merely ordinal probability information, appear to be good candidate strategies in decisions from experience but not in decisions from description. The only model that appears to be simultaneously a good model for both kinds of decisions is cumulative prospect theory (Tversky and Kahneman 1992). However, in Hau et al.'s competition, it only performed well assuming linear probability weighting (Fig. 2). And with linear weighting (and ignoring the value function that did not much affect predictive performance), the natural-mean heuristic is mathematically equivalent to cumulative prospect theory. In terms of the implied cognitive processes, however, they are radically different.

### 3.5 Summary

Humans are unique in their ability to make risky choices based on symbolic descriptions of the alternatives. Convenient descriptions of outcomes and probabilities, however, are rarely available. Most of the time, we need to content ourselves with the chance of exploring the alternatives, thus getting a sense of the alternatives' possible outcomes and their statistical probabilities. In recent research, these different formats of risky information have consistently yielded diverging choices when the alternatives involve rare events. What causes this description–experience gap? Several factors have been investigated. Undoubtedly, sampling error and people's default reliance on small samples account for a substantial portion of the gap but not for all of it. Larger samples do not appear to make the gap disappear. Three non-exclusive factors may prevent extensive experience from closing the gap. First, due to order effects such as recency or primacy, only part of extensive experience may guide our evaluations. Second, tallied relative frequencies and a priori probabilities may be weighted differently. Third, the different information formats underlying decisions from experience and description may trigger different cognitive strategies. It is too early to tell which of these factors (or any other) accounts for that part of the description–experience gap that cannot be explained in terms of sampling error. Once we fully understand the causes of the gap, however, we will also understand the cognitive processes recruited in experience-based choice.

## 4 How rational are decisions from experience?

It was not long after Von Neumann and Morgenstern (1947) formalized expected utility theory that the first investigations were conducted to test whether people's choices conform to the theory's predictions. Evidence promptly emerged that people's choices systematically violate expected utility theory (e.g., Allais 1953; Ellsberg 1961; MacCrimmon 1968; Mosteller and Nogee 1951; Preston and Baratta 1948). In subsequent decades, more demonstrations of violations were entered onto a growing list (Edwards 1968; Kahneman and Tversky 2000). Although the specific violations and their normative implications are still under debate (e.g., Allais 1979), a widespread consensus has emerged among decision scientists in psychology, that human behavior is often inconsistent with the predictions of expected utility theory.[7]

How does the brief history of research on decisions from experience compare to the long saga of violations in research on decisions from description? How rational or boundedly rational (Simon 1982) do people appear when rendering decisions from

---

[7] The extent to which people's choices are inconsistent with expected utility theory and expected value theory, however, is quite unclear. In a reanalysis of 450 choice problems with a large variability in expected values (studied by Mellers et al. 1992), Brandstätter et al. (2006) found that with larger ratios between the expected values of the two gambles per problem, the expected value theory predicted people's choice extremely well. Depending on whether one analyzes selected problems (opportunistic sampling) or a large set of randomly or systematically constructed problems, one may arrive at quite different conclusions. In evaluating the validity of models of risky choice, it is important to keep in mind that it is always possible to construct problems that trigger choices that a given model—be it the expected utility calculus, cumulative prospect theory, or cognitive heuristics—can and cannot explain (see Brandstätter et al. 2008).

experience? To attempt an answer to this question, let us turn to one substantial benchmark of rationality, namely, the assumption that rational agents form, on average, correct beliefs about events in their environment. In investigations on decisions from experience, people fall short of this benchmark. As Table 1 shows, several studies have demonstrated that agents are content with relatively small samples—samples so small that rare events go unnoticed or are underrepresented. To the extent that people form beliefs about the likelihood of events, people relying on small samples cannot help but form inaccurate beliefs about rare events (and by extension, common events). Possibly related to these inaccurate likelihood beliefs, people chose the option with the higher expected value in, for instance, Hertwig et al.'s (2004) study in merely 41% of cases.

There are at least three interpretations of people's apparent failure to form correct beliefs about event probabilities in their environment. One is that they are *naïve intuitive statisticians*. We are intuitive statisticians insofar as our cognitive machinery enables us to extract exquisitely accurate information from the samples drawn from the world (e.g., Zacks and Hasher 2002). However, we are naïve insofar as we appear to take "the information input for granted, failing to correct for selectivity and constraints imposed on the [sampled] input" (Fiedler and Juslin 2006, p. 4). That is, we have little understanding that sampled information can be systematically distorted due to factors such as sampling error. On this view, we are boundedly rational—here bounded by our lack of insight into the limited generalizability of the narrow sets of information that we sample.

