



Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Decisions from experience: Why small samples?

Ralph Hertwig^{a,*}, Timothy J. Pleskac^b^a University of Basel, Department of Psychology, Missionsstrasse 60/64, 4055 Basel, Switzerland^b Michigan State University, Department of Psychology, 231 Psychology Building, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:

Received 4 July 2007

Revised 14 December 2009

Accepted 16 December 2009

Keywords:

Sampling

Decisions from experience

Choice

Small samples

Amplification effect

ABSTRACT

In many decisions we cannot consult explicit statistics telling us about the risks involved in our actions. In lieu of such data, we can arrive at an understanding of our dicey options by sampling from them. The size of the samples that we take determines, *ceteris paribus*, how good our choices will be. Studies of decisions from experience have observed that people tend to rely on relatively small samples from payoff distributions, and small samples are at times rendered even smaller because of recency. We suggest one contributing and previously unnoticed reason for reliance on frugal search: Small samples amplify the difference between the expected earnings associated with the payoff distributions, thus making the options more distinct and choice easier. We describe the magnitude of this amplification effect, and the potential costs that it exacts, and we empirically test four of its implications.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Being freed of tasks such as mating and territorial defense, worker bees are particularly well suited for evolutionary studies of choice (Real, 1992). Time and again, they need to decide which area of potential food resources to exploit—a choice rendered thorny due to changing ecological conditions. Using an artificial patch of flowers with varying colors signaling different amounts and probabilities of nectar rewards, Real (1991, 1992) investigated how a bee made choices in its uncertain habitat. He found that if a bee were maximizing expected utility, it chose as if it “misjudges the objective probabilities and underestimates the rare event”—a finding that is in “contrast with the conclusions of Kahneman and Tversky” (Real, 1992 p. S132). Kahneman and Tversky (1979) asserted that rare events—at least for humans—have more impact than they deserve according to their objective probabilities.

Investigations giving rise to the conclusion of opposite perceptions of rarity, however, have dealt their agents—

bees versus humans—very different cards. In Real's studies (1991) bees were initially ignorant of the habitat's distribution of nectar and learned about their environment while foraging. What Real described as bees' misjudgments of rare events relates to the discrepancy between bees' implicit estimates of the likelihood of the rewards, inferred from their behavior, and actual probabilities. In contrast, during studies on how humans make decisions under risk, people receive perfect information about the payoff associated with each option and the probability of those payoffs (e.g., 3 with certainty versus 32 with probability of .1, 0 otherwise). They thus make what Hertwig, Barron, Weber, and Erev (2004) referred to as *decisions from descriptions*, rendering foraging for information and estimating probabilities superfluous.

2. Perception of rarity and small samples

What happens when people, like bees, sample information from uncertain environments, thus making *decisions from experience* (Hertwig et al., 2004)? Such decisions represent situations in which, being ignorant of the payoffs, agents may resort to experience: experience garnered

* Corresponding author. Tel.: +41 61 267 0611; fax: +41 61 267 0441.
E-mail address: ralph.hertwig@unibas.ch (R. Hertwig).

through sampling from the payoff distributions and eventually arriving at what Knight (1921) called “statistical probabilities” based on the empirical classification of instances (see Hau, Pleskac, & Hertwig, 2010). Such sampling can occur in two fundamentally different ecologies, one with an inherent trade-off between exploiting and exploring options (see Barron & Erev, 2003; Berry & Fristedt, 1985; Erev & Barron, 2005), and one in which the agent’s only objective, at least initially, is exploration for information (Hertwig et al., 2004; Weber, Shafir, & Blais, 2004). In the former ecology, the sampled outcomes simultaneously provide information and payoffs to the agent. In contrast, the sampled outcomes in the latter exploration-only environment merely provide information, much like attending a free wine-tasting fair, perusing the Gault-Millau or the Michelin Guide to select a restaurant, or taking a quick look at the online traffic cams. Exploitation of the options for payoffs—for example, dining at one of the acclaimed gastronomic temples—only comes after search for information was terminated.

We focus on this second ecology. Hertwig et al.’s (2004) study illustrates how it can be investigated. Participants were asked to choose between two gambles (payoff distributions). Lacking knowledge about them, they could explore each distribution by sampling from them. Specifically, people saw two boxes on a computer screen representing two possible gambles or outcome distributions. Clicking on a box triggered a random draw of an outcome from the associated distribution. They were encouraged to sample until they felt confident enough to decide which box was “better”, in the sense that they would prefer to draw from it during a final trial involving real monetary payoffs. Hertwig et al. compared respondents’ choices in this final trial with those of a group who made decisions from description from the same problems. Choices differed drastically and systematically. In description-based choices, rare outcomes were consistent with prospect theory’s assumption of overweighting of rare events (Kahneman & Tversky, 1979). In experience-based choices, however, people behaved as if rare outcomes had less impact than they deserved according to their objective probabilities.¹ This description–experience gap has been replicated across a wide range of studies (e.g., Erev, Gluzman, & Hertwig, 2008; Hau et al., 2010; Hau, Pleskac, Kiefer, & Hertwig, 2008; Rakow, Demes, & Newell, 2008; Ungemach, Chater, & Stewart, 2009; Weber et al., 2004; for a review of experience-based decision making see Hertwig & Erev, 2009, and for a special issue on experience-based decision making see Rakow & Newell, 2010).

For bees and humans, the source of the relative lack of appreciation of rare events appears to be the same: *the small samples on which they base their choice*. Based on investigations across different floral reward distributions, Real (1992) concluded that “bees frame their decisions on the basis of only a few visits” (p. S133). Focusing on hu-

mans, Hertwig et al. (2004) found that the typical number of draws that respondents made was approximately seven from each deck (and a median of 15 across both decks). Hau et al. (2010) reviewed the sample sizes in five subsequent decisions from experience studies in which people were instructed to sample until they felt confident enough to decide between the two payoff distributions, and in which monetarily only their final choice mattered (unlike in Barron & Erev, 2003). Consistent with Hertwig et al.’s original observation, sample sizes fell in a surprisingly small range from 9 to 19 draws, amounting to nearly 7 ± 2 draws from each deck. The only outlier was Hau et al.’s (2008) observation of 33 draws (Study 2) in a condition involving an order of magnitude larger payoff than in other studies.

Based on these observations, it seems fair to say that so far, *ceteris paribus*, the finding of relatively small sample sizes in decisions from experience studies is robust. Admittedly, referring to these sample sizes as “small” is somewhat arbitrary, as, of course, samples could have been even smaller. Yet, if one keeps in mind that each draw requires an investment of merely a few seconds and that people’s small samples systematically and noticeably misrepresented the probabilities of rare events, the conclusion that people in these studies framed their decisions on relatively “small” samples appears justified. For example, in Hau et al.’s (2008) Study 1, respondents who sampled a median of 11 draws across both distributions did not even encounter the rare event in 50.3% of all trials, and the median difference between the experienced relative frequency and the objective probability of the nonzero outcomes was 10.9 percentage points.

