Chapter 10

The game of life: how small samples render choice simpler

Ralph Hertwig

University of Basel, Basel, Switzerland

Timothy J. Pleskac

Michigan State University, East Lansing, MI, USA

> Life is a gamble at terrible odds. If it was a bet you wouldn't take it.
>
> Tom Stoppard (1967), 'Rosencrantz and Guildenstern are dead' (p. 115)

> Bet your shirt on a horse, your retirement on GE, your premiums on disaster, your tuition on a college ... your energy on a book. Birth is just anteing up; every action thereafter is a bet.
>
> Michael and Ellen Kaplan (2006) 'Chances are ...' (p. 82)

> In the interest of fairness, the CPS [Chicago Public School system] resorted to a lottery. .... Imagine two students, statistically identical, each of whom wants to attend a new, better school. Thanks to how the ball bounces in the hopper, one goes to the new school and the other stays behind.
>
> Steven Levitt & Stephen Dubner, (2005), 'Freakonomics' (p. 158)

Arguably one of the most successful metaphors guiding psychological theorizing is not so much a metaphor of the mind, but of how the world presents itself to us. As Goldstein and Weber (1997) pointed out the *life as a gamble* metaphor that enters psychology via the work of von Neumann and Morgenstern (1947) and Savage (1954) has roots dating back to Jacob Bernoulli's (originally published 1713) *Ars Conjectandi*. Relying strongly on monetary gambles, he pursued probability as 'a general theory of rational decisions under uncertainty'—a theory that would apply to 'choices about witness (or creed) to believe, which venture to invest in, which candidate to elect, what insurance premium to charge, which scientific theory to endorse' (Daston, 1988, pp. 40, 50). Savage proposed that virtually all alternative courses of actions have risky, not perfectly predictable consequences, and that therefore everyday choices could be seen as choices between monetary gambles. By making this analogy, Savage sanctioned the future ubiquitous use of a very simple and convenient experimental tool in economists' and psychologists' laboratories, namely, the investigation of choices between monetary gambles.

Choice between monetary gambles is the topic of this chapter. We will distinguish between two different choice contexts, one in which a person can persue convenient descriptions of gambles' outcomes and probabilities, and one in which a person can sequentially sample outcomes thus being able to form an impression of the gambles' stakes and their probabilities. The former choice context gives rise to *decisions from descriptions*, the latter to *decisions from experience*. We will be concerned with decisions from experience, and offer a simple explanation as to why people appear to be content with small samples of experiences before they render their choice. First, however, we illustrate both choice contexts.

Decisions from description and decisions from experience

If the Wason selection task is, as psychologists' wisdom has it, indeed the most studied 'fruit fly' in cognitive psychology, then choice between monetary gambles must be a close runner up. This particular *Drosophila melanogaster* can be studied in many different strands. One involves representations in which people are provided with convenient descriptions of possible alternatives, the possible consequences associated with each alternative, and their respective likelihoods. Faced with such descriptions, we can afford making what Knight (1921) called *decisions under risk*. Decisions under risk are almost exclusively studied via *decisions from description* (Hertwig *et al.*, 2004), namely, situations in which individuals are fully informed about the possible outcomes of each option and their probabilities, in either a visual or a numerical format.[i]

Take Maurice Allais' (1953/1979) famous article on 'The foundations of a positive theory of choice involving risk' for illustration. Herein he described a person who tenders successive choices between four independent prospects, which are described as follows (p. 41):

$$(P_0)\begin{cases}\frac{1}{2} & \$100\\ \frac{1}{2} & 0\end{cases}\qquad (P'_0)\begin{cases}1 & \$100\end{cases}\qquad (P_1)\begin{cases}\frac{1}{2} & \$1000\\ \frac{1}{2} & 0\end{cases}\qquad (P'_1)\begin{cases}1 & \$300\end{cases}$$

That is, the person encounters four prospects, each one involving explicitly stated outcomes and likelihoods. There are real world analogies to these comprehensively described choices in information sources such as weather forecasts, actuarial tables, and mutual-fund brochures. But all the same, compare such fully described prospects with those that we *typically* encounter once we step outside the confines of the experimental laboratory. An admittedly drastic example is that of a patient who considers participating in clinical trials on which basis the Food and Drug Administration decides whether or not to approve an experimental drug. The development and test of new drugs is a fiscal gamble of colossal proportions. By the time a drug reaches the market the manufacturer is estimated to have spent nearly a billion dollars on its

[i] In a recent meta-analysis of all studies involving decisions between a two-outcome risky prospect and a sure thing (with equal expected value), Weber *et al.* (2004) found that all 226 choice situations called for decisions from description.

development, and nearly nine out of ten experimental drugs that enter the first stage of a three-stage regime of clinical trials are eventually abandoned (Groopman, 2006). For the drug manufacturers, the gamble of developing a drug is just about money, albeit enormous amounts of money, and guesses of the likelihoods of success can be garnered from the past.

For a patient, in contrast, entering a clinical trial can be a life-or-death wager, and one in which the likelihoods remain unknown. For illustration, consider *Torcetrapib*, an experimental cholesterol drug developed by Pfizer, the world's largest drug maker. Hailed as the company's most promising experimental drug and—should it work—potentially the 'largest-selling pharmaceutical in history' (Groopman, 2006), all Torcetrapib trials were halted in December 2006 after it was found that it actually caused an increase in deaths and heart problems. Eighty-two people had died so far in a Stage III trial, versus 51 people in the same trial who had not taken it (Berenson, 2006). Partaking in the experimental trial is a gamble, but one in which, unlike in Allais' choice prospects, patients cannot simply consult a list of outcomes and precisely specified likelihoods. Being unaware of the outcomes' probabilities, their decision was thus—in Knight's (1921) terms—a *decision under uncertainty*. Uncertainty is a defining property of human existence, and unfortunately, more often than not there is no reliable probability and outcome information (see also Knight, 1921; Lopes, 1983).

In the twilight of uncertainty, there are at least two strategies to fill the void of knowledge. One is to turn to previous experience—experience that was gathered by us, or others, when facing similar decisions—to thus simulate and anticipate the future (e.g., Dudai & Carruthers, 2005). This ubiquitous strategy is used, for instance, when we remember previous meetings to anticipate what might happen during tomorrow's big meeting. Or take one of the most uncertain of all businesses: predicting the intentions of one's enemy. The political scientist Alexander George (1959) described in his book *Propaganda Analysis* how small groups of intelligence analysts from the Allies successfully applied this strategy, reliance on past experience, during World War II to uncover German secrets.

Cut off from normal sources of information about events in enemy territory, the analysts' task was to monitor everything that came over the radio waves from Nazi Germany. In domestic broadcasts in 1943, Nazi propaganda boasted that the German military had developed a secret 'super weapon.' Were those claims to be believed? The analysts predicated their analysis on their past experience that 'German propaganda never deliberately misled the German people in questions involving an increase of German power' (George, 1959, p. 143). The likely reason was that propaganda was supposed to boost morale. If the propagandists' claims had been bluffs, their attempts at manipulation would have swiftly become ineffective. So, if they broadcast that Germany had a secret super weapon it meant, quite likely, that indeed they had one. Starting from that premise, the intelligence analysts mined other German public pronouncements for more insight, and as George described thus arrived at a surprisingly accurate sequence of inferences about the super weapon. It turned out to be Nazi's fabled V-1 rocket.

Resorting to past experience to anticipate the future, however, only works if there is a past. Sometimes there is none. Contemplate, for instance, the analysts' work at the beginning of the war when they had no experience with the German propaganda machine.

In other cases, the current problem is truly unique making the past a poor model for the present or the future. In such cases, another strategy that, given its feasibility, enables a person to deal with uncertainty is to garner novel experience before making a decision. Novel experience can be collected through sampling information from the payoff distributions under consideration. As Bernstein (1996) pointed out, many critical decisions would be impossible without sampling. And it is a fact of life's limitations (time, resources) that sampling needs to be limited as well. On basis of samples we can make an educated guess as to which option is preferable (see also Gladwell, 2005). For instance, one sip of wine, perhaps even as little as a sniff or a look, determines whether the contents of the bottle are drinkable. Enjoying a two-week free trial of a daily newspaper tells us whether it is to our political taste. A quick look on the online traffic cams tells us which route to choose for the morning commute. Glancing at the plates of the other guests helps us to determine which dish to choose from a restaurant's menu. Polling friends about which of two new movies they enjoyed renders our night at the movies more pleasurable.