According to a second interpretation, people have perfectly rational reasons to terminate sampling early. Such good reasons may consist in them reckoning with opportunity costs when they calculate the costs and benefits of search for each further piece of information, and stopping search as soon as costs outweigh benefits. According to this interpretation, people optimize search with respect to the time, opportunity costs, money, cognitive effort, and other resources being spent. Such optimization under constraints (e.g., Sargent 1993; Stigler 1961) sounds plausible at first glance but can require even more knowledge and computation than unbounded rationality (e.g., Vriend 1996). Yet, one experimental finding is consistent with this view: When Hau et al. (2008) increased the monetary stakes by an order of magnitude (relative to Hertwig et al. 2004), participants responded to the increased incentives and doubled their search efforts. Not surprisingly, their larger samples permitted them to estimate more accurately the events' true probabilities (Hau et al. 2008, Fig. 3).

According to a third interpretation, people—and other animals, for that matter—may be frugal samplers because small samples have adaptive benefits. In bumblebees, for instance, the adaptive value of small samples could be related to the clumpy structure of their natural habitat. Real (1992) explained the argument as follows:

> Short-term optimization may be adaptive when there is a high degree of spatial autocorrelation in the distribution of floral rewards. In most field situations, there is intense local competition among pollinators for floral resources. When "hot" and "cold" spots in fields of flowers are created through pollinator activity, then such activity will generate a high degree of spatial autocorrelation in nectar rewards. If information about individual flowers is pooled, then the spatial

structure of reward distributions will be lost, and foraging over the entire field will be less efficient. In spatially autocorrelated environments ("rugged land-scapes"), averaging obscures the true nature of the environment. (p. S135)

Arguments concerning the evolutionary adaptiveness of small samples have also been made in humans (for brief reviews see e.g., Hertwig and Todd 2003; Juslin et al. 2006). Kareev (1995, 2000), for example, proposed that working memory might have evolved so as to increase the chance for early detection of covariation. He argued that the experienced sample size most conducive to the detection of useful correlations (i.e., value $\geq$ .5) is close to Miller's (1956) estimate of the limited capacity of working memory ($7 \pm 2$). Specifically, he proposed that because of the skew of the sampling distribution of certain correlation measures (e.g., the $\varphi$ coefficient), correlations observed in small samples more often exceed the population correlation than correlations estimated from larger samples. That is, small samples *amplify* the correlations, thus enabling their early detection. This conjecture of the evolutionary advantage of small samples has fueled a controversial debate (see Juslin et al. 2006).

Could there be any advantage to frugal sampling in decisions from experience? As pointed out earlier, Hertwig and Pleskac (2008, 2010) proposed one possible advantage that, like Kareev's (2000), rests on the notion of amplification. Unlike Kareev, however, they argued that amplification proffers a cognitive rather than an evolutionary benefit. Through mathematical analysis and computer simulation, Hertwig and Pleskac showed that small samples amplify the difference between the options' average rewards. That is, drawing small samples from payoff distributions results in experienced differences of sample means that are larger than the objective difference. Take, for instance, the choice between the following two gambles:

$A$:  $26 with probability .76, 0 otherwise,
$B$:  $48 with probability .55, 0 otherwise

Gamble $A$ has an expected value of $19.76, whereas $B$'s value is $26.40, which amounts to a difference of $6.64 in the gambles' expected values. Choosing between them appears not overly difficult—especially if the choice comes in the form of a decision from description. But how difficult is the choice on the basis of a sample of, say, five draws from either gamble? In a sample of size five, the expected difference between the sample means is $10.98, 1.7 times larger than the difference of $6.64 (see Hertwig and Pleskac 2008, for the details of this analysis). In other words, this choice points even more clearly to one alternative. That is, *amplified* absolute differences make the choice between gambles simpler, thus explaining the frugal sampling behavior observed in investigations of decisions from experience—a conjecture for which Hertwig and Pleskac (2010) found empirical evidence.

Of course, as the old adage of economists goes: "There is no such thing as a free lunch". In line with this wisdom, decision makers who are content with small samples do not get to enjoy for free the simultaneous advantages of small search costs, low opportunity costs, and amplified differences. The price comes in terms of somewhat inaccurate representations of the gambles' parameters. Specifically, the amplified absolute difference will sometimes point toward the objectively inferior gamble. But how large is that price? Interestingly, more accurate knowledge derived from larger

samples yields surprisingly modest gains in terms of the probability of selecting the higher-value distribution (due to the diminishing return of the value of information). As Hertwig and Pleskac (2008) showed in an analysis involving an environment of 1,000 pairs of randomly generated two-outcome gambles, choices based on small samples, albeit not optimal, are surprisingly competitive. For instance, with a sample as minute as *one* draw from each gamble, a person has an approximate 60% chance of selecting the higher-expected value gamble. With *five* draws from each gamble, this chance already amounts to 78% (assuming the natural-mean heuristic as the choice strategy). To now reach about the same increase in accuracy—from a 78% to a 95% chance of selecting the higher-expected value gamble—a person would have to draw *one hundred* observations. That is, accuracy continues to increase with each further draw, but at a diminishing rate.

