האוניברסיטה העברית בירושלים

THE HEBREW UNIVERSITY OF JERUSALEM

ADVANTAGES OF COGNITIVE LIMITATIONS

By

YAAKOV KAREEV

Discussion Paper # 611 June 2012

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Advantages of Cognitive Limitations

Yaakov Kareev

Center for the Study of Rationality

The Hebrew University of Jerusalem

Advantages of Cognitive Limitations[1]

Yaakov Kareev

Abstract

Being a product of evolutionary pressures, it would not be surprising to find that what seems to be a limitation of the cognitive system is actually a fine-tuned compromise between a set of competing needs. This thesis is demonstrated using the case of the limited capacity of short-term memory, which is often regarded as the prime example of a cognitive limitation.

Short-term memory (STM) can hold only a small number of items; originally estimated at 7±2 items, it is now believed to be closer to 4 or 5. Because STM is the part of the cognitive system that holds the information available for conscious processing, its capacity sets an upper limit on the size of the sample that may be considered to determine characteristics of the environment. Obviously, the smaller that size, the higher the risk of obtaining inaccurate estimates of important parameters. However, at the same time, it can be shown that the very same limitation also carries with it a number of advantages.

First, small samples lead to systematic, and arguably beneficial, biases in the estimation of parameters that bear on the perception of regularity: Correlations are likely to appear stronger than they actually are and variances (i.e., risks) smaller than they

[1] This paper is a chapter to appear in Hammerstein, P. and J. R. Stevens, eds. 2012. Evolution and the Mechanisms of Decision Making. Strüngmann Forum Reports, vol. 11. Cambridge, MA: MIT Press (due out October, 2012). Work reported in this chapter was supported by Israel Science Foundation grants 539/2007 and 121/2011.

actually are. People have been found not to correct for these statistical biases, further implying that it is beneficial to perceive the environment as more regular than it actually is. Second, because decisions based on small samples are bound to result in occasional judgment errors and inconsistent behavior, the somewhat erratic behavior that inevitably results from the use of small-sample, error-prone data can benefit the organism in situations in which it is to the organism's advantage to behave in an unpredictable manner (e.g., prey-predator interactions, repeated choices between service providers, other situations in which a mixed strategy is called for and auto-correlation between successive decisions is undesirable). Third, by relying on only a limited number of recent experiences, the organism is better attuned to changes in its environment.

It is therefore maintained that what might look upon first sight as a major cognitive limitation, actually represents a fine-tuned compromise between a set of competing needs.

Introduction

Defining a characteristic of an information processing system to be a limitation is a tricky business. On one hand, one could easily propose how essentially any characteristic of the system could be improved: having three or four eyes would give humans a more complete picture of their immediate surroundings, faster neuronal transmission would speed up reactions, etc. In this respect, the current value of any characteristic of an information processing system might be regarded, almost by definition, as a limitation. On the other, increasing the amount of information available to the system—whether by way of the amount of input transmitted by a sensory organ, the amount of output resulting from internal calculations, or just the speed with which the same amount of information is made available—may require system-wide changes, such as increased storage capacity or increased processing speed, which may require compromising some other aspects of performance. As a result of such considerations, many characteristics of the human information processing system are not considered limitations: neither having only two eyes nor the speed of transmission through the nervous system is regarded as a limitation.

Here I will focus on what it takes for a characteristic of an organism's information processing system to be considered a cognitive limitation. No rules for this enterprise exist, of course, but I would like to propose a few. First, for a characteristic to qualify as a cognitive limitation, its value should impose a clear cost in terms of functioning relative to the level of functioning that could be achieved with other, more desirable (and potentially reachable) values of the same characteristic. Second, there should be some variability—inter-species, intra-species, or both—in the values of the characteristic, and there should be at least one way of moving from the undesirable range of values to a more desirable range, such as through actions by the individual (exertion of effort, learning), situational factors (education, changes in the environment), or easily envisioned evolutionary pressures.