Hertwig *et al.* (2004) referred to decisions based on thus sampled experience as *decisions from experience.* A number of studies have demonstrated that decisions from experience are distinct from decisions from description (e.g., Hau *et al.*, in press; Hertwig *et al.*, 2004; Gottlieb *et al.*, 2007; Weber *et al.*, 2004), but see Fox and Hadar (2006). Hertwig *et al.* argued that the key to their difference is people's sampling behavior—especially their tendency to rely on relatively small samples of information. Next, we describe the context in which people rely on small samples of experience.

Decisions from experience: heeding small samples' call

What do we know about the psychology of decisions from experience? Relatively little. As mentioned earlier, this is because studies of human risky choice have almost exclusively examined decisions from description. This situation, however, has been changing rapidly. Recently, a number of studies have turned to environments in which people are confronted with choices between initially unknown payoff distributions where they must overcome their ignorance by sampling from these distributions (e.g., Barron & Erev, 2003; Busemeyer, 1985; Denrell, 2007; Hertwig *et al.*, 2004; March, 1996; Weber *et al.*, 2004). By sampling they can form an impression of the distributions' respective attractiveness.

In decisions from experience the environment and its structure play a substantial role as opposed to the minimal role they play in decisions from description. As a result, models of decisions from experience must adopt more of an ecological approach (Anderson, 1991; Brunswik, 1943; Simon, 1956). To that end we must distinguish between two ecologies when people sample from both distributions. One is characterized by an inherent trade-off between exploiting and exploring options (see Berry & Fristedt, 1985; Daw *et al.*, 2006; Erev & Barron, 2005; March, 1996), and one in which the agent's only objective initially is exploration for information (Hertwig *et al.*, 2004; Weber *et al.*, 2004). In the former ecology, the sampled outcomes simultaneously provide information and payoffs to the agent. He thus has to strike a balance between exploring alternatives and exploiting them. This trade-off adds a different dynamic to decision making, and makes such choices distinct from decisions from description. Specifically, a risk-neutral person may come to prefer a certain alternative to an uncertain one with identical expected values, thus exhibiting risk aversion (see Denrell, 2007; March, 1996). In addition, properties of the decision maker—such as the tendency to probability match, memory constraints, and loss aversion—are likely to shape such choices (Erev & Barron, 2005).

In this chapter we focus on the second information ecology, in which, much like perusing the Gault-Millau or the Michelin Guide to select a restaurant, the sampled outcomes only provide information. Exploitation of the options—dining at one of the acclaimed two- or three-star gastronomic temples—only comes after the agent has terminated search for information. Therefore, people can afford—disregarding opportunity costs for the moment—to further explore a distribution even if the previously sampled outcomes were anything other than attractive. Sampling is thus like testing the water without footing the bill: the sampled outcomes educate about the possible payoffs but do not yet constitute actual earnings. This sampling dynamic in turn makes decisions from experience different from decisions from description.

For illustration consider a study by Hertwig *et al.* (2004) where participants were presented with several choices between two gambles. Not being told anything about the gambles, they were afforded the opportunity to sample information about them. Each respondent saw two boxes on a computer screen, with each box representing a payoff distribution. Clicking on a given box triggered a random draw (with replacement) of an outcome from its distribution. Respondents could sample outcomes in whatever order and as often as they desired. They were encouraged to sample until they felt confident enough to decide from which of the two boxes ('decks') they would like to make one final draw with monetary consequences. After they had stopped sampling they indicated this preference. On this choice and this choice only, respondents received the monetary payoffs identified by their draw.

Hertwig *et al.* (2004) compared the choices found when a group of participants made decisions from experience for six problems with those of a second group who, although responding to the same problems, made decisions from description. Although respondents in the *experience* and *description* groups faced structurally identical problems their choices were drastically different. For instance, respondents faced Problem 1:

Deck A: 4 with probability 0.8, 0 otherwise,
or
Deck B: 3 with certainty.

In this case, a large majority of respondents (88%) in the experience group selected *A*, whereas only about a third of respondents in the description group did so. Choices in the description group were consistent with prospect theory's postulate that people choose as if small-probability events (here $0 with probability 0.2 in option *A*) receive more weight than they deserve according to their objective probabilities of occurrence (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). In contrast, in decisions from experience rare (small-probability) events appear to have less impact than they deserve according to their objective probabilities.

Hertwig *et al.* (2004) argued that rare events are not duly appreciated in experience-based decisions because decision makers tend to rely on small samples to make their choices. Two recent studies of decisions from experience have documented this tendency. Hertwig *et al.* found that the typical number of draws respondents made—with the outcome of each draw appearing immediately—was approximately 7 from each deck (and 15 across both decks). In a similar study, Weber *et al.* (2004) reported a median size of 17 total draws, collapsed across both decks. Reliance on *genuinely* small samples proved to be complemented by reliance on *functionally* small samples due to recency. In Hertwig *et al.*'s study, the most recently experienced outcomes tended to shape the final choice more than previously sampled ones (for more evidence for this *recency* effect see Barron & Erev, 2003; Busemeyer & Stout, 2002; Erev & Barron, 2005; Hertwig *et al.*, 2006; Yechiam & Busemeyer, 2005).

Such small samples interact with the environment in interesting ways. Events in the typical decisions from experience design (as described above) are binomially distributed, where n is the number of draws from a particular gamble or deck, and p is the probability of the maximum outcome in the gamble. When n is small (i.e., few draws) and p—or by implication, $1-p$—is small (i.e., the event is rare) the binomial distribution is skewed for the number of times this rare outcome will be observed in n independent trials. This, in turn, increases the probability that a frugal searcher will not come across the distribution's rare event, keeping her ignorant of its existence. But even if she encounters the rare event in her small sample, she will tend to observe it less frequently than expected (given its objective probability; see Hertwig *et al.*, 2004, for the detailed argument).

The Hindu fable of the elephant and the six blind men is famous precisely because each one took a tiny sample of the entire animal (e.g., the tusk, the trunk, the ear) and arrived at rather erroneous conclusions (e.g., the elephant was very much like a spear, a snake, or a fan). By analogy, if small samples tend to give rise to inaccurate mental models of events—for instance, by raising the risk of overlooking rare events—why then do people rely on such small samples? Next, we put forth an explanation. It complements but does not rule out other possible reasons such as memory constraints, opportunity costs, and lack of motivation (see also Hau *et al.*, in press). However, rather than looking into people's minds to explain their frugal sampling, we suggest that their behavior is a reflection of how gambles present themselves to the player through the windows of small samples.

Gambling in Smallville: why intuitive sampling may be destined to be frugal[ii]

In many domains, sampling in itself is a form of risk-taking. How representative are the samples from which we draw inferences? When are the samples large enough to render possible a good guess? When does further sampling just becomes a waste of time? In Hertwig *et al.* (2004) and Weber *et al.*'s (2004) studies, respondents could have sampled for as long as they wanted without penalty, and each draw required merely a few seconds. By sampling more they could have markedly enhanced their chances of identifying the distribution with the higher expected value. For example,

ii The following exposition of the argument is an adapted extension of the analysis reported by Hertwig and Pleskac (2007).

in Problem 1 if respondents had calculated the observed sample mean for each deck and deterministically chosen the gamble with the larger mean, the median sample size of seven draws from each deck would have offered a meek chance of 58% to select the higher expected value deck.[iii] Can small samples nevertheless be beneficial on some other dimension? Not denying its cost, reliance on small samples can indeed have benefits too. One is that they make it easier to choose. Moreover, to foreshadow a later result, the risk of selecting the inferior gamble in a representative environment is by no means as pronounced as one might fear, based on the analysis of Problem 1.

To appreciate the argument that small samples render choice simpler let us first turn to a distinction discussed by Griffin and Tversky (1992), namely, that between the strength and the weight of evidence. To illustrate both concepts, they used the following example that also involves a decision from experience:

> [s]uppose we wish to evaluate the evidence for the hypothesis that a coin is biased in favor of heads rather than tails. In this case, the proportion of heads in a sample reflects the strength of evidence for the hypothesis in question, and the size of the sample reflects the credence of these data. The distinction between the strength of evidence and its weight is closely related to the distinction between the size of an effect (e.g., the difference between two means) and its reliability (e.g., the standard error of the difference). (Griffin & Tversky, 1992, p. 412)

Griffin and Tversky also suggested that 'people focus on the strength of the evidence—at least, as they perceive it—and then make some adjustment in response to its weight' (p. 413). Typically, this adjustment is insufficient and therefore 'the strength of the evidence tends to dominate its weight in comparison to an appropriate statistical model' (p. 413).

In choosing between two gambles about which one needs to sample information, we also suggest that people focus primarily (but not exclusively) on the size of the effect or the strength of the evidence.