To sum, there is more than one interpretation of the rationality or lack thereof of people's reliance on small samples in decisions from experience. Dependent on one's view, one could see it as the naïveté of boundedly rational intuitive statisticians, as rational behavior of optimizers under constraints, or as the consequence of an amplification effect that renders choice easier at a surprisingly small cost.

Let us conclude with one last remark on the (bounded) rationality of decisions from experience. Our discussion so far has been concerned with the accuracy of people's beliefs about the likelihood of events in the world. It could turn out that their calibration may not be key for how good or bad people's decisions from experience are. The reason is that if people evaluate gambles using simple outcomes heuristics such as the maximax heuristic, then event probabilities will not even enter their calculations. If associative learning strategies describe the processes best, then factors such as recency will limit the amount and accuracy of information regardless of how much information about event frequencies a person sampled. If, however, the natural-mean heuristic proves to be a good model of decisions from experience, then this simple strategy will permit people to act in accordance with expected value theory, thus conforming to two processes—weighting and summing—that are often used to describe the core of rational choice (on which numerous descriptive and normative theories of human behavior rest, including expected value theory, expected utility theory, prospect theory, Benjamin Franklin's moral algebra, and theories of moral sense such as utilitarianism and consequentialism; see Brandstätter et al. 2006). With weighting and summing, the rational decision maker can make a tradeoff between, say, high returns with low probabilities and low returns with high probabilities. Ironically, the natural-mean heuristic achieves this tradeoff without forming an explicit representation of the probabilities of monetary outcomes, and without weighting the outcomes by their probabilities.

## 5 Conclusions

Let us at last return to Mount Vesuvius and to the puzzle that most local residents have a strong impulse not to budge. Why is that? Perhaps it has become clear by now why the distinction between decisions from description and experience may be key in understanding this and other puzzling risk-taking behavior. Their personal experience tells residents living in the vicinity of Mount Vesuvius that violent eruptions are extremely

rare; in fact, in most people's lifetime, they have been non-existent. Unless catastrophes have occurred recently, the relative indifference with which citizens and politicians often consider rare but high-consequence events like bursting levies, catastrophic earthquakes and eruptions of volcanoes may be owed to the experience of their rarity (Weber in press). Just as residents in the vicinity of Mount Vesuvius have ignored incentives to relocate, people living in flood plains who make decisions about insurance based on their personal experience with floods—a rare event—have tended to turn down even federally-subsidized flood insurance (Kunreuther 1984). We will not understand how people respond to rare but high-consequence events unless we understand the psychology and rationality of people's decisions from experience.

# References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica, 21*, 503–546.

Allais, M. (1979). The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school. In M. Allais & O. Hagen (Eds.), *Expected utility hypotheses and the Allais paradox*. Dordrecht: Reidel.

Arló-Costa, H., & Helzner, J. (2009). Iterated random selection as intermediate between risk and uncertainty. In T. Augustin, F. P. A. Coolen, S. Moral, & M. C. M. Troffaes, *Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*.: Retrieved August 25, 2009, from http://repository.cmu.edu/philosophy/17.

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

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research, 33*, 1–27.

Berry, D.A., & Fristedt, B. (1985). *Bandit problem: sequenrial allocation of experiments*. London: Chapman and Hall.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review, 113*, 409–432.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2008). Risky choice with heuristics: Reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck & Willemsen (2008) and Rieger and Wang. *Psychological Review, 115*, 281–290.

Bruni, F. (2003, Aug 26). San Giuseppe Vesuviano Journal: Who's afraid of Vesuvius? (Pompeii is history). *New York Times*. Retrieved August 25, 2009, from http://www.nytimes.com/2003/08/26/world/san-giuseppe-vesuviano-journal-who-s-afraid-of-vesuvius-pompeii-is-history.html.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 538–564.

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General, 121*, 177–194.

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

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

Denrell, J. (2007). Adaptive learning and risk taking. *Psychological Review, 114*, 177–187.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380–417.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review, 69*, 109–135.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

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

Epstein, L. G., & Wang, T. (1994). Intertemporal asset pricing under Knightian uncertainty. *Econometrica, 62*, 283–322.

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

Fiedler, K., & Juslin, P. (2006). Taking the interface between mind and environment seriously. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 3–29). New York: Cambridge University Press.

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

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risks. *Journal of Political Economy, 56*, 279–304.

Gärdenfors, P., & Sahlin, N.-E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese, 53*, 361–386.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon and Schuster.

Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis, 25*, 623–629.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.

Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty information is presented affects decision biases. *Psychological Science, 18*, 240–246.

Hau, R., Pleskac, T. J, & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making, 23*, 48–68.

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

Hershey, J. C., & Schoemaker, P. J. H. (1980). Risk taking and problem context in the domain of losses: An expected utility analysis. *Journal of Risk and Insurance, 47*, 111–132.

Herszenhorn, D. M. (2008, Oct. 7). Bailout plan wins approval; Democrats vow tighter rules. *New York Times*. Retrieved August 25, 2009, from http://www.nytimes.com/2008/10/04/business/economy/04bailout.html.

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

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2006). Decisions from experience: Sampling and updating of information. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 72–91). New York: Cambridge University Press.

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

Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 621–642.

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for rational models of cognition* (pp. 209–235). Oxford, UK: Oxford University Press.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*, 225–237.

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking. Psychological perspectives on reasoning, judgment, and decision making* (pp. 213–231). Chichester, UK: Wiley.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57*, 116–131.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*, 1–55.

Juslin, P., Fiedler, K., & Chater, N. (2006). Less is more in covariation detection—or is it? In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 92–123). New York: Cambridge University Press.

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

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. New York: Cambridge University Press.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*, 263–269.

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

Katz, L. (1962). Monetary incentive and range of payoffs as determiners of risk taking. *Journal of Experimental Psychology, 64*, 541–544.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Boston: Hart, Schaffner & Marx; Houghton Mifflin Company.

Kunreuther, H. (1984). Causes of underinsurance against natural disasters. *Geneva Papers on Risk and Insurance, 9*, 206–220.

Levi, I. (1974). On indeterminate probabilities. *Journal of Philosophy, 71*, 391–418.

Loomes, G., & Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *Review of Economic Studies, 53*, 271–282.

Lopes, L. L. (1983). Some thoughts on the psychological concept of risk. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 137–144.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions*. New York: Dover Publications Inc.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. III, pp. 249–410). New York: Wiley.

MacCrimmon, K. R. (1968). Descriptive and normative implications of the decision-theory postulates. In K. H. Borch & J. Mossin (Eds.), *Risk and uncertainty* (pp. 3–23). New York: St. Martin's Press.

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

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy, 60*, 151–158.

Mastrolorenzo, G., Petrone, P., Pappalardo, L., & Sheridan, M. F. (2006). The Avellino 3780-yr-B.P. catastrophe as a worst-case scenario for a future eruption at Vesuvius. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 4366–4370.

Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin, 126*, 910–924.

Mellers, B. A., Chang, S.-J., Birnbaum, M. H., & Ordonez, L. D. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 347–361.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy, 59*, 371–404.

Myers, J., & Katz, L. (1962). Range of payoffs and feedback in risk taking. *Psychological Reports, 10*, 483–486.

Myers, J. L., & Sadler, E. (1960). Effects of range of payoffs as a variable in risk taking. *Journal of Experimental Psychology: General, 60*, 306–309.

Preston, M. G., & Baratta, P. (1948). An experimental study of the auction value of an uncertain outcome. *American Journal of Psychology, 61*, 183–193.

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes, 106*, 168–179.

Real, L. A. (1991). Animal choice behavior and the evolution of cognitive architecture. *Science, 253*, 980–986.

Real, L. A. (1992). Information processing and the evolutionary ecology of cognitive architecture. *American Naturalist, 140*, S108–S145.

Runde, J. (1998). Clarifying Frank Knight's discussion of the meaning of risk and uncertainty. *Cambridge Journal of Economics, 22*, 539–546.

Sargent, T. J. (1993). *Bounded rationality in macroeconomics*. New York: Oxford University Press.

Savage, L. J. (1954/1972). *The foundations of statistics* (2nd ed.). New York: Dover Publications.

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy, 69*, 213–225.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge; MA: The MIT Press.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102*, 269–283.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

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

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Vriend, N. J. (1996). Rational behavior and economic theory. *Journal of Economic Behavior and Organization, 29*, 263–285.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall.

Weber, E. U. (in press). Doing the right thing willingly: Using the insights of behavioral decision research for better environmental decisions. In E. Shafir (Ed.), *The behavioral foundations of policy*. Princeton: Princeton University Press.

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Behavioral decision theory and environmental policy. *Psychological Review, 111*, 430–445.

Wehrung, D. A. (1989). Risk taking over gains and losses: A study of oil executives. *Annals of Operations Research, 19*, 115–139.

Wilford, J. N. (2006, Mar 7). Long before burying Pompeii, Vesuvius vented its wrath. *New York Times*. Retrieved August 25, 2009, from http://www.nytimes.com/2006/03/07/science/07volc.html.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *etc.: Frequency processing and cognition* (pp. 21–36). Oxford: Oxford University Press.