Of the characteristics that are commonly considered limitations of the cognitive system (e.g., processing speed, computational complexity or lack thereof, and memory capacity), none looms larger than the limited capacity of short-term memory (STM). This limit, reflecting a structural characteristic of the cognitive system, was first pointed out by G. A. Miller (1956) in his classical study “The Magic Number Seven, Plus or Minus Two.” There are numerous sophisticated measures of STM capacity (e.g., Engle et al. 1999; Just and Carpenter 1992), but the uninitiated can easily get a feel for this limit by considering the digit-span task, which is a component of the Wechsler Intelligence Scale for Children (Wechsler 1949). In this task, a tester reads a list of single digits at a rate of one digit per second and instructs the test taker to repeat the list in exactly the order it was read. The first list is short and easy to repeat (e.g., “3, 8, 6”). Another, longer list (e.g., “4, 5, 9, 2”) follows a correct repetition. Increasingly longer lists are read until two successive failures occur. The number of digits in the longest list correctly remembered serves as the estimate of the test taker's STM capacity. Analyzing people’s performance in a number of converging tasks, Miller estimated the STM capacity of normal adults to be around seven items, ranging from slightly less to slightly more depending on the characteristics of the individual and the specific task. Miller’s original observation generated enormous interest and follow-up research, and has become a cornerstone of cognitive psychology (Simon 1990). Although estimates of STM capacity vary—in fact, modern estimates place its value at four or five items (Cowan 2001)—it is uncontested that STM capacity is limited to a small number of items.

Before getting further into a discussion of whether or not a limited STM capacity constitutes a cognitive limitation, it should be pointed out that, by imposing a relatively rigid constraint on information processing, the limited capacity of STM renders information processing quite robust: An important aspect, the amount of information likely to be considered when a judgment, decision, or choice is being made is unlikely to be much affected by extraneous factors such as context or (to some extent) even the ability or mood of the decision maker.

Is STM Capacity a Cognitive Limitation?

The limited capacity of STM has obvious implications when one considers that STM is regarded to be the workspace of the cognitive system. Only items in it—whether selected from sensory input or retrieved from long-term memory—are actively available for processing. Once STM is full to capacity, any new information added necessarily pushes some information out. Therefore, a clear implication of this capacity limit is that one can simultaneously consider only a small number of items—whether they are observed in the outside world, retrieved from memory, or both. This, in turn, implies that the cognitive system can act on only a small sample of the information available at any time. Because the variance of the sampling distribution of any statistic is larger the smaller the sample, small-sample data are likely to result in inaccurate estimates, and give rise to disagreements between different people observing the same phenomena. The built-in inaccuracy that comes with small samples is orthogonal to the lack of bias in estimates of certain population parameters (more on that later).

If we assume that having accurate estimates of our environment is essential for efficient functioning, the limited capacity of STM appears to be a major handicap, one that increases the chances of erroneous estimates, which would lead to sub-optimal decisions and actions. Furthermore, it is also a limitation by the other criterion I suggested earlier: First, there is natural variability in built-in STM capacity—more intelligent people have, on average, larger STM capacity than less intelligent people (e.g., de Jong and Das-Small 1995; Engle 2002; Jurden 1995), adults have larger capacity than children (e.g., Case et al. 1982; Huttenlocher and Burke 1976)—and STM capacity also varies with situational factors, such as cognitive load (Gilbert et al. 1988) or emotional stress (e.g., Diamond et al. 1996). Second, there are ways to reduce the impact of the limitation: Although the number of items in STM is limited, the items themselves are chunks—units that have a reference in long-term memory. For example, the sequence G O L D B A R, read a letter at a time, may constitute 7 one-letter chunks for a person whose knowledge of English is limited to the names of the letters, but only two chunks (or just one) for people familiar with English words. Thus, whereas the number of chunks is limited, the amount of information per chunk can vary, depending on a person’s knowledge. This point was demonstrated in studies comparing laypersons' and chess masters’ memory for briefly presented board positions (Chase and Simon 1973; De Groot 1965).

To sum up, by the criteria suggested earlier, the limited capacity of STM is definitely a cognitive limitation: Its effects are costly, it has a range of values that vary in the severity of their implications for functioning, and there are well-defined processes that affect the value of this characteristic.

However, although it is clear and uncontestable that the limited capacity of STM has some detrimental effects on performance, I will argue that the very limited capacity of STM, and the use of small-sample data that it imposes, has some effects that not only are not detrimental, but may in fact be outright beneficial. Some of these effects are obvious, and a brief discussion of them may suffice, while others are more subtle and require a more elaborate discussion.