Small samples interact with the statistical structure of the environment in systematic ways. To see this, consider a gambling environment with binomially distributed outcomes, and 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/or 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. For such distributions, one is more likely to encounter the rare event less frequently than expected (np) than more frequently than expected. For illustration, let us assume that a person samples 10 times from a distribution in which the critical event has a probability of .1 and estimates the event’s probability to be the proportion in the sample that she observes. The probabilities that she will observe the critical event more than once, less than once, or exactly once are .26, .35, and .39, respectively. That is, the person is more likely to underestimate than to overestimate the frequency of the rare event (.35 versus .26). This asymmetry decreases the larger n becomes, for example, from 9 percentage points to 7, 6, and 5 percentage points for sample sizes of 10, 20, 30, and 40, respectively.

Next, we consider possible reasons why people appear content with small samples, and how the fact that the rare event is more likely to be under- than overrepresented in small samples changes the options that people actually experience.

¹ Fox and Hadar (2006) correctly pointed out that this weighting pattern in experience-based choice need not be in conflict with prospect theory, and that prospect theory can account for decisions from experience if its input is the sampled rather than the experienced probabilities (see also Hau, Pleskac, Kiefer, & Hertwig, 2008).

3. Why do people rely on small samples?

There are a number of possible and not exclusive factors that may contribute jointly to people's reliance on small samples. One could be opportunity costs. Searching for information takes time—time during which people cannot pursue other activities. In other words, in a context in which exploration and exploitation do not coincide, sampling is costly in terms of the things that one has to forgo. This fact matters even more if the payoffs are such that not making optimal choices (say in terms of expected value or utility maximization) does not exact a substantial cost (see Harrison, 1992; Smith, 1982, on the precept of payoff dominance; and von Winterfeldt & Edwards, 1986, for a discussion of flat maxima). Indeed, Hau et al.'s (2008) finding that increasing payoffs by an order of magnitude can triple sample size indicates that opportunity costs and payoff magnitude shape people's search efforts.

Another possible contributory factor to the termination of search is limited short-term memory. Specifically, such a limit could provide a natural stopping rule for information acquisition (Kareev, 2000). In support of this prediction, Rakow et al. (2008) found that people's working memory capacities were related to the sample sizes that they used to make decisions from experience (Rakow et al., 2008).²

Not ruling out these and other possible factors, we suggest that another reason why respondents' sampling is economical may be because small samples make it easier to choose between the options. For illustration, consider a person who draws seven observations from deck A that (in the long run) offers \$32 with probability .1, \$0 otherwise, and seven from deck B that offers \$3 for sure (a decision problem studied by Hertwig et al., 2004). She may encounter a "0" seven times from deck A, and a "3" each time from deck B, amounting to a sample mean of 0 and 3, respectively. Let us also assume that the person determines her choice on the basis of the absolute difference between the sample means, an assumption to which we return later. In our example, the absolute difference between the sample means, 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. This amplification was the rule rather than the exception in Hertwig et al. In 81% of choices, experienced differences were amplified—on average, by a factor of 10.8—relative to description differences. Similarly, in Rakow et al.'s (2008) study, the median absolute difference between the sample means was 2, relative to the median difference between the expected values of 0.125 based on the objective probabilities—an amplification by a factor of 16 (see Hau

et al., 2010). As a consequence of small samples, options appeared more dissimilar and therefore, *ceteris paribus*, more easily discernable than they would be on the basis of their objective properties, and (some) people may make decisions sooner than they would do otherwise.

To avoid misunderstandings, we *do not* argue that amplified differences make it easier to detect which gamble has the objective highest payoff. Instead, we assert that—much like Griffin and Tversky (1992) proposed with confidence judgments—decision makers focus, but not exclusively, on the *strength of evidence* as opposed to the *weight of evidence* in decisions from experience and consequently terminate search early. Griffin and Tversky illustrated the distinction between the strength and weight using another sequential process, repeated coin flipping:

[S]uppose we wish to evaluate the evidence for the hypothesis that a coin is biased in favor of heads rather than in favor of 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., a difference between two means) and its reliability (e.g., the standard error of the difference). Although it is not always possible to decompose the impact of evidence into the separate contributions of strength and weight, there are many contexts in which they can be varied independently. . . . the same proportion of heads can be observed in a small or large sample (p. 412).

Analogously, in choosing between gambles during decisions from experience, the strength would be the difference between sample means and the weight would be the standard error of the difference. If people encounter a large difference in a relatively small sample, and if they value strength of evidence more than weight of evidence, this difference may be "good enough" for them to stop search. Just as in Simon's (1983) theory of search, if the difference between sample means surpasses an *aspiration level* (i.e., is good enough) no further information will be required and a choice is rendered.

In fact, assuming the same decision problem, the expected strength of evidence will always be as large or larger in experienced-based than in description-based choice. Specifically, the *absolute expected difference* between the sample means, SM_A and SM_B , will always be as large or larger than the expected or description difference, $EV_A - EV_B$. Hertwig and Pleskac (2008, pp. 216–217) provided a simple proof for this fact (see Appendix A). The intuition behind this proof is that the distribution of absolute experienced differences is asymmetrical and small samples tend to increase this asymmetry by increasing the proportion of large experienced differences that point to either gamble. As a result, the expected absolute difference is pulled away from the long run expected difference, $EV_A - EV_B$, or the *description difference* based on the gamble's expected values. The proof, however, leaves several questions unanswered, including: What is the magnitude of the amplification effect in a given uncertain environ-

² The effect of a limited short-term memory on people's final choice has been less consistent. Hertwig et al. (2004) reported evidence for a recency effect, whereas Rakow et al. (2008) found mixed evidence and Hau et al. (2008) could not replicate it. Hau et al. (2010) also found no difference in choice during experience-based decision making whether participants were provided with a running record of their past sampled observations, or not. The former condition should eliminate the need to maintain items in memory when making a choice.

ment? And how often does the absolute sampling difference point the decision maker to the inferior gamble?

4. The amplification effect: how substantial and how costly is it?

In order to answer these questions, Hertwig and Pleskac (2008) developed an analytical framework that enables one to determine both the magnitude and the costs of the amplification effect in any given gambling environment. The details of this framework are described in Hertwig and Pleskac; here we review the main findings. The analysis assumes that a person samples from two payoff distributions (decks), each one consisting of two outcomes of the type “a probability p to win amount x ; otherwise win amount y ” (x, p, y). The size of the sample that the person draws ranges from 1, 2, ..., to 50, from either deck. The gambling environment consists of 1000 pairs of randomly generated gambles. 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.