As we will show the size of the effect is larger with smaller samples, thus rendering choice easier. Before explaining this relationship, let us first propose a simple choice rule that rests on the strength of evidence, measured in terms of the difference between means. The *natural mean heuristic* consists of two steps:

> Step 1. Calculate the sample mean for each deck, SM_A and SM_B by summing all the experienced (sampled) outcomes in the respective decks, and dividing by respective sample sizes n_A and n_B, respectively.[iv]
>
> Step 2. Choose the deck with the strictly larger sample mean; otherwise guess.

This choice strategy accounts well for people's choices between payoff distributions (henceforth *decks*). It predicted a total of 77% of people's choices in Hertwig *et al.* (2004), relative to 67% for cumulative prospect theory (Fox & Hadar, 2006). For a set *n*, the strategy even maximizes expected earnings *given* the knowledge of the person.

[iii] The sample mean equals the decks' expected value as calculated on the basis of the sampled outcomes and likelihoods.

[iv] If the sample sizes for each deck are equal, then the rule only requires computing the sum of the outcomes per deck.

In addition, with increasingly larger *ns*, the rule grows more likely to select the gamble with the highest objective expected value. With smaller samples, in contrast, the heuristic runs the risk of selecting the inferior gamble. But resting one's choice on the sample means renders choice easier with small samples. To appreciate this, consider a user of the natural mean heuristic who draws seven observations from deck *A* that offers \$32 with probability 0.1, \$0 otherwise, and seven from deck *B* that offers \$3 for sure. She may encounter seven times a '0' from deck *A*, and a '3' each time from deck *B*, amounting to a sample mean of 0 and 3, respectively. The *experienced difference* is thus 15 times as large as the *description difference* (i.e., the objective difference between the gambles' expected values), 3 versus 0.2. That is, the experienced difference is *amplified*, relative to the description difference.

Amplification was the rule rather than the exception in Hertwig *et al.*'s (2004) study. In a reanalysis of their data, we found that in 121 (81%) of the total of 150 choices the experienced differences were greater than those based on the gambles' descriptions. In merely 12 cases (8%), they were identical, and in 17 choices (11%) the experienced difference was less than the description difference. On average, the experienced difference was 10.8 times larger than the description difference. This amplification is not a coincidence but, as the following formal proof shows, a necessity. Before we turn to the proof, however, one clarification is in order. We do not conjecture that people *exclusively* attend to the strength of evidence and terminate sampling when the experienced difference is largest, thus completely ignoring the reliability of the difference. Rather, we suggest that if—as Griffin and Tversky (1992) proposed—the strength of the evidence tends to dominate its weight, strength can eclipse weight even more by means of the amplification effect. As a consequence, early termination of search is fostered.

The amplification effect: a simple proof

If we assume two options, *A* and *B*, of which *A*'s expected value is greater than *B*'s, then the *absolute* expected difference between the sample means (or strength of evidence), SM_A and SM_B, will always be as large or larger than the expected or description difference, $EV_A - EV_B$. To arrive at this finding, we take the following steps: Setting $Y = SM_A - SM_B$, the expected value of Y, $E(Y)$, can be calculated as follows:

$$E(Y) = P(Y \geq 0)E(Y \mid Y \geq 0) + P(Y < 0)E(Y \mid Y < 0) = EV_A - EV_B. \quad (1)$$

The expected *absolute* difference of $E(|Y|)$ can be found because using the absolute values is tantamount to moving the area below '0,' representing all 'erroneous' differences (i.e., suggesting *B*'s expected value to exceed *A*'s) in the distribution of differences onto the positive reals. Consequently, the expected absolute difference $E(|Y|)$ can be stated:

$$E(|Y|) = P(Y \geq 0)E(Y \mid Y \geq 0) - P(Y < 0)E(Y \mid Y < 0). \quad (2)$$

Because $E(Y \mid Y < 0)$ is by definition negative (the expected value of *Y* given that *Y* is smaller than 0), the second term in (2) becomes positive. Therefore, $E(|Y|)$ is at least as large as $E(Y)$. Put differently, the experienced difference is, on average, larger than the objective or description difference.

This proof also suggests what factors are moderating the amplification effect. Specifically, anything reducing $P(Y<0)$ will result in $E(|Y|)$ approaching $E(Y)$. Three factors that do this are (a) increasing sample size *n*; (b) increasing the difference between the expected values of the two options; and (c) reducing the pooled variance across the two options. This simple proof, however, leaves several key questions unanswered: What is the magnitude of the amplification effect in a given uncertain environment? How robust is it across different choice rules that people may use? What price does the amplification effect and, by extension, the benefit of easier choice exact? To find answers, we turned to a simulated environment of gambles.

The amplification effect: an analysis based on simulation

We assume that a person chooses between simple two-outcome payoff distributions of the type 'a probability *p* to win amount *x*; a probability $(1 - p)$ to win amount *y*' $(x, p; y)$, using the natural mean heuristic. We analyzed a total of 12 different sample sizes, consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, and 25 draws from either deck. For each of these sizes, we calculated several statistics including the expected absolute value and median absolute value of the experienced difference. For illustration, consider choosing between the following two randomly generated decks (Problem 2):

A: 26 with probability 0.76, 0 otherwise,
or
B: 48 with probability 0.55, 0 otherwise.

Deck *A* has an expected value of 19.76, whereas *B*'s value is 26.4. With a difference of 6.64 points between the gambles' expected values, choosing between them appears fairly easy—especially if the choice comes in the form of a decision from description. But how hard is the choice on the basis of a sample of, say, five draws from either deck? In such a sample, a person could encounter $k = 0, 1, 2, 3, 4$, or 5 occurrences of the maximum gain (26 and 48, respectively), and $n - k$ observations of the minimum gain (i.e., 0 in each gamble). The binomial distribution provides the probability of obtaining each *k* (and $n - k$, respectively). Table 10.1 lists the probability of each *k*, and the resulting sample mean. For instance, the probability of the *A*'s maximum gain (26) occurring twice in a sample of five draws equals 0.08. The sample mean for this particular sample would be 10.4. Note that the likelihood of this mean mirrors the likelihood of $k = 2$ in five draws.

In the second step of the analysis, we determined all differences between all possible sample means a person might experience. Table 10.2 lists for Problem 2 all possible absolute experienced differences (assuming sample size of five), and the likelihood with which they occur (values in parentheses). For example, the probability of experiencing a (absolute) difference of 1.6 equals 0.11. The latter value results from multiplying the probability of observing a sample mean of 20.8 in deck *A*, namely, $p(SM_A = 20.8) = 0.40$, with the probability of observing a sample mean of 19.2 in option *B*,

Table 10.1. The probability of encountering the maximum outcome exactly *k* times out of *five draws* and the resulting empirical means in samples of two payoff distributions: Deck *A* (26 with probability 0.76, 0 otherwise) and deck *B* (48 with probability 0.55, 0 otherwise), respectively. The values are rounded.

	Deck *A*		Deck *B*	
Frequency of maximum outcome *k*	Probability of *k*	Sample Mean ($n = 5$)	Probability of *k*	Sample ($n = 5$)
0	0.00	0.00	0.02	0.00
1	0.01	5.20	0.11	9.60
2	0.08	10.40	0.28	19.20
3	0.25	15.60	0.34	28.80
4	0.40	20.80	0.21	38.40
5	0.25	26.00	0.05	48.00

namely, $p(SM_B = 19.2) = 0.28$. By collapsing across the probabilities of observing each absolute difference, one can determine the expected absolute value of the experienced difference between two decks for a given sample size. In a sample of size 5, for instance, this value is 10.98, 1.7 times larger than the description difference of 6.64. The median absolute value of the experienced difference, 7.92, also exceeds the description difference. Indeed, for 70% of the possible observations in our example the experienced difference is greater than the description difference. Note that for the purpose of analyzing the magnitude of the amplification effect we focus on the absolute difference and ignore the direction of the amplification. Respondents who sample outcomes from the two decks of cards, of course, experience a difference that points toward one or the other. Moreover, we will shortly address how often the difference points toward the objectively better gamble.

To examine the degree to which this amplification effect generalizes to other gambles it is necessary to establish an environment of gambles. To that end 1,000 pairs of

Table 10.2. The absolute values of all possible sample differences (assuming sample size of five), and the likelihood with which they occur (values in parentheses) for Problem 2.