Saving Search Cost and Time

It is a truism that the expected accuracy of an estimate increases with an increase in sample size. It is therefore generally believed that better decisions are reached on the basis of a larger sample. However, making do with small samples saves search time and effort. Saving on search time translates into earlier choice and more time for enjoying the benefits of that choice. Such savings are even more important when a decision is taken in the context of a race for the possession or right of use of a unique, indivisible good (e.g., buying or renting a house, buying a used car, choosing a mate). In such cases, competitors who make do with less information, who spend less time on search and accept the risk of having made a non-optimal decision, may win possession or right of use of what would have been one's eventual choice after collecting all the data necessary to reach a well-justified decision. Thus, shorter search and decision times are obvious advantages of the constraints imposed by the limited capacity of STM. Although the benefits in this respect are difficult to assess, it is clear that they mitigate the costs incurred by the inaccuracy of the estimates themselves.

Furthermore, whereas search time and costs usually go up linearly with sample size, the expected accuracy of estimates increases linearly only with the square root of the size of the sample. As a result, the benefits of extra items diminish quickly. Indeed, analyses have shown that in judgment and choice tasks, accuracy quickly approaches an asymptote with a surprisingly small number of sample items (Hertwig and Pleskac 2010; Hertwig and Todd 2003; Johnson et al. 2001). Other indications that good performance can be built on relatively little data come from work on the effect of observing the behavior of a small number of neighbors and imitating the behavior of a successful neighbor (Aktipis 2004; Morales 2002; Nowak and May 1992; Schlag 1999). Work on the implications of adopting the "Win—Stay, Lose—Shift" decision rule, a rule that bases decisions on the most recent outcome only (e.g., Nowak and Sigmund 1993; Posch 1999; Posch et al. 1999), also demonstrates that high levels of performance may be achieved while relying on little data. Other analyses of performance with relatively little data (Kandori et al. 1993; Sandholm 1998) also point to the fact that using small amounts of data is not very costly, if indeed it is costly at all.[2]

Beneficial Effects of Biased Estimates

Although sample statistics provide unbiased estimates of a population mean (and hence, also the difference between two means), they provide biased estimates of other population parameters—most notably variance and correlation. Moreover, for these latter parameters, the degree of bias is negatively related to sample size: The smaller the sample, the larger the bias. These biases are a statistical fact. Unlike the biases so much studied in the behavioral sciences (e.g., Kahneman et al. 1982; Gilovich et al. 2002), which are assumed to result from the biased processing of unbiased input, the biases that I am referring to here apply when the data are well sampled and the processing unbiased: By using sample data, one ends up seeing a systematically distorted picture of the world.

---

[2]In this respect, it is also appropriate to mention work that demonstrates, theoretically and empirically, that simple decision rules can lead to good performance. These studies include the initial work within the heuristics and biases framework, particularly work on the availability heuristic (for summaries of early studies, see (for summaries of early studies, see Kahneman et al. 1982; Tversky and Kahneman 1974), its later emphasis on biases notwithstanding; work in the tradition of the fast and frugal heuristics (Gigerenzer et al. 1999); work on simple learning rules in complex environments (Houston et al. 1982; McNamara and Houston 1985, 1987); and work on a simple heuristic that leads to regret minimization (Hart 2005; Hart and Mas-Colell 2000). Mention of this work has been relegated to a footnote because many of these simple rules or heuristics still rely on the accumulation of much data.

The two parameters for which sample statistics provide a biased estimate are the variance of a distribution and the correlation between variables. Both involve aspects of the regularity of the environment, and if their bias is not corrected for, the world is perceived as more regular than it really is.[3] I argue that the biases these statistics introduce to one's picture of the world are likely to be beneficial.

Assessment of Variance

The variability, or heterogeneity, of a distribution—measured by the second moment around the mean[4]—is of great import for the organism: It provides an indication, for example, of the error (squared deviation) that is expected when predicting the mean, the risk that is assumed when taking an action, and the variety of strategies that may be called upon when having to deal individually with all items in a distribution (e.g., an instructor of a class of students).