As a function of sample sizes, Fig. 1 plots the expected (and the median) absolute values of the experienced differences, across all gambles. The straight line represents the average description difference. Small samples substantially amplify the difference between gambles. For example, with two draws from each deck, the average expected experienced difference is 23.1, 1.5 times larger than the description difference of 15.2. Because of the law of large numbers, as sample sizes increase the expected experienced differences converge to the description difference. With 10 draws per deck, the expected experienced difference (17.2) is merely 1.1 times larger than the description difference, and with 25 draws, the differences are nearly the same. The same regularity emerges for the median experienced difference, except that the amplification effect is smaller relative to the expected absolute values.

Fig. 1 also clarifies another important point. People generally do not terminate sampling when the absolute difference is largest, that is, after two draws per deck. Rather, they appear to stop after 7 ± 2 draws per deck. This is consistent with the notion that the decision of whether to terminate search is by no means only determined by the strength of evidence (the difference between the sample means) but to some extent also by its reliability (the stan-

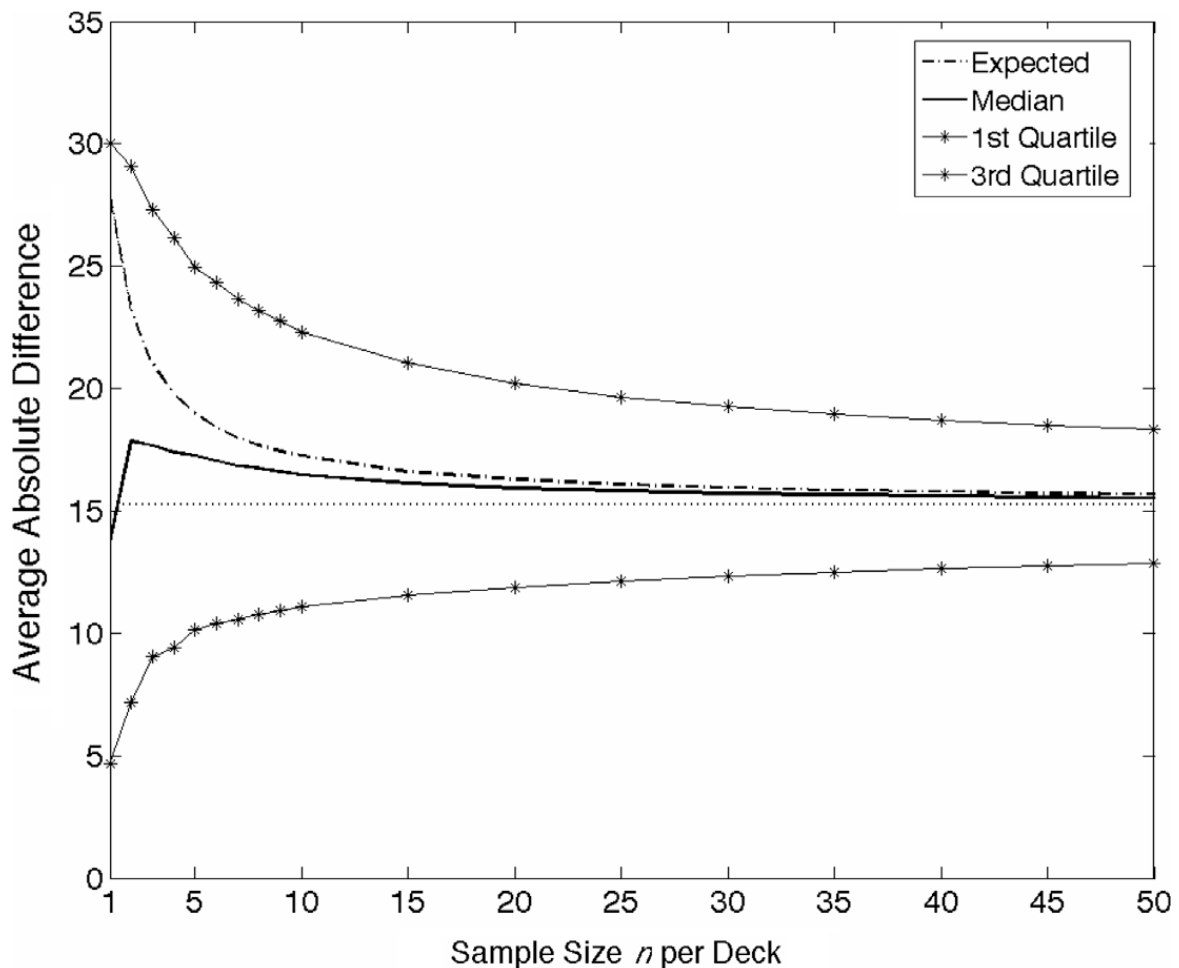


Fig. 1. Experienced differences across 1000 pairs of gambles as a function of sample size (per deck). The curves represent (a) the mean of the expected absolute difference, (b) the median of the experienced absolute difference and (c) the first and third quartiles of the experienced absolute difference. The straight horizontal line represents the average description difference based on expected value (15.2).