SM_A/SM_B	0	9.6	19.2	28.8	38.4	48
0	0 (0.00)	9.6 (0.00)	19.2 (0.00)	28.8 (0.00)	38.4 (0.00)	48 (0.00)
5.2	5.2 (0.00)	4.4 (0.00)	14 (0.00)	23.6 (0.00)	33.2 (0.00)	42.8 (0.00)
10.4	10.4 (0.00)	0.8 (0.01)	8.8 (0.02)	18.4 (0.03)	28 (0.02)	37.6 (0.00)
15.6	15.6 (0.00)	6 (0.03)	3.6 (0.07)	13.2 (0.09)	22.8 (0.05)	32.4 (0.01)
20.8	20.8 (0.01)	11.2 (0.05)	1.6 (0.11)	8 (0.13)	17.6 (0.08)	27.2 (0.02)
26	26 (0.00)	16.4 (0.03)	6.8 (0.07)	2.8 (0.09)	12.4 (0.05)	22 (0.01)

randomly generated two-outcome gambles were created (e.g., Problem 2). For each gamble the probability values were randomly sampled from a uniform distribution between 0 and 1. One payoff value was set to 0 whereas the other payoff value was drawn from a uniform distribution between 0 and 100. None of the gamble pairs were allowed to have stochastically dominating options.

How robust is the amplification effect?

As a function of increasing sample sizes, Fig. 10.1 plots the expected absolute values of the experienced differences for the natural mean heuristic, across all gambles. The straight line represents the average description difference (15.2). Assuming the natural mean heuristic, small samples substantially *amplify* the difference between gambles. For instance, with two draws from each deck, the average expected experienced difference is 23.1, 1.5 times larger than the description difference. Not surprisingly, as sample sizes increase the expected experienced differences converge to the description difference. With 10 draws the expected experienced difference (17.2) is merely 1.1 times larger than the description difference, and with 50 draws, they are nearly the same. Figure 10.2 shows the same finding for the median experienced difference, except that the amplification effect is smaller relative to Fig.10.1.

The basis for the amplification effect, as the formal proof above illustrates, is the positive skew of the distribution of possible experienced differences. The top panel in

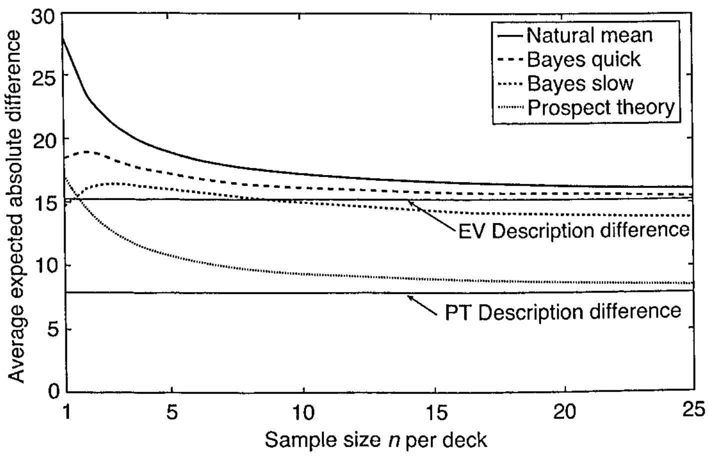

Fig. 10.1. Experienced differences across 1,000 pairs of gambles as a function of sample size (per deck). The four curves represent the mean of the expected absolute difference for the natural mean heuristic, two Bayesian updating strategies (slow versus quick updating), and cumulative prospect theory, respectively. The upper straight horizontal line represents the average description difference based on expected value (15.2). The lower straight horizontal line represents the average description difference after values and probabilities have been transformed according to prospect theory's value and weighting function (7.8).

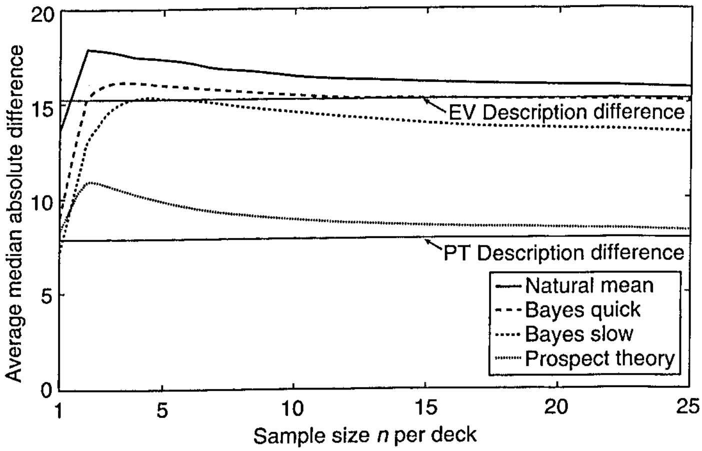

Fig. 10.2. Experienced differences across 1,000 pairs of gambles as a function of sample size (per deck). The four curves represent the median of the experienced absolute difference for the natural mean heuristic, two Bayesian updating strategies (slow versus quick updating), and cumulative prospect theory, respectively. The upper straight horizontal line represents the average description difference based on expected value (15.2). The lower straight horizontal line represents the average description difference after values and probabilities have been transformed according to prospect theory's value and weighting function (7.8).

Fig. 10.3 illustrates this skew for Problem 2 (*A*: 26 with probability 0.76, 0 otherwise, or *B*: 48 with probability 0.55, 0 otherwise) when $n = 5$ samples are drawn from *A* and *B*, respectively. As the figure shows, the expected absolute value of the experienced difference is markedly larger than the description difference. Moreover, the range of possible differences is bounded at zero (i.e., the smallest experienced difference is zero), whereas the largest experienced difference is over seven times larger than the description difference.

The primary reason for the change in the amplification effect as sample size increases, as our proof suggested, is that as sample size increases a larger proportion of the distribution's mass gathers around the description difference, decreasing the distribution's positive skew. Figure 10.3 shows this reduction in the skew for Problem 2. Comparing the bottom panel to the top panel shows that the skew systematically decreases as sample size increases. The other two factors that determine the amplification effect, namely the magnitude of the difference between the expected values of the two gambles and the pooled variance across the two options, can help explain how changes in our environment of gambles would influence the amplification effect. For example, what would happen if the gambles in each problem of our environment had two non-zero outcomes rather than having one zero outcome and one non-zero outcome? If both outcomes are equally likely to fall between 0 and 100 then each of the gambles on average would have less variance in the experienced differences. This is tantamount to decreasing the skew in the distribution of experienced differences.

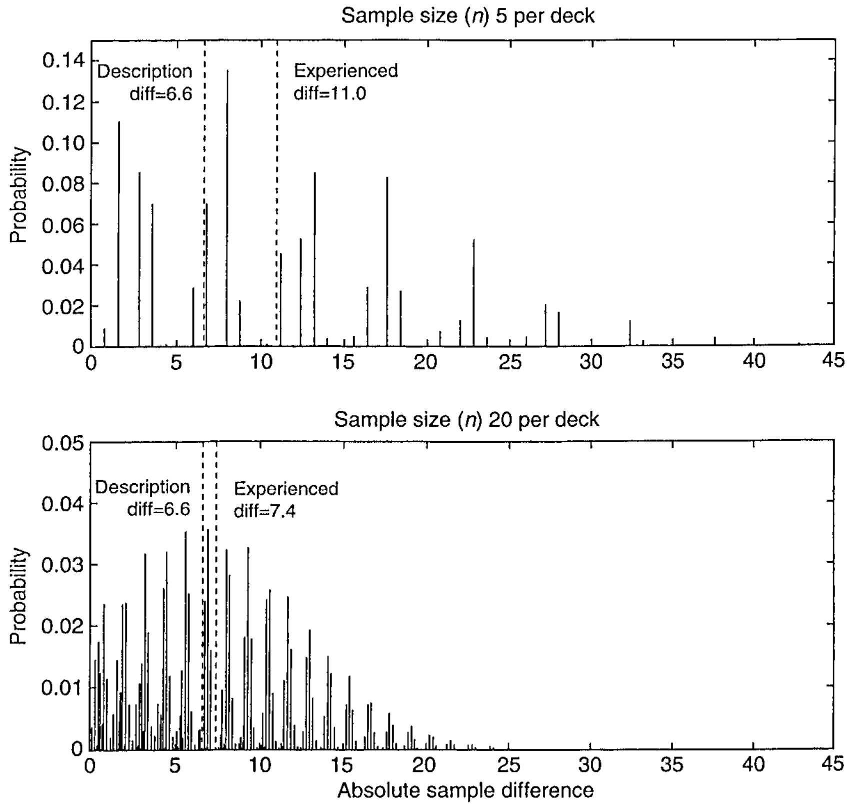

Fig. 10.3. The distribution of all possible experienced differences for Problem 2 (*A*: 26 with probability 0.76, 0 otherwise, or *B*: 48 with probability 0.55, 0 otherwise) when $n = 5$ samples are drawn from *A* and *B*, respectively (see upper panel). The distribution has a positive skew causing the expected and median (absolute) value of the experienced difference to be markedly larger than the description difference. Over 70% of the possible experienced differences are larger than the description difference. The bottom panel is the distribution for Problem 2 for $n = 20$ samples drawn from each deck. The larger sample size decreases the skew of the distribution.