It is a well-known statistical fact, although one whose behavioral implications have hardly been considered, that sample variance is an attenuated estimator of the variance of the population from which that sample has been drawn. The attenuation, very familiar to practitioners of the t-test, is by a factor of $(N–1)/N$, where N denotes the sample size. To obtain an unbiased estimate of the population variance, σ^2_x, one needs to apply the following correction: $\text{est}(\sigma^2_x)=(N/(N–1))\, S^2_x$, where S^2_x is the sample variance. The degree of the bias is small when N is large, but substantial for the size of the sample

[3]Interestingly, as long ago as 1620, Francis Bacon, in describing the "idols"—the bad habits of mind that cause people to fall into error—noted that people expect more order in natural phenomena than actually exists: "The human understanding is of its own nature prone to suppose the existence of more order and regularity than it finds" (Bacon 1620/1905:265).

[4]In using this definition, I assume that variables are measured on interval or ratio scales, and that variance is the measure of variability. My arguments, however, also apply for variables measured on ordinal and nominal scales, or for measures of variability other than variance.

imposed by the limited capacity of STM: If one literally accepts the classical estimate of STM capacity, then adults observe, on average, a world that is only 6/7, or 86% as variable as it really is. Another way to consider this effect is to note that, for samples drawn from a normal distribution and consisting of 7 items, the estimated variance will be lower than the population variance 65% of the time. Thus, if uncorrected, small-sample-based estimates of variability are likely to be lower than the true variability—to project a picture of a world that is more homogeneous, less risky, than it actually is.

Of course, subjective perception need not reflect the (distorted) picture observed. The cognitive system could have evolved, or could learn, to apply some correction to estimates of the variability observed in samples, much as statisticians correct their estimates of variability. However, empirical studies addressing this question (Kareev et al. 2002) indicate that this is not the case: These studies compared performance in consequential decision-making tasks between people who had, or were likely to have at their disposal, samples differing in their size: In one study people with small STM capacity were compared with people with large STM capacity; in another people presented with small samples were compared to people presented with large samples; in a third study people who had a whole population in full view were compared to people who saw only a sample of the population. In all cases, people who saw, or were likely to rely on smaller samples behaved as if the variability of the population in question was smaller. Such results would not have been observed had people corrected for the bias introduced by the use of small samples (or if they had been altogether inattentive to variability).

One implication of these findings is that people tend to underestimate risk and errors of prediction, and consequently feel more in control of their environment than they actually are. The bias, which affects everyone, will be more pronounced among the less intelligent, children, and people under stress than among the more intelligent, adults, and people who are relaxed. The bias also implies that individuals will perceive others as more homogeneous than themselves, and out-groups as more homogeneous than in-groups (Linville et al. 1989).

Would such a bias benefit or harm the individual? It is difficult to answer this question. On the one hand, it is possible to claim that any bias is dysfunctional. On the other hand, consider the fact that feelings of control contribute to organisms' physical and psychological well-being (Alloy and Tabachnik 1984; Langer 1975; Seligman 1975) or the claim that "cautious optimism" is a key to the risk-taking behavior and successful entrepreneurship that are so important for society (Selten 2001). It may be the case that the distorted view resulting from the use of small-sample data to assess variability provides just the right amount of distortion.

Detection of Correlation

The detection of correlations is one of the more important undertakings of the cognitive system. It underlies classical and operant conditioning, it promotes the emergence of concepts and provides the basis for learning about causal relationships. When correlations exist and are detected, they help understand the past, control the present, and predict the future (Alloy and Tabachnik 1984; Shanks 1995). Organisms could hardly be expected to negotiate the vast complexity of tasks facing them if they did not detect and utilize the relationships existing in their environment. What relationship could one expect to observe when drawing a sample of bivariate observations and calculating the correlation between them? As is the case with other statistics, the sampling distribution is more variable the smaller the size of the sample on which it is based. Thus, when variables are actually unrelated, the use of small samples could well result in a false alarm—the "detection" of a correlation that does not exist. However, a very different picture emerges when one considers the sampling distributions of non-zero correlations and tries to infer relationships between variables that are in fact related. It is a well-established, though little known, fact that the sampling distribution of non-zero correlations is skewed, and increasingly so the smaller the samples on which they are based (e.g., Hays 1963). Figure 10.1, based on the tables compiled by David (1954), presents the sampling distributions for a population correlation, ρ, of .50 for sample sizes of 4, 7, and 10. The skew is striking: The medians—at .59, .54, and .52, for the three sample sizes, respectively—and even more so the modes of the sampling distributions are more extreme than the true values, and the bias is stronger the smaller the sample size. To be sure, the means of the sampling distributions are all slightly less extreme than the population values, and the more so the smaller the sample. But, because initial impressions not only carry greater weight than later impressions (Hogarth and Einhorn 1992), but also often bias the interpretation of further information (Nickerson 1998), the fact that small-sample data, more often than not, produce estimates that are more extreme than the population parameters, but deviate in the right direction, is of great import. It points to the possibility that the limited capacity of STM acts as an amplifier of correlations (Kareev 1995).[5]