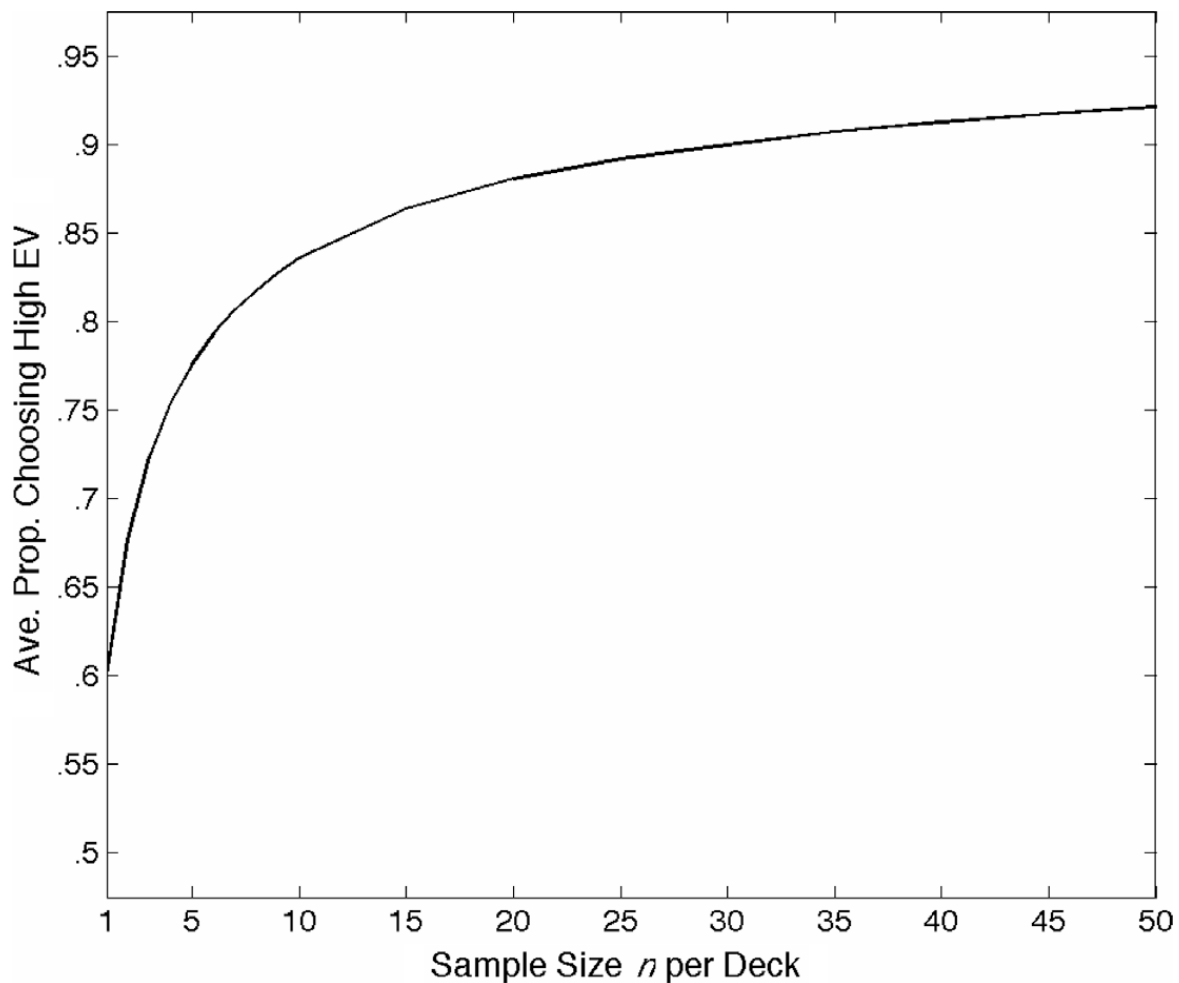


Fig. 2. Choice proportions of the deck with the higher expected value (EV) as a function of sample size (per deck), averaged across 1000 pairs of gambles. The assumption is that for each n the deck with the larger sample mean is selected.

dard error of the difference, or a simple proxy of it, the sample size). That people have some notion of the reliability of parameter estimates is both consistent with the assertion by Jacob Bernoulli, who formulated the first version of the law of large numbers, that “even the stupidest man knows by some instinct of nature *per se* and by no previous instruction” that the greater the number of confirming observations, the surer the conjecture (cited in Daston, 1988; Gigerenzer et al., 1989, p. 29), and with the available evidence regarding people’s intuition about sample size (see Sedlmeier & Gigerenzer’s, 1997, review and distinction between frequency and sampling distributions).

The old adage of economists goes, “There’s no such thing as a free lunch”. Indeed, 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 a somewhat inaccurate representation of the gambles’ parameters, and as a possible consequence, less than optimal choices. How big is this price? To find out how costly relying on small samples is, Hertwig and Pleskac (2008) calculated how often a respondent who invariably selected the option with larger sample mean would succeed in choosing the objectively better gamble (accuracy) as a function of sample size (effort). Fig. 2 summarizes the results of this effort–accuracy analysis.

Although choices derived from small samples are not optimal, they are surprisingly good. With a sample as tiny as *one* draw the chance to select the better gamble is approximately 60%. Drawing as few as seven times from each deck offers an 81% chance of selecting the better gamble.³ Moreover, accuracy continues to increase at a diminishing rate. A person can, for example, increase the likelihood of selecting the higher expected value gamble from 60% to 78% by increasing sample size from 1 to 5. By doubling sample size from 10 to 20 draws, accuracy increases by merely 2 percentage points.

In sum, several studies that employed the sampling design as used by Hertwig et al. (2004) observed that people relied on relatively small samples, amounting to about 7 ± 2 draws from each deck. Yet, the amplification effect is only one key factor explaining why people tend to rely on relatively small samples. Several factors are likely to shape people’s search, including opportunity costs, magni-

³ Of course, how good or bad this level of accuracy is depends on several factors including a person’s payoff function. However, compared to the level of accuracy that people reach in various inferential domains—for instance, inferring which of two risks claims more lives per year (e.g., Hertwig, Pachur, & Kurzenhäuser, 2005; Pachur & Hertwig, 2006)—this level of accuracy is high.

tude of payoffs, memory constraints, and possibly the options' variance. Moreover, although a person's ability to choose the higher expected value gamble was diminished with small samples, our analysis showed that small samples are not ruinous to choosing the higher expected value gamble. Johnson, Budescu, and Wallsten (2001) reported an analogous result when aggregating probability judgments to form a new probability judgment: judges can be quite accurate when averaging very few judgments. Before we describe testable implications of the amplification effect, let us briefly turn to how the difference between options relates to difficulty of choice and to choice strategies.

5. Differences in sample means, choice difficulty, and choice strategies

Several factors make the choice between gambles easy or difficult, including the options' outcome variance or lack thereof (see Weber et al., 2004) and the extent to which outcomes and their probabilities are negatively correlated (thus causing the experience of conflicts and difficult trade-offs, e.g., Shanteau & Thomas, 2000). In light of these and other properties, how important a determinant of choice difficulty is the difference between the expected values of gambles (decisions from description) and the difference between samples' means associated with gambles (decisions from experience), respectively?

There are two converging answers to this question: one stems from psychophysics, the other from early research on risky choice. In psychophysics it is well established that as the objective difference between two stimuli (e.g., difference in line lengths or weights) grows larger, a person's judgments become more and more reliable, giving rise to the typical S-shaped psychometric function. In other words, with larger differences in the physical stimuli, it becomes easier to discriminate between them and the correct stimulus is identified more often (Link, 1992; Urban, 1910). Second, Mosteller and Noguee (1951; see also Edwards, 1955) established that an analogical psychometric function exists for choices between gambles so that as the difference between the expected values of two options grew so did the probability of choosing one of the options. When the options were of essentially equal expected value the probability was close to .5. In other words, the strength of preference grows as a function of the difference in expected values of gambles (Busemeyer & Townsend, 1993).

In experienced-based choice, the amplification effect is likely to play a role because the difference between the sample means is a cue relevant to several choice strategies that people may exploit. One is the *natural-mean heuristic* (Hertwig & Pleskac, 2008). It was originally proposed in the context of *n*-armed bandit problems (Sutton & 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 make action selection decisions. According to Sutton and Barto, "the true value of an action is the mean reward received when that action is selected. One natural way to estimate this is by averaging the re-

wards actually received when the action was selected" (p. 27). Applied to decisions from experience, the natural-mean heuristic consists of two steps:

Step 1. Calculate the natural mean of experienced observations in both decks by summing, separately for each deck, all *n* experienced outcomes and then dividing by *n*.