As a result there would be a smaller amplification effect. Our own analysis shows this decreases the amplification effect from, for instance, 1.3 to 1.1 times the description difference at $n = 5$. A similar result holds for mixed gambles with the same range of possible outcomes.

How does more processing affect the amplification effect?

The natural mean heuristic bets on the observed sample and nothing else. Alternatively, a person may aim to combine prior beliefs about the environment in which she finds herself with the newly sampled evidence. Or, alternatively, rather than taking the sample at face value the sample's parameters (i.e., outcomes and probabilities) may be interpreted through prospect theory's value and weighting function. In what follows, we examine both of these computationally taxing alternatives to the natural mean heuristic. We first turn to the possibility that a person combines new data and existing beliefs through use of Bayes' theorem.

Bayesian probability updating: will the amplification effect persist?

To address this question we assumed that searchers started with the initial assumption that in each deck there are two possible outcomes, a minimum and maximum. Believing that there were two outcomes but having no *a priori* reason to believe that one outcome was more likely than another, people could use Laplace's principle of indifference and treat the likelihoods for either payoff in both decks as equally likely. To model this we worked with the probability of the maximum outcome for deck A and B, $\hat{p}_A$ and $\hat{p}_B$.

Specifically, we modeled the person's belief in $\hat{p}_A$ with a beta distribution—often used in Bayesian statistics because it is conjugate to the binomial distribution (see Gelman *et al.*, 2003)—that had a uniform distribution on the interval [0, 1]. Beta distributions are described with two parameters α and β and setting them equal to 1 reproduces this assumption for each deck. We used the mean of the beta distributions as the person's best estimate of $\hat{p}_A$ and $\hat{p}_B$. For deck A, for instance, the estimate is

$$E(\hat{p}_A) = \frac{\alpha}{\alpha + \beta} = 0.5, \tag{3}$$

and the chances of the minimum outcomes are then $1 - E(\hat{p}_A)\ 0.5$ and $1 - E(\hat{p}_B) = 0.5$, respectively. Note that this is only an approximation of a Bayesian learner. In decisions from experience decision makers are also ignorant of the payoffs. To bypass this problem we stipulated that until they experienced the payoffs they assumed them to be zero. More precise models would also account for the decision maker's uncertainty in the payoffs.

Under this model, two things occur when people draw a sample of size n from each deck and observe k_A draws of the maximum outcome from deck A and k_B from deck B, respectively. First, they learn about the value of the maximum outcome. Second, they use their sampling experience to update their prior opinions according to Bayes' theorem. Because the beta distribution is conjugate to the binomial distribution, the updated beliefs for both decks can be expressed as a function of the prior distributions and the experienced sample. In light of her sampled experience, an individual's updated estimate $\hat{p}_A$ about deck A is

$$E(\hat{p}_A \mid k_A, n) = \frac{\alpha + k_A}{\alpha + \beta + n}. \tag{4}$$

The same procedure takes place for deck B using the number of draws of the maximum outcome from it, k_B. The chances of the minimum outcomes are then $1 - E(\hat{p}_A \mid k_A, n)$ and $1 - E(\hat{p}_B \mid k_B, n)$. Using their updated estimates people then determine the expected value for both decks, SEV_A and SEV_B, and choose the deck with the (strictly) higher expected value (otherwise they guess).

Figures 10.1 and 10.2 show the experienced differences for two Bayesian updaters, a slow and a quick updater. The quick updater represents a Bayesian with $\alpha = \beta = 1$. With such low values the person would be very sensitive to the observed data and quickly update her beliefs in light of new evidence. By and large the experienced difference quickly tracks the description difference (see the horizontal line), the reason being that the Laplacian assumption counteracts the possible extreme observations seen in small samples by initially anchoring the decision maker at 50/50. We can also assess how different levels of confidence in one's prior beliefs might impact the amplification effect. By setting $\alpha = \beta = 10$, the variance of the beta distribution decreases and becomes more peaked around the values of $p_A = p_B = 0.5$. In other words, the individual is now more confident in her beliefs, and changes her beliefs at a slower rate than the less confident Bayesian ($\alpha = \beta = 1$). Consequently, the slow updaters lean more heavily toward their prior beliefs. As the slow updater lines in Figs. 10.1 and 10.2 show, the prior beliefs of the slow updater remove the amplification effect and in some cases reverse it, making it harder for him to distinguish between the two decks. It takes him over 100 observations to counteract his dulled discrimination.

Weighting of outcomes and probabilities: will the amplification effect persist?

Cumulative prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) is a Neo-Bernoullian theory. It retains the original expected utility scaffolding, that is, the assumption that human choice can or should be modeled in the same terms that Daniel Bernoulli (1738/1954) postulated: that people behave as if they multiplied some function of probability and utility, and then maximized. Resting on this core assumption, cumulative prospect theory assumes that the impact of probabilities on choice is nonlinear such that small probabilities are overweighted and moderate and large probabilities are underweighted. It also posits that the value (or utility) function is S-shaped, and that the function is steeper for losses than for gains, thereby yielding 'loss aversion' (i.e., the fact that a loss of size k has greater impact than does a gain of size k). Cumulative prospect theory was developed to model decisions from description, that is, risky choices involving stated probabilities. By entering the sample probabilities and outcomes that Hertwig *et al.*'s (2004) respondents encountered into cumulative prospect theory's weighting and value function, Fox and Hadar (2006) showed that this two-stage model (Tversky & Fox, 1995) could also be applied to decisions from experience. In doing so, one can predict 67% of people's choices (compared with 77% by assuming the natural mean heuristic; see Hau *et al.*, in press).

Does the amplification effect continue to exist even if the sampled experience is fed into cumulative prospect theory's value and decision weighting functions? To investigate this issue, we turned to all 1,000 pairs of gambles investigated earlier. As a function of increasing sample size, we used the observed relative frequencies of minimum and maximum outcomes, respectively, as a proxy for people's subjective beliefs. In addition, we calculated prospect theory's value for each gamble (and given sample size), assuming Tversky and Kahneman's (1992) median estimates for the three parameters describing the value function and the two parameters describing the decision weighting function. Across all pairs of gambles, we then calculated the expected experienced difference and the median experienced difference, respectively. Because the value and weighting function are transformations on the original values the amplification effect remains and in fact approximates the one seen for the natural mean heuristic (Figs. 10.1 and 10.2). However, all the values are shifted down for both the experienced differences and the description differences. This is a result of the value and weighting functions mitigating the impact of extreme payoffs and probabilities.

Our concern in this section was whether the amplification effect continues to exist even if people process the sampled information in a cognitively more taxing way. We found two results. First, if respondents update their beliefs starting with a Laplacian assumption, then the amplification effect will disappear. Consequently, Bayesian updaters will forego the 'benefit' of having more dissimilar alternatives. Second, if respondents behave as though they weighted outcomes and probabilities according to prospect theory, then the amplification will continue to exist.

Fine: small-sample choices are easier, but how competitive are they?

The old adage of economists goes, 'There's no such thing as a free lunch.' In line with this wisdom, searchers do not get to enjoy for free the simultaneous advantages of small search costs, small opportunity costs, and amplified differences. The price comes in terms of unreliable representations of the gambles' parameters, and less than optimal choices. As Fig. 10.1 shows, the larger the sample, the more veridically the experienced difference reflects the objective difference between the gambles. Therefore, searchers who sample extensively—say, 25 draws per deck—could thus maximize their earnings by being able to consistently select the gamble with the larger objective payoff. To find out how good or, perhaps, how bad choices based on small samples are, we calculated, using the aforementioned framework, per sample size and across all 1,000 pairs of gambles the proportion of times a respondent would choose the deck with the objectively larger expected value.