[5]The bias discussed is larger, the more extreme the true correlation.

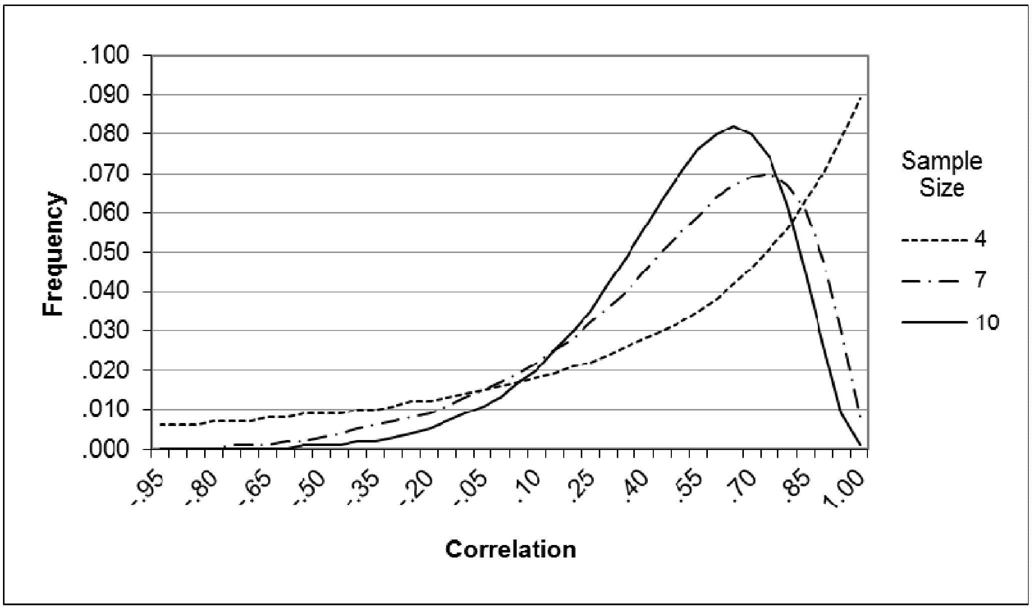

Figure 10.1 The sampling distributions of the correlation coefficient, for ρ=.50, and for sample sizes of 4, 7, and 10.

Empirical research (Kareev et al. 1997) lent support to this hypothesis, and further theoretical analyses (Kareev 2000) showed samples in the range of 7±2 items—the very range originally identified by Miller (1956)—to be of particular import in the detection of useful relationships between binary variables. These results point to the intriguing possibility that the limited capacity of STM in fact facilitates the detection of important relationships. The premise that organisms are intent on detecting effects when they exist (i.e., avoiding misses), even at the cost of making false alarms, sits well with recent findings that show people invest misses with greater value than false alarms (Kareev and Trope 2011; Wallsten et al. 1999; Wallsten and González-Vallejo 1994) and "design" their sampling procedures in ways that would increase the power of a statistical test, at the cost of observing inaccurate, inflated sample values that they do not correct (Kareev and Fiedler 2006; Soffer and Kareev 2011).[6] Theoretical analyses as well as experimental findings also support the notion that in reaching decisions, people rely on a small number of instances retrieved from memory (e.g., Selten and Chmura 2008; Stewart et al. 2006). The claims about the beneficial effects of small samples with respect to the detection of correlations also sit well with earlier accounts of the benefits of small capacity for perceptual (Turkewitz and Kenny 1982) and language (Elman 1993; Newport 1988, 1990) development.[7]

Being Less Predictable

As has been pointed out time and again, when sample size is small, sample statistics – hence, estimates of population parameters—are likely to vary. To an outside observer, behavior based on repeated small samples may look erratic and unpredictable. In previous sections of this paper, I assumed that it is to the organism's benefit to obtain accurate estimates; in the present section, I point out that, in many situations—situations that involve interactions between agents whose interests do not completely match—being less than perfectly predictable may in fact be in the organism's best interest. In such situations, the interests of the individual may be well served by the unavoidable reliance on small, error-prone estimates.