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

As Hau et al. (2008) pointed out, the natural-mean heuristic has several interesting characteristics. For one, the natural-mean heuristic makes the same outcome or choice predictions as the expected value calculus applied to sampled probabilities and outcomes. However, the heuristic arrives at the choice through a shortcut. Rather than keeping track of all outcomes and their respective probabilities, multiplying each by the other and summing up, the heuristic tots up all experienced rewards (or losses) per gamble and then divides this sum by the sample size per deck to yield the "natural mean." In a model competition between 15 choice models, Hau et al. found that the natural-mean heuristic was among a small set of models that performed best in predicting people's decisions from experience. Of course, the natural-mean heuristic will always choose the option with the favorable sample mean, regardless of how large this advantage is. However, it appears plausible to assume that a user of the natural-mean heuristic will find the choice the easier, the larger the difference in sample means.

The natural-mean heuristic, however, is not the only choice model subject to the amplification effect. Hertwig and Pleskac (2008) also showed that the amplification occurs when the experienced outcomes and probabilities per sample are entered into cumulative prospect theory's value and weighting function (see their Fig. 10.1). This is not surprising because a sample's natural mean is quantitatively (although not computationally) identical to the sample's expected value (assuming the experienced probabilities and outcomes). More generally, any choice theory that retains the Bernoullian framework scaffolding (see Brandstätter, Gigerenzer, & Hertwig, 2006)—namely, the assumption that people behave as if they multiplied some function of probability and value, and then maximized—will entail the amplification effect.

6. The amplification effect: an empirical test

Our thesis is that, relative to large samples, small samples amplify the difference between the expected earnings associated with the payoff distributions, thus rendering the options more distinct and choice easier, and that ease is one of several factors that determine termination of search in experienced-based choice. This thesis has several testable implications:

Prediction 1. Assuming identical problems, people who make experienced-based choices and rely on small

samples will perceive choice as easier than respondents who make description-based choices.

This prediction assumes the difference between the sample means to be one determinant of choice difficulty, and builds on the fact that smaller samples will, *ceteris paribus*, result in larger differences between the experienced sample means (Fig. 1 and Appendix A), thereby making the options more distinct.

Prediction 2. If experienced-based choices are based on small samples and this sample information is being described to yoked partners, then they will perceive the resulting decisions from description as easier than respondents who make decisions from descriptions involving the objective probabilities.

This prediction is derived as follows: One group's sampled experience can be summarized and described to another group. Thereby, this yoked-description group responds to the same probabilities and outcomes that the experience group saw (see Hau et al., 2010; Rakow et al., 2008). Moreover, the yoked-description group responds to gambles in which the differences between the expected values are, on average, larger (because of reliance on small samples) than in the description group involving the objective probabilities.

Prediction 3. Frugal searchers, that is, respondents who rely on small samples, will find choice easier than avid searchers.

This prediction is derived as follows: Several factors have an effect on the decision to terminate search; ease of choice is only one among them. Therefore, there will be people that sample avidly in decisions from experience.

They, in turn, will not benefit from the amplification effect and consequently, so we predict, perceive choice as more difficult than the frugal searchers.

Prediction 4. Frugal searchers, that is, respondents who rely on small samples, will arrive at greater preference strength than avid searchers.

One measure of strength of preference is in terms of the absolute difference of the choice proportions from the indifference point (i.e., 50%). If frugal searchers find choices easier than avid searchers, then—*ceteris paribus*—their preference may be stronger. Clearly, many factors can affect preference strength; ease of choice is only one among them. One can find support for Prediction 4 in comparisons across studies. In Hau et al.'s (2008) Studies 1 and 2, the median sample size was 11 and 33 draws (caused by the order of magnitude higher payoffs). Consistent with Prediction 4, the average absolute difference of the choice proportions from 50% was 18 and 8.8 percentage points, respectively. That is, with larger samples, respondents became more indifferent between the options. In what follows, we test whether this difference can be replicated within the same study.

These four predictions were tested in the following experiment.

6.1. Methods

A total of 265 students at Michigan State University participated in the study for class credit in addition to a monetary reward based on their choices. Eighty-nine participants in the description group saw twelve problems (Table 1) described on a computer screen. Eighty-eight participants in the experience group saw two buttons on a computer screen and were told that each button was asso-

Table 1
Summary of the decisions problems and results.

Decision problem	Option		Description group	Percentage choosing <i>H</i>	
	<i>H</i>	<i>L</i>		Experience group	Yoked-description group
1	<u>4.8</u>	3, 1 ^F	26	60*	50*
2	-3, 1.0	<u>-4.8^F</u>	46	62*	60
3	<u>4.2^F</u>	3, .25	74	47*	46*
4	-3, .25 ^F	<u>-4.2</u>	64	65	60
5	<u>32, .10^F</u>	3, 1	44	18*	33
6	<u>16, .2^F</u>	3, 1	36	24	38
7	-3, 1.0 ^F	-32, .1	64	24*	47*
8	<u>32, .025^F</u>	3, 1.0	33	21*	28
9	<u>10, .1^F</u>	1, 1.0	39	22*	36
10	10, .9	9, 1.0 ^F	18	78*	63*
11	<u>-10, .9^F</u>	-9, 1.0	47	72	70
12	<u>10, .05^F</u>	1, .5	34	35	28
Mean experience–description gap ^a				18	11

Note: Underlining indicates the options including rare events. *H* = option with the higher expected value (as calculated by probability times monetary value); *L* = option with the lower expected value. For each option only one outcome is given, followed by its probability; the second outcome, which is not stated, was 0 and occurred with a probability complementary to the stated one. F = option favored by the rare event.

* *p* < .05.

^a The mean experience–description gap was calculated as the difference in the predicted direction between proportions of participants choosing *H* in the experience group versus the description and yoked-description groups.

ciated with a different payoff distribution (money machine). They could explore the distributions by pressing one of two color-coded buttons on a keyboard, thus eliciting the sampling of one outcome (with replacement) at a time. The outcome appeared for one second. Respondents were instructed to sample until they felt confident enough to decide from which single distribution to draw for a real payoff. At that point, they indicated with a button press their preferred money machine and proceeded to the next problem. Eighty-eight participants in the *yoked-description* group saw described gambles on the computer screen, with the objective probabilities replaced with the observed relative frequency of the outcomes from a corresponding yoked partner in the experience group.

Respondents in the two description groups were asked, after each choice: “How difficult was it for you to make your final choice between the two money machines?”. In order to focus participants in the experience group on their choice rather than on the process of search, respondents in the experience group were asked: “Once you obtained enough information to make a decision, how difficult was it for you to make your final choice between the two money machines?”. All respondents rated choice difficulty on a scale from 1 (“extremely easy”) to 6 (“extremely difficult”). After a person had completed all twelve choices and ease ratings, the computer played out the chosen gambles and the participant was paid accordingly. Each point was worth \$0.01 (thus replicating the incentive structure in Hertwig et al., 2004). To make sure that nobody would lose money, all participants were given \$1.00 at the beginning of the study.