Figure 10.4 depicts the results of this analysis. Several striking results emerge. First, choices derived from small samples are clearly not *optimal*, but they are also far from being disastrous. With a sample as tiny as *one* draw from each deck a person has an approximate chance of 60% of selecting the higher-expected value gamble. Second, the information value garnered from each draw is subject to a diminishing return. Using the natural mean heuristic a person can, for instance, reap an 18-percentage point performance increase from 60% to 78% by increasing the sample from 1 to 5. By doubling the sample size from 5 to 10 draws, one achieves another 6-percentage point performance boost to 84%. By again doubling the sample size from 10 to 20 draws, accuracy increases by a mere 4 percentage points. If a person drew 100 rather than 20 observations from each deck, the expected accuracy would amount to

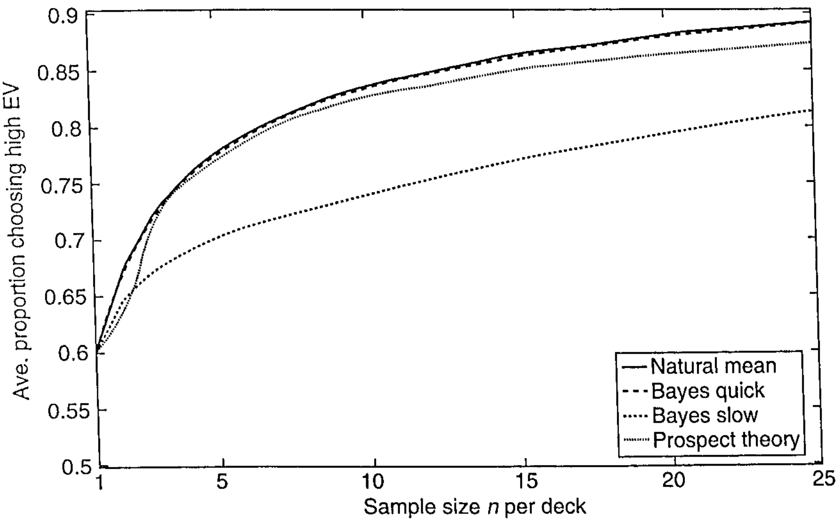

Fig. 10.4. Choice proportions across 1,000 pairs of gambles of the deck with the higher expected value as a function of sample size (per deck) for the four choice rules analyzed here. In the long run, the natural mean heuristic and the Bayesian strategies will choose the better gamble 100% of the time, relative to prospect theory's long run accuracy of 92%.

95% (compared to 81% at $n = 7$). Although accuracy continues to increase with each further draw, it increases at a diminishing rate.

Third, as Fig. 10.4 also shows, with small samples a person who does without complex calculations and bets on the sample means to choose between the decks either exceeds (slow updating) or equals (quick updating or cumulative prospect theory) the level of accuracy reached by a Bayesian updater. After a large number of draws from each deck the Bayesian and the natural mean strategies converge. This occurs for the two Bayesian strategies and the natural mean heuristic at around 100 draws from each deck where all three have an expected accuracy of approximately 95%. In the limit these strategies would choose the higher expected value gamble 100% of the time. Interestingly, people who perceive values through the lenses of prospect theory and draw larger and larger samples do not converge to this accuracy. Instead, they make systematic mistakes in the gambling environment and converge at 92% accuracy in the long run, the proportion of times the higher expected value gamble is chosen in decisions from description.

To conclude, betting on small samples exacts costs. At worst, small samples can give rise to mental representations in which the higher-value gamble appears inferior, and the lower-value gamble superior. In our environment of randomly selected gambles, however, small samples—even as small as one draw per deck—resulted in markedly more accurate than inaccurate choices. Drawing as few as seven times from each deck secured an 86% chance to select the gamble with the higher expected value. To avoid misunderstandings: We do not conjecture that reliance on small samples is the optimal strategy—and in fact, it is not clear to us how to compute the optimal solution for decisions from experience. Yet, as our analysis has shown, relatively small samples afford a surprisingly high probability of choosing the gamble with the higher expected value.

If life is a gamble, then it is not one but many

At any given time, our ancestors played a myriad of gambles. They had to wager on what they could eat, what they could capture, what could capture them, and whom they could trust, to name just a few (see Buss, 2005, for many others gambles). In light of the many dicey choices that our ancestors faced, and, by the same token, we face every day, a key issue emerges: How much information should we forage in each of the choices? And, when does the cost of more choice outweigh the benefits? There is no simple answer because any answer depends on multiple factors such as the costs of wrong decisions, and the cost of search. Without the pretension of having an answer to these questions, our analytical frame enables us to take a first, simple step toward an answer.

Let us assume the following simplified world. People encounter in the course of their lives a total of 100 pairs of payoff distributions, each pair representing one of life's gambles. Before making a decision, they can sample from the distributions (about which they initially are ignorant) as many times as they wish before deciding which one they prefer to play once for real money. In realistic models, search must be limited because real people have only a finite amount of time, knowledge, attention, or money to spend on a particular decision. In our microworld, limited search is represented in terms of a finite budget of draws that each person can afford to sample across all gambles. How should people employ the budget across decisions?

There is no simple answer. Because the parameter of the payoff distributions can be estimated better the larger the number of draws, people stand to increase their expected payoffs *per decision* with more observations. However, the increase in expected payoffs diminishes with larger sample sizes. In addition, the strategy of maximizing accuracy per decision may ultimately decrease average earnings across many decisions, simply because once the budget of draws is used up one cannot help but to choose randomly.

To investigate this trade-off between maximizing accuracy per decision and across decisions, we investigated the expected earnings of players who used the natural mean heuristic in 100 pairs of gambles in the gain domain with a zero-outcome as a second payoff in both gambles. In addition, we let the players set different aspiration levels for identifying the gamble with the highest expected value ranging from 50% to 95% accuracy. The aspiration levels determined the sample size agents took based on the analysis presented in Fig. 10.4. So, for instance, recall that with $n \approx 100$ participants can expect 95% accuracy and setting $n = 0$ they can expect 50% accuracy. We also assumed that once the budget of draws is exhausted players resort to choosing randomly between the options. This budget was varied and ranged from $R = 100$, 200, ..., to 1,000 draws per decks A and B, respectively.

Table 10.3 summarizes the average earnings as a function of the aspiration level (and corresponding sample size), and the available number of draws per deck.[v] To interpret the results let us first point out that if respondents chose to learn nothing about the gambles and guessed they would, on average, earn 2,390 points. If instead, they had access to the gamble descriptions and chose according to expected value, they would, on average, earn 3,164 points. Three results deserve to be highlighted.

First, the analysis gives new meaning to the phrase, 'Jack of all trades, master of none, though ofttimes better than master of one.' That is in our analysis agents do best—regardless of how tight or generous their resources are—by distributing the draws equally among the gambles and being content with limited knowledge rather than allocating resources to just a few selected gambles. This conclusion is supported by the values in the diagonal: They represent the highest average earnings within a given budget, and they coincide with a strictly equal distribution of draws across decks. Take, for instance, the value 2,929. It corresponds to making three draws per deck for each of the 100 *A* and *B* decks, respectively. Any other strategy that aims to invest more resources in some gambles (thus striving for a high level of accuracy) and, as a consequence, fewer in others is outperformed by an equal-sampling policy (compare 2,929 to all other values in the '300' row). Second, the results again demonstrate that the increase in information value across sample size is subject to diminishing return. Whereas the average earnings increase by 6 percentage points when increasing the budget from 100 to 200 draws for decks *A* and *B*, respectively, this increase shrinks to 0.3 percentage points when increasing the budget from 900 to 1,000 draws. Third, the average earnings, assuming seven draws per deck (and a corresponding accuracy of identifying the higher-expected value option of about 80%), amount to 96% of the maximum earning, 3,039 versus 3,164 points, respectively.

Our implementation of the life-as-a-gamble metaphor suggests that relying on a small sample proves to be a competitive strategy in the long run. One reason for this result in our microworld of gambles, and, perhaps, by extension in life's gambles, is that players are faced with the problem of flat maximum (Winterfeldt & Edwards, 1986). The problems in which decision makers stand to gain the most via more observations are also the ones in which sampling produces the least marginal gain. The reason is that these problems involve gambles that are close in their expected values. In some cases, two gambles might be so close together that one might just as well flip a coin to decide. In contrast, the natural mean heuristic only needs a small

v The expected earnings for each problem can be calculated in the following manner. For each problem we can calculate the probability of choosing the higher expected value gamble, $C(n)$, for a player who uses the natural mean after drawing a sample size n from each gamble. With some level of resources, R, the player can sample from R/n of the 100 gambles, where $n \leq R/100$. In the long run this means that for each problem there is a probability of $r = R/(100n)$ of using natural mean with sample size n to make a choice, and a probability of $(1-r)$ to guess. The expected earnings for all 100 gambles is then $[C(n)r + 0.5(1-r)]EV_A + [(1-C(n))r + 0.5(1-r)]EV_B$, where EV_A and EV_B are the expected values of the two gambles. Gamble A is assumed to have a higher expected value then B.

Table 10.3. The average earnings in our life-is-a-gamble simulation, assuming budget of draws ranging between 100 and 1,000 draws for decks *A* and *B*, respectively, and confidence level ranging between 0.5 and 0.95 (corresponding to sample sizes of 0–134 per deck). The confidence level denotes the chances of selecting the gamble with the higher-expected value. The values in the diagonal represent the highest average earning with a given budget size.