[6]With respect to the analysis put forward by Error Management Theory (Haselton & Buss, 2000; Haselton & Nettle, 2006; Nettle, in press), an implication of the limited capacity of STM is that over the whole range of decisions, Misses (False Negatives) are more costly that False Alarms (False Positives). Interestingly, in all the examples used by Nettle (in press) – examples apparently capturing natural and important decision-making scenarios – this is indeed the case.

[7]It should be noted that the claims about the correlation-amplifying effect of small samples have been challenged (Anderson et al. 2005; Gaissmaier et al. 2006; Juslin and Olsson 2005), but also that these challenges have been responded to (Kareev 2005).

Many interactions are of the pursuit-avoidance type, in which one agent wishes to catch another agent, whereas the latter wishes to avoid the former. Predators and prey, police and criminals, players of matching pennies (and of other games modeled by it)—all engage in such interactions. The prescription in such situations is to behave in a random, unpredictable manner—to employ, so to speak, a mixed strategy, or to engage in Protean defense (Driver and Humphries 1988; Humphries and Driver 1967, 1970). Use of small-sample data renders one's behavior noisier, and thus more difficult to characterize and predict. In this respect, the limited capacity of STM endows behavior automatically with a desirable characteristic.

Erratic, unpredictable behavior is of even greater value in another important class of situations—those involving choices between adaptive agents who are eager to be selected by the decision maker. Examples of such agents include service providers, who are selected on the basis of the quality of their service, and employees who are competing for a bonus to be awarded on the basis of their productivity, as well as flowers, which are selected by pollinators on the basis of the quality and quantity of their nectar. In all cases, the decision maker's main interest is to maximize his or her utility—to obtain the best service, the highest production, or the most nectar. Which monitoring regime should one employ to achieve that goal? A straightforward suggestion would be to employ full scrutiny: to observe each option numerous times, in order to obtain an accurate estimate of its value, and then choose the option with the highest value. However, when it comes to choices between adaptive agents that are eager to be selected by the decision maker, it can be argued that larger samples do not necessarily lead to better decisions. To see why, consider the owner of a small shoe factory, who has two employees producing shoes.

Wishing to increase production, the owner decides to offer a monthly bonus, to be awarded to that employee who produced more shoes during the month. How should production be assessed? Should the owner take stock of each day's production and award the bonus on the basis of the monthly sum total? What effect would such a monitoring scheme have on the workers' motivation? Given that the workers are likely familiar with each other's productivity, it could turn out that neither of them would exert greater effort: The less productive worker would see no chance of outperforming the more productive one and would continue to produce at the previous level; the more productive worker would be likely to realize that and therefore not exert higher effort either. With such a scenario, the owner would indeed award the bonus to the truly deserving worker, but might end up paying the extra amount of the bonus without reaping any benefit. Now consider an alternative monitoring regime—minimal scrutiny. Suppose the owner announces that production will be assessed on the basis of one day's production, with the crucial day chosen at random, separately for each worker, at the end of the month. With such a regime, if the distributions of daily production of the two workers overlap, the less productive worker does stand a chance of occasionally winning the bonus. Realizing that, he or she may work harder, to increase the overlap between the two distributions. Sensing the danger, the more productive worker may also try harder, to maintain or even reduce the overlap between the two distributions. With such a monitoring scheme, the principal will occasionally (but infrequently) commit an error, awarding the bonus to the less deserving worker, but may also bring about a more desirable situation.