Respondents in all three groups saw 12 decision problems (see Table 1). Half of the problems were identical to the problems investigated by Hertwig et al. (2004); the other half were studied by Rakow et al. (2008); see also Weber et al., 2004. Eight problems present options with small difference between their expected value (as calculated by probability times monetary value), and four problems offer options with identical expected values (problems 9–12). Moreover, eight problems offer positive and four offer negative prospects.

7. Results

Before turning to the predictions, we report to what extent previous results were replicated. Across all 12 problems, 14 participants in the experience group made only one draw from each payoff distribution before proceeding with their choice. Because our set-up required everybody to make at least one draw from each option, they either misunderstood the instruction or were unwilling to invest any effort. We removed these participants and their corresponding “partners” in the yoked-description group. Their removal, however, did not alter the qualitative patterns of the results.

As in previous studies, decisions from experience differed from decisions from descriptions. As shown in Table 1, the percentage of respondents who chose option *H* (i.e., the option with the higher expected value, as calculated by probability times monetary value) differed mark-

edly between the groups. In 11 of the 12 problems, the ordinal difference was consistent with Hertwig et al.’s (2004) hypothesis that in decisions from experience, rare events carry less impact than they deserve according to their objective probability than in decisions from description (the only exception was problem 11). In three out of 11 problems, the description–experience gap was small (problems 4, 6, and 12), and in eight problems the differences were significantly different according to a *t*-test of proportions. Across all problems, the average (absolute) description–experience gap amounted to 18 percentage points (correcting for the hypothesis-inconsistent difference in problem 11).

Also replicating previous results, some of the impact of rare events in decisions from experience is attributable to the small samples that participants drew and the ensuing sampling error that they experienced. The median total sample from both decks was 12 (Inter-Quartile Range (IQR) = 18; $M = 18$; $std = 17$). Sampling was equally dispersed between the *H* and *L* (lower expected value) options with a median of 6 and 6 observations, respectively. Comparing choices in the experience and the yoked-description group reveals the effect of sampling error on choice. Here, the mean description–experience gap was 11 percentage points, and in only four problems was the difference significantly different. This suggests that sampling error is one source behind the description–experience gap but others are contributing as well (see Hertwig and Erev, 2009, for a review of the possible factors). Let us now turn to Predictions 1–4.

7.1. Were choices based on small samples perceived easier?

According to Prediction 1, people who rely on small samples in experienced-based decisions will find choices easier than respondents who face structurally identical problems and receive perfect information about outcomes and their probabilities. Respondents in our experience group relied on samples about as small as those observed in Hertwig et al. (2004) and others studies. In line with Prediction 1, they rated their choice easier than the description group. Across all decision problems and people, the median ease ratings for decisions from description and experience were 3 (IQR = 2) versus 2 (IQR = 3)—a significant difference (Mann–Whitney test: $z = 2.32$, $p = .021$).

7.2. Were choices based on yoked descriptions perceived easier?

According to Prediction 2, when people receive the sampled information of the experience group, condensed into descriptions of outcomes and probabilities (yoked descriptions), they will find choice easier than people facing descriptions of the objective probabilities and outcomes. In line with this prediction, the ease rating of the description and yoked-description group differed strongly, with median ease ratings of 3 (IQR = 2) and 1 (IQR = 1) in the description and yoked-description group, respectively ($z = 14.8$, $p < .001$). Note that the ease ratings in the yoked-description group are also lower than those in the experience group ($z = 11.8$, $p < .001$). One possible reason

is that despite our request to only focus on the final choice when assessing ease, the effort people mustered for sampling and updating information spilled over into their ease judgments. In contrast, the yoked-description group benefited from amplification without the additional effort of sequential search. If indeed such spill-over occurred, the observed difference in the ease ratings of the experience and description groups would be even larger without it.

7.3. Did frugal searchers find choice easier?

Because the amplification effect becomes smaller the more that people sample, those who explore (sample) more, *ceteris paribus*, should find choice more difficult. In order to test Prediction 3, we performed a median split based on sample size (across both options) in the experience group. The frugal searchers (taking 12 or fewer draws) had a median ease rating of 2 (IQR = 2), whereas the avid searchers arrived at a rating 3 (IQR = 2; $z = 2.9$, $p < .01$).⁴ This difference was also echoed in the yoked-description group. Those who saw the experience of avid searchers found choices more difficult than those who looked at the option through the eyes of frugal searchers, with ease ratings of 1 (IQR = 1) and 2 (IQR = 2), respectively ($z = 9.5$, $p < .0001$). This suggests that the reason for why avid searchers find choice more difficult than frugal searchers is not additional search *per se* but the change in the experienced difference in sample means.

The previous analysis classified respondents in the experience group according to their sampling effort. Regardless of sample size, however, one can classify them also according to whether their experienced differences exceeded the description differences. Consistent with the impact of the amplification effect, respondents whose experienced differences were larger than the respective description differences found choice less difficult (2, IQR = 3) than those who experienced equally large or smaller differences (3, IQR = 2; $z = 2.5$, $p = .01$).⁵ Again, this finding is also mirrored in the yoked-description group. Those who saw the “larger” differences found choice easier, relative to those who saw the “as large or smaller” differences, with median ease ratings of 1 (IQR = 1) versus 2 (IQR = 2), respectively ($z = 3.8$, $p < .001$).

7.4. Did frugal searchers arrive at stronger preferences?

To the extent that frugal searchers, relative to avid searchers, find choice easier, they may also arrive at stronger preferences. In order to test Prediction 4, we employed the same median split as before, and used the absolute difference of the choice proportions from the indifference point (i.e., 50%) as a measure of preference strength. In 9 of 12 decision problems, the frugal samplers had more ex-

treme choice proportions than the avid samplers ($p = .07$), and, on average, their choice proportions were 6 percentage points further from 50% than those of the avid searchers. In the yoked-description group, in 9 out of 12 cases, the yoked-frugal searchers displayed more extreme choice proportions than yoked-avid samplers ($p = .07$), and their choice proportions were, on average, 9 percentage points further from 50% than those of the yoked-avid searchers. Thus, there is evidence for a link between ease and preference strength, albeit not quite as strong as in the aforementioned reanalysis of Hau et al.'s (2008) studies. One reason could be that the difference in people's sample sizes was larger between Hau et al.'s studies, relative to the present difference between frugal and avid samplers.