Sample size per deck (aspiration level)												
Global budget per deck	1(0.60)	2 (0.67)	3 (0.73)	4 (0.76)	5 (0.77)	6 (0.79)	7 (0.81)	8 (0.82)	9 (0.83)	10 (0.83)	31 (0.90)	134 (0.95)
100	2,666	2,614	2,573	2,541	2,518	2,504	2,487	2,483	2,471	2,465	2,419	2,405
200	–	2,830	2,753	2,685	2,639	2,604	2,583	2,561	2,544	2,532	2,441	2,405
300	–	–	2,929	2,829	2,760	2,710	2,673	2,647	2,617	2,599	2,471	2,412
400	–	–	–	2,973	2,881	2,817	2,763	2,726	2,691	2,667	2,493	2,420
500	–	–	–	–	3,002	2,917	2,853	2,811	2,771	2,734	2,515	2,428
600	–	–	–	–	–	3,023	2,949	2,890	2,844	2,801	2,537	2,428
700	–	–	–	–	–	–	3,039	2,975	2,917	2,869	2,566	2,435
800	–	–	–	–	–	–	–	3,054	2,991	2,936	2,588	2,443
900	–	–	–	–	–	–	–	–	3,064	3004	2,610	2,450
1000	–	–	–	–	–	–	–	–	–	3,071	2,632	2,450

number of observations to choose relatively accurately between gambles when the difference between their expected values is large.

Some potential benefits of small samples

Why do people rely on small samples? Standard accounts implicate (cognitive and economic) costs such as those involved in internal and external search, opportunity costs, lack of appreciation for the empirical law of large numbers, or they attribute frugal search to limits in our cognitive architecture. It is only recently that cognitive psychologists and cognitive ecologists have begun to explain frugal information search in terms of potential benefits of small samples. Some of these benefits are less disputed than others. One advantage is, for instance, the enhanced ability to detect environmental change. An organism remembering only a small number of recent events—tantamount to drawing small samples from memory rather than from the environment—is better equipped to detect a change in its environment than it would be if it remembered all of its history (e.g., Heinrich, 1979; McNamara & Houston, 1985, 1987; Shafir & Roughgarden, 1998). The optimal number of items to be remembered depends on the rate of the changes in the environment, but perfect memory appears to be a liability rather than an advantage in a world that continues to change.

Real (1992) investigated the foraging behavior of bees across different floral reward distributions, and concluded that 'bees frame their decisions on the basis of only a few visits' (p. 133). He then argued that calculating reward probabilities based on small frame lengths could prove advantageous under several scenarios, one of which takes the structure of bees' natural habitat into account:

> Short-term estimation 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 landscapes'), averaging obscures the true nature of the environment (p. 135).

In psychology, Kareev (1995, 2000) advanced the argument that the cognitive system—more precisely, working memory—may have evolved so as to increase the chance for early detection of covariation. In Kareev (2000), he argued that the experienced sample size most conducive to the detection of *useful* binary correlations (i.e., value ≥ 0.5) is close to Miller's (1956) estimate of the limited capacity of working memory. This conjecture of the evolutionary advantage of small samples (or, more precisely, a limited working memory) has fueled a controversial debate (Anderson *et al.*, 2005; Juslin *et al.*, 2006; Juslin & Olsson, 2005; Kareev, 2005; and see also Fiedler & Kareev, 2006). Among other issues, it focuses on the question of what elements should enter the payoff function in determining the benefits or lack thereof of small samples (e.g., false alarms).

We do not conjecture that small samples provide a more veridical picture of the gambles in our environment. They do not (see Fig. 10.4). Our key points are: (a) Drawing small samples from payoff distributions leads to experienced differences of sample means that are larger than the objective difference; (b) such biased differences may make the choice between payoff distributions simpler, and may thus explain the frugal sampling behavior observed in decisions from experience investigated by Hertwig *et al.* (2004) and Weber *et al.* (2004); (c) sampling larger samples, *ceteris paribus*, gives rise to more accurate knowledge of the objective parameters of the payoff distributions; (d) more accurate knowledge derived from larger samples, however, yields surprisingly modest gains in terms of the probability of selecting the higher-value distribution (diminishing return of the value of information); and (e) choices based on small samples although not optimal are surprisingly competitive. Last but not least, our analysis, although conducted in the domain of monetary gambles, generalizes to other realms of life that can be modeled in terms of the binomial distribution.

Sample size and choice difficulty: empirical evidence

Small samples amplify the difference between the options' average rewards, thus, we suggest, easing the difficulty of choosing between them. This thesis gives rise to several testable predictions. One is that if people sample large samples before they choose, the options will appear more similar and choice will be more difficult. One indicator of more difficult choices is if choice proportions are closer to 50%. In another study (Hau *et al.*, in press) we conducted two experiments where people used larger samples to make a choice. In the first experiment, as in Hertwig *et al.* (2004), participants decided when to stop sampling and make a choice. Choice accuracy, however, was stressed with two procedural changes from Hertwig *et al.* Stakes were ten times higher and people were required to explicitly estimate the relative frequencies of outcomes. People drew a median sample of 33 draws (relative to 15 draws in Hertwig *et al.*) from both decks. In the second experiment, participants were required to make 100 draws from both decks. Both experiments used the same six problems as Hertwig *et al.* Consistent with the prediction that larger samples render choice more difficult, in 10 out of 12 problems choice proportions were—relative to Hertwig *et al.*—closer to 50% in Hau *et al.* $Pr(10 \text{ or more by chance}) < 0.02$. Choice proportions were 12 percentage points (median) closer.

The present account also sheds new light on research regarding people's proclivity to reason statistically in everyday problems. Fong *et al.* (1986) asked respondents to consider everyday problems to which the law of large numbers could be brought to bear. In one problem, people were asked to decide which of two car manufacturers, Volvo or Saab, was more likely to produce cars free of troublesome repairs. Respondents could resort to two pieces of information, namely, (a) objective information in terms of the consensus of *Consumer Reports*' experts and readers, and (b) personal experience of three friends who owned one of the cars. Fong *et al.* found that most people did not spontaneously reason statistically, thus being guided more by personal experience rather than the aggregate consensus. The amplification effect offers one possible explanation as to why. The large sample, represented by the aggregate consensus in *Consumer Reports*, suggests that 'both cars were very sound mechanically, although the Volvo felt to be *slightly* superior on some dimensions'

(p. 285, emphasis added). The small sample, represented by the friends' experience, in contrast, reports a huge difference between both cars: 'Both Saab owners reported having had a few mechanical problems but nothing major. The Volvo owner exploded when asked how he liked his car. 'First that fancy fuel injection computer thing went out: $250 bucks. [...]. I finally sold it after 3 years for junk".' (p. 285). One way to interpret the lure of personal experience is that the small sample amplifies the difference between the two options, whereas the large sample reports only a slight difference. The small sample thus renders the choice easier, albeit pointing to the option that is likely to be inferior.

The generality of our analysis

One challenge to the generality of the current analysis is that it is predicated on the assumption that people derive their choice from differences in the samples' average rewards. There are, however, numerous choice strategies that do not exploit this dimension. Take, for instance, the minimax heuristic according to which the decision maker selects the payoff distribution with the highest minimum payoff, irrespective of any probability information. This heuristic was originally suggested for choice under ignorance (see Luce & Raiffa, 1957), that is, for choice when probabilities are unknown. In decisions from experience, probabilities are indeed initially unknown and small samples can result in substantial estimation biases. Therefore resorting to a strategy that ignores probabilities may be a justifiable strategy. Although the minimax heuristic may not be unreasonable, people do not employ it, or many other strategies for that matter in decisions from experience (Hau *et al.*, in press).

As our analysis has revealed the amplification effect, however, is not restricted to samples' average rewards. It also occurs when the experienced outcomes and probabilities per sample are entered into cumulative prospect theory's value and weighting function (Fig. 10.4). This is not surprising because a sample's natural mean is quantitatively (though not computationally) identical to the sample's expected value (assuming the experienced probabilities and outcomes). Any choice theory that retains the Bernoullian framework scaffolding (see Brandstätter *et al.*, 2006)—namely, the assumption that people behave as if they multiplied some function of probability and value, and then maximized—will entail some kind of the amplification effect. In other words, all contemporary models that Selten (2001) subsumed under the label 'repair' program—that is, models that introduce psychological variables such as emotions, reference points, decision weights to rescue the original multiplication scaffolding—imply amplification effects. Examples of such models include disappointment theory (Bell, 1985; Loomes & Sugden, 1986), regret theory (Bell, 1982; Loomes & Sugden, 1982), the transfer-of-attention exchange model (Birnbaum & Chavez, 1997), decision affect theory (Mellers, 2000)—to name a few.