A formal analysis of similar situations has been advanced within the framework of labor markets and tournament theory (Cowen and Glazer 1996; Dubey and Haimanko 2003;

Dubey and Wu 2001). In all these studies it has been concluded that less, or even minimal, scrutiny can result in an outcome that is more desirable from the viewpoint of the decision maker because of the very uncertainty that a lower level of scrutiny introduces into the resolution process. An experimental test of the viability of this line of reasoning (Kareev and Avrahami 2007) demonstrated that in a two-person competition involving the performance of a routine, simple task, minimal scrutiny indeed resulted in higher performance; the effect depended on the existence of a bridgeable initial gap between competitors and showed up in competitive situations only.

Increased productivity is not the only possible benefit of minimal scrutiny. As pointed out by Kareev and Avrahami (2007), by leading to the occasional reward of a weaker competitor, minimal scrutiny also helps maintain competition and sustain diversity. In contrast, the consistent choice of a currently superior option may eliminate weaker competitors altogether, eventually leaving the decision maker at the mercy of a monopoly or a provider with a limited pool of characteristics.

Thus, by necessarily using small-sample data and therefore being susceptible to error, decision makers signal that they are error prone, or even “irrational”—that knowledge of the objective, prevailing conditions does not necessarily lead to successful predictions of their behavior. It is as if the decision makers announce, “I am not, I cannot be, fully rational. Take note of that and now act as you please.” The notion that irrationality can breed rational behavior and desirable outcomes sits well with work in game theory on equilibrium refinements (Myerson 1978; Selten 1975; see also the related work of Kreps et al. 1982). That work “indicates that rationality in games depends critically on irrationality. In one way or another, all refinements work by assuming that irrationality cannot be ruled out, that the players ascribe irrationality to each other with a small probability. True rationality needs a ‘noisy,’ irrational environment; it cannot grow in sterile soil” (Aumann and Sorin 1989:37–38). Game-theoretic work on the role of bounded recall in fostering cooperation (Aumann and Sorin 1989) demonstrates another benefit of limited capacity. Thus, a cognitive limitation that introduces a grain of inconsistency or irrationality into behavior may act to bring about an environment that is more beneficial to the constrained decision maker.

Detection of Change

Thus far, I have assumed a static environment. In such an environment, the inaccuracy that necessarily results from limited capacity may still be viewed as a drawback, even in view of the benefits I have discussed. However, environments are hardly ever static; with the advent of time, many characteristics change in value: Whether constituting abrupt changes in quality—an inferior service provider becomes superior, a foe becomes a friend, a cooperator starts defecting—or gradual, periodic shifts—the yield of one type of flowers diminishes while that of another increases—changes abound. To function properly, to make sound decisions that respond to current conditions, organisms must continuously monitor the environment and act in a way that is responsive to changes in absolute and relative values (Cohen et al. 2007; Gross et al. 2008; Speekenbrink and Shanks 2010). It is obvious that using numerous observations and averaging over a large number of items may provide an inaccurate estimate of the current situation. As a result, the detection of change would be slow. Indeed, it has long been recognized by students of behavior that in changing environments, remembering too many past events is counterproductive (McNamara and Houston 1985, 1987; Rakow and Miler 2009; Shafir and Roughgarden 1994). Furthermore, both human (e.g., Massey and Wu 2005) and animal (Gallistel et al. 2001; Shettleworth et al. 1988) studies of behavior in changing environments reveal fast adaptation to change. The adaptation is so fast, in fact, that Gallistel et al. (2001) explicitly state that it poses a challenge to any learning theory. In contrast to learning theories, which postulate that behavior reflects response propensities accumulated over all past experience, models of behavior that rely on small samples, observing the environment through a narrow window of the most recent events, monitor the environment continuously and adjust to change automatically. In fact, the detection of change is faster the narrower the window. To illustrate that last point, Figure 10.2a–d presents the predictions derived from two models of choice behavior. Both models—"Win—Stay, Lose—Shift" (e.g., Nowak and Sigmund 1993; Posch 1999; Posch et al. 1999) and "No-Regret—Stay, Regret—Shift" (Kareev, Avrahami, & Fiedler, in preparation)—take into account the last trial only. The graphs show how choice probabilities for two or three options change over time, both before and after a change in value is introduced. The important point is that although the two models employ different dynamics, and neither involves learning, both predict smooth changes with time and quick adaptation to change. Here, too, it is the apparent limitation which insures efficient and robust performance.