In sum, we empirically tested four implications of the amplification effect. Consistent with Prediction 1, we found that the experience group (who sampled frugally) found choice easier than the description group. In line with Prediction 2, we found that the yoked-description group perceived choice to be easier than the description group. In support of Prediction 3, we found that, relative to frugal searchers, avid searchers in the experience group found choice more difficult, and people who made the same choices based on the condensed experience of avid and frugal searchers exhibited parallel differences in ease ratings. Finally, we also found indications that those who sample less and find choice easier arrive at stronger preferences.

8. General discussion

Recent research on decisions from experience using the sampling paradigm observed that respondents consistently rely on relatively small samples of information. Hertwig and Pleskac (2008) proposed a first step toward an explanation of this phenomenon. In this article, we tested implications of their amplification account. Next, we outline a framework for the termination of search, discuss how amplification may contribute to other judgments than risky choices, and review the potential benefits of small samples.

8.1. A framework for termination of search

We have focused on one possible factor, perceived ease of choice, impacting on search and its termination. Yet, as stressed before, several factors are likely to be at work in a person's decision to end exploration. Some of them foster further search, whereas others weigh in to stop. One way to conceptualize this trade-off is in terms of Coombs and Avrunin's (1977) framework of single-peaked functions. They pointed out that various dependent variables such as “preferences” or “exploratory behavior” are frequently observed to be single-peaked functions of independent variables. Recently, Reutskaja and Hogarth (2009) extended their framework to the number of alternatives available when making a choice. Coombs and Avrunin's single-peaked framework explains these functions in terms of a “compromise between good and bad attributes of a single quantity” (p. 219), and the psychological

⁴ The Goodman–Kruskal rank order correlation (G) between ease rating and sample size in the experience group was $G = .09$, $p < .01$. The rank order correlation between the ease rating of the yoked-description participants and the sample size of their partner in the experience group was $G = .37$, $p < .001$.

⁵ The Goodman–Kruskal rank order correlation between ease rating and the experienced difference was $G = -.12$, $p < .001$. The corresponding correlation for the yoked condition was $G = -.22$, $p < .0001$.

principle that “Good things satiate and bad things escalate” (p. 224).

Let us assume sample size (n) to be the single dimension in the domain of exploratory behavior. The good attributes that come with search are primarily accuracy (e.g., a more veridical picture of the world and better odds to identify the more attractive options). When such good things satiate, the utility of the benefits follow a concave, negatively accelerated function. Indeed, our accuracy analysis in Fig. 2 revealed such a concave function. Of course, search also exacts costs, for example, in terms of time, opportunity costs, and increasing memory load and retrieval efforts. When bad things escalate, such costs, minor at first, grow more and more rapidly as a function of n , so that the combined utility of such costs is also negatively accelerated. Following Coombs and Avrunin (1977), the sum of the utility functions for costs and benefits of search will be a single-peaked function over n . Thus, depending on a person's exact cost and benefit utility functions, there is an optimum in that the total utility is maximum for that n . Fig. 3 illustrates the utility curves for costs and benefits of exploration, and their combination. The peak is reached when the marginal costs per draw begin to exceed the marginal benefits. Note that a single peak would still emerge even if the costs did not escalate but increased at a constant rate.

This framework classifies the factors that impact search and termination of search, and can help explain differences in exploration in terms of either (a) differences in cost and benefit functions or (b) differences in the combination schemes for both functions. Based on this framework, experimenters could consider how various manipulations strengthen or weaken the impact of costs and benefits of exploration, respectively. For example, providing larger payoffs and requiring confidence judgments (e.g., “how certain are you that you selected the better gambles?”) may slow down the negative acceleration of the utility

curve for benefits. In contrast, imposing additional cognitive load or making time scarce may spur the negative acceleration of the costs of exploration.

8.2. Amplification beyond gambles

The amplification effect can also shed new light on classical research regarding people's proclivity to reason statistically. Fong, Krantz, and Nisbett (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 the 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 was 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 ren-

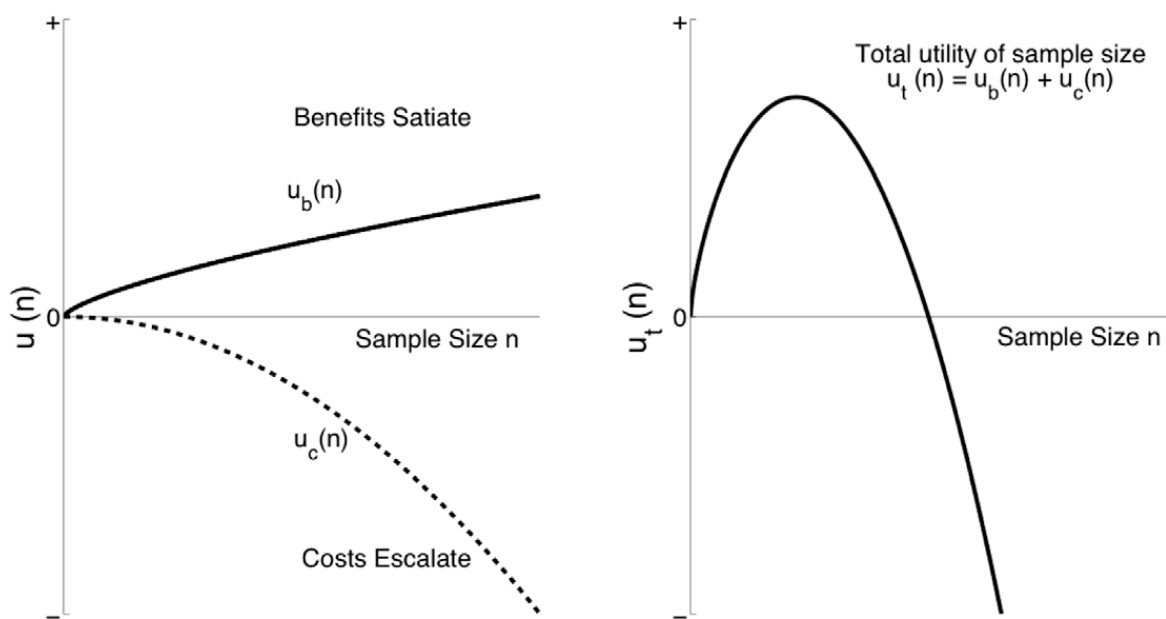


Fig. 3. Good things satiate and bad things escalate: The single-peakedness in an exploration domain (n = sample size).

ders the choice easier, albeit pointing to the option that may be inferior.

8.3. Some potential benefits of small samples

Only recently have cognitive psychologists and cognitive ecologists begun to explain frugal information search in terms of potential benefits of small samples. Traditional accounts of why people rely on small samples 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. Our results suggest that small samples make choice easier—although not more accurate (though consider Vul, Goodman, Griffiths and Tenenbaum's, 2009, argument that locally suboptimal decisions based on few samples are globally optimal)—and contribute to a growing appreciation of the various benefits of small samples. Some of the benefits, however, are less disputed than others. One advantage is, for example, 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.