The amplification effect also generalizes to another class of descriptive choice models, namely, experiential-learning models (e.g., Busemeyer & Stout, 2002; March, 1996). They conceptualize human choice not in terms of a calculated expected utility, but as a response gauged from experience. On this assumption people choose one deck over the other on the basis of the experienced outcomes, which are integrated through an averaging operation and in which recent outcomes can have more weight than previous ones. Due to this recency weighting (for example see Hertwig *et al.*, 2006), experiential-learning models would yield amplified differences in the gambles' values regardless of the actual number of observations drawn. To conclude, the occurrence of the amplification effect can be expected to occur across a wide range of descriptive choice models, ranging from Neo-Bernoullian modifications of expected utility theory to experiential-learning models.

Admittedly, the difference in the samples' average reward is not the only dimension from which people can derive their decision. Alternatively, they can encode the frequency of occurrence of different outcomes in each payoff distribution, and rest their decision on those. Indeed, the human mind is often conceptualized as acutely sensitive to frequency information (e.g., Zacks & Hasher, 2002). Of course, with event frequencies, differences between them will, on average, increase as the sample grows larger. In other words, here large rather than small samples would amplify differences. Event frequencies, however, can hardly be the sole basis for the decision between two payoff distributions for at least two reasons. One is that in payoff distributions with more than one outcome (and each of our gambles had two outcomes), frequency information may be inconclusive: For instance, one of gamble *a*'s two outcomes may be more frequent and one less frequent than both of gamble *b*'s two outcomes. So, which gamble is better? Second, a very frequent outcome may be small in say monetary value, and a rare outcome's value may be vast, thus ignoring outcome information altogether can paint a quite misleading image of a gamble's value. Sample means, in contrast, integrate both value and frequency information into one dimension. In addition, a person can arrive at sample means without the cognitively demanding operation of weighing outcomes by the likelihood but merely by summing and dividing by the sample size; if sample sizes are identical, then just summing suffices.

Conclusion

The life-is-a-gamble metaphor suggests that life's decisions have the same structure as games of chance. If they indeed do, investigating how people play those games promises insight into how they play the game of life. Hundreds of studies in decisions under risk have provided people with convenient descriptions of games of chance, that is, their monetary outcomes and probabilities, thus giving rise to *decisions from description*. Life's decisions, however, do not often come nicely packaged. Investigating how people make decisions in the absence of complete knowledge of the gambles' properties, that is, *decisions from experience*, turns out novel questions and revives old unanswered ones. Understanding the psychology of decisions from experience requires an ecological approach to cognition (see Anderson, 1991; Brunswik, 1943; Simon, 1956), encompassing models of both the cognitive processes and their interaction with the statistical structures of the environment. Early termination of information search and the tendency to rely on small samples in decisions from experience can at least partly be understood in terms of the statistical structures of the world.

Authors' note

We are grateful to Nick Chater, Klaus Fiedler, Robin Hau, and Robin Hogarth for their helpful comments. We also thank Laura Wiles for editing the manuscript, and the Swiss National Science Foundation for a grant to the first author (100013-107741/1).

References

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, The Netherlands: Reidel (originally published in 1953).

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences,* **14,** 471–517.

Anderson, R. B., Doherty, M. E., Berg, N. D., & Friedrich, J. C. (2005). Sample size and the detection of correlation—a signal detection account: Comment on Kareev (2000) and Juslin and Olsson (2005). *Psychological Review,* **112,** 268–279.

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. (1982). Regret in decision making under uncertainty. *Operations Research,* **30,** 961–981.

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

Berenson, A. (2006, December 4). End of drug trial is a big loss for Pfizer and heart patients [electronic version]. *The New York Times.* Retrieved 2007 May 31 from: http://www.nytimes.com/2006/12/04/health/04pfizer.html

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica,* 22, 23–36 (Original work published 1738).

Bernoulli, J. (1969–1975). *Ars conjectandi.* In *Die Werke von Jacob Bernoulli* (Vol. 3, pp. 107–259). Basel, Switzerland: Basel Naturforschende Gesellschaft. (Original work published 1713.)

Bernstein, B. L. (1996). *Against the gods: The remarkable story of risk.* New York: Wiley & Sons.

Berry, D., & Fristedt, B. (1985). *Bandit problems.* London: Chapman and Hall.

Birnbaum, M., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes,* **71,** 161–194.

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

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review,* **50,** 255–272.

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., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the bechara gambling task. *Psychological Assessment,* **14,** 253–262.

Buss, D. M. (Ed.). (2005). *The handbook of evolutionary psychology.* New York: Wiley.

Daston, L. (1988). *Classical probability in the enlightenment.* Princeton, NJ: Princeton University Press.

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.

Dudai, Y., & Carruthers, M. (2005). The janus face of Mnemosyne. *Nature*, **434**, 567.

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

Fiedler, K., & Kareev, Y. (2006). Does decision quality (always) increase with the size of information samples? Some vicissitudes in applying the law of large numbers. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **32**, 883–903.

Fong, T. G., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, **18**, 253–292.

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.

George, A. L. (1959). *Propaganda analysis: A study of inferences made from Nazi propaganda in World War II*. Evanston, IL.: Row, Peterson and Co.

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. New York: Little, Brown and Company.

Goldstein, W. M., & Weber, E. U. (1997). Content and discontent: Indications and implications of domain specificity in preferential decision making. In W. M. Goldsmith & R. M. Hogarth (Eds.), *Research on judgment and decision making: Currents, connections and controversies* (pp. 566–617). Cambridge, England: Cambridge University Press.

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.

Griffin, D., & Tversky, A. (1992). The weighting of evidence and the determinants of confidence. *Cognitive Psychology*, **24**, 411–435.

Groopman, J. (2006, December 18). The right to a trial. *The New Yorker*, p. 41.

Hau, R., Pleskac, T., Kiefer, J., & Hertwig, R. (2007, in press). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*.

Heinrich, B. (1979). 'Majoring' and 'minoring' by foraging bumblebees, *Bombus vagans:* An experimental analysis. *Ecology*, **60**, 245–255.

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). The role of information sampling in risky choice. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 72–91). New York: Cambridge University Press.

Hertwig, R., & Pleskac, T. (2007). *Decisions from experience: Small samples and the perception of differences*. Manuscript submitted for publication.

Juslin, P., Fiedler, K., & Chater, N. (2006). *Less is more in covariation detection—or is it?* New York: Cambridge University Press.

Juslin, P., & Olsson, H. (2005). Capacity limitations and the detection of correlations: Comment on Kareev (2000). *Psychological Review*, **112**, 256–267.

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

Kaplan, M., & Kaplan, E. (2006). *Chances are ... Adventures in probability.* New York: Viking.

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.

Kareev, Y. (2005). And yet the small-sample effect does hold: Reply to Juslin and Olsson (2005) and Anderson, Doherty, Berg, and Friedrich (2005). *Psychological Review*, **112**, 280–285.

Knight, F. H. (1921). *Risk, uncertainty, and profit.* New York: Sentry Press.

Levitt, D. S., & Dubner, S. J. (2005). *Freakonomics: A rogue economist explores the hidden side of everything.* New York: Harper Collins.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, **92**, 805–824.

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: John Wiley & Son.

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

McNamara, J. M., & Houston, A. I. (1985). Optimal foraging and learning. *Journal of Theoretical Biology*, **117**, 231–249.

McNamara, J. M., & Houston, A. I. (1987). Memory and the efficient use of information. *Journal of Theoretical Biology*, **125**, 385–395.

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

Miller, G. A. (1956). The magical number 7, plus or minus 2—some limits on our capacity for processing information. *Psychological Review*, **63**, 81–97.

Real, L. A. (1992). Information processing and the evolutionary ecology of cognitive architecture. *American Naturalist*, **140**, 108–145.

Savage, L. J. (1954). *The foundations of statistics.* New York: John Wiley & Sons.

Selten, R. (2001). What is bounded rationality? In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 13–36). Cambridge, MA: MIT Press.

Shafir, S., & Roughgarden, J. (1998). Testing predictions of foraging theory for a sit-and-wait forager, *anolis gingivinus*. *Behavioral Ecology*, **9**, 74–84

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, **63**, 129–138.

Stoppard, T. (1967). *Rosencrantz & Guildenstern are dead.* New York: Grove Press Books.

Tversky, A., & Fox, C. R. (1995). Weighting risk and uncertainty. *Psychological Review*, 95, 269–283.

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

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press (1980).

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

Winterfeldt, D. von., & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge, England: Cambridge University Press.

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin and Review*, 12, 387–402.

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