Investigating the foraging behavior of bees across different floral reward distributions, Real (1992) 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 optimization may be adaptive when there is a high degree of spatial autocorrelation in the distribution of floral resources. 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. S135).

In cognitive 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 $\geq .5$) is close to Miller's (1956) estimate of the limited capacity of working memory. The detailed argu-

ment runs as follows: To determine whether two variables covary—for example, whether fleeing behavior triggers a predator's chase behavior—one typically relies on data sampled from one's environment (and, perhaps, prior expectations). If the assessment of a covariation has to be made “on the fly,” the limited capacity of working memory imposes an upper bound on the size of the information sample that can be considered at one time. Kareev (2000) then showed that samples of size 7 ± 2 increase the likelihood of encountering a sample that indicates a stronger correlation than that of the population. Thus, the limited working memory functions as an amplifier of correlations, leading to biased estimates. According to Kareev (2000, p. 401):

a biased estimate may better serve the functioning of the organism than an unbiased one. By providing such a biased picture, capacity limitation, may have evolved so as to protect organisms from missing strong correlations and to help them handle the daunting tasks of induction.

This conjecture of the evolutionary advantage of small samples (or, more precisely, a limited working memory) has fueled a controversial debate (Anderson, Doherty, Berg, & Friedrich, 2005; Juslin, Fiedler, & Chater, 2006; Juslin & Olsson, 2005; Kareev, 2005). Juslin and Olsson, for example, stressed that small samples not only amplify existing contingencies (i.e., hits) but also advocate nonexistent contingencies (i.e., false alarms). Moreover, if one takes both hits and false alarms into account, the posterior probability (based on a Bayesian analysis) that reality proves consistent with a sample contingency will always be higher for large than for small samples. In response, Fiedler and Kareev (2006) demonstrated that if the decision maker's aim is to select the better of two options rather than estimating the precise size of a contingency in a sample, and if this choice is only made when the evidence in the sample exceeds some decision threshold, then performance (defined in terms of hits and false alarms) can indeed decrease with increasing sample size.

9. Conclusions

People and bees both appear to rely on relatively small samples in their experience-based choices. Although our analysis regarding the effects of small samples does not confer an evolutionary advantage onto small samples, it suggests that an organism's economical search can at least partly be understood in terms of the statistical structures of an organism's niche. We showed that in an environment of gambles, drawing small samples from payoff distributions leads to experienced differences of sample means that are larger than the objective difference (Hertwig & Pleskac, 2008; Fig. 1). Moreover, for humans such biased differences make the choice between payoff distributions simpler. Our analysis also showed that although small samples increase the chance of choosing the lower-valued gamble, choices in our environment would not be all that bad. With as few as seven observations drawn from each gamble people would choose the higher expected value

gamble about 81% of the time. Moreover, accurate knowledge derived from larger samples yields surprisingly modest gains in terms of the probability of selecting the higher-value distribution (diminishing return of the value of information; Fig. 2). Together, the increase in difficulty with larger samples as well as other costs and the increase in accuracy at a diminishing rate help explain the frugal sampling behavior of people. More generally, we believe that our analysis calls for investigations of risky decision making to consider not only the cognitive processes underlying choice, but also how those processes interact with the ecological and statistical structures of risky environments (see also Hau et al., 2010).

Acknowledgements

We thank J. Ryan Brunton for his help in data collection. We are grateful to Robin Hau, Robin Hogarth, Lael Schooler, and Peter Todd for their helpful comments. We also thank Laura Wiles for editing the manuscript, and the Swiss National Science Foundation for two Grants to the first author (100014-118283 and 100014-126558).

Appendix A. ⁶

If one 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 (see Hertwig & Pleskac, 2008): Setting $Y = SM_A - SM_B$, the expected value of Y , $E(Y)$, can be calculated as follows:

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

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:

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

Because $E(Y|Y < 0)$ is by definition negative (the expected value of Y given that Y is smaller than 0), the second term in Eq. (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. Several factors can moderate 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.

References

- 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.
- Berry, D., & Fristedt, B. (1985). *Bandit problems*. London: Chapman and Hall.
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409–432.
- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.
- Coombs, C. H., & Avrunin, G. S. (1977). Single-peaked functions and the theory of preference. *Psychological Review*, *84*, 216–230.
- Daston, L. J. (1988). *Classical probability in the enlightenment*. Princeton, NJ: Princeton University Press.
- Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, *112*, 912–931.
- Erev, I., Glozman, I., & Hertwig, R. (2008). What impacts the impact of rare events. *Journal of Risk and Uncertainty*, *36*, 153–177.
- 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, and Cognition*, *32*, 883–903.
- Fong, G. T., 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.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance. How probability changed science and everyday life*. Cambridge, England: Cambridge University Press.
- Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411–435.
- Harrison, G. W. (1992). Theory and misbehavior of first-price auctions: Reply. *American Economic Review*, *82*, 1426–1443.
- 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*, 46–68.
- 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.
- 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., & 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*, *31*, 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 Bayesian cognitive science* (pp. 209–235). Oxford, England: Oxford University Press.
- Johnson, T. R., Budescu, D. V., & Wallsten, T. S. (2001). Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *Journal of Behavioral Decision Making*, *14*, 123–140.
- 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.
- 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.
- Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition*, *56*, 263–269.

⁶ We thank Dr. Maarten Speekenbrink for suggesting this proof to us.

- 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*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.
- Link, S. W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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.
- 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. *The Journal of Political Economy*, *59*, 371–404.
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 983–1002.
- Rakow & Newell (2010). Degrees of uncertainty: An overview and framework for research on experience-based choice. *Journal of Behavioral Decision Making*, *23*, 1–14.
- 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 (Behavioral mechanisms in evolutionary ecology).
- Reutskaja, E., & Hogarth, R. M. (2009). Satisfaction in choice as a function of the number of alternatives: When “goods satiate”. *Psychology & Marketing*, *26*, 197–203.
- Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, *10*, 33–51.
- Shafir, S., & Roughgarden, J. (1998). Testing predictions of foraging theory for a sit-and-wait forager, *Anolis gingivinus*. *Behavioral Ecology*, *9*, 74–84.
- Shanteau, J., & Thomas, R. P. (2000). Fast and frugal heuristics: What about unfriendly environments? *Behavioral and Brain Sciences*, *23*, 762–763.
- Simon, H. A. (1983). *Reason in human affairs*. Stanford, CA: Stanford University Press.
- Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, *72*, 923–955.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: The MIT Press.
- Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted, when rare outcomes are experienced (rarely)? *Psychological Science*, *20*, 473–479.
- Urban, F. M. (1910). The method of constant stimuli and its generalizations. *Psychological Review*, *17*, 229–259.
- von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. New York: Cambridge University Press.
- Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? Optimal decisions from very few samples. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 66–72). Austin, TX: Cognitive Science Society.
- 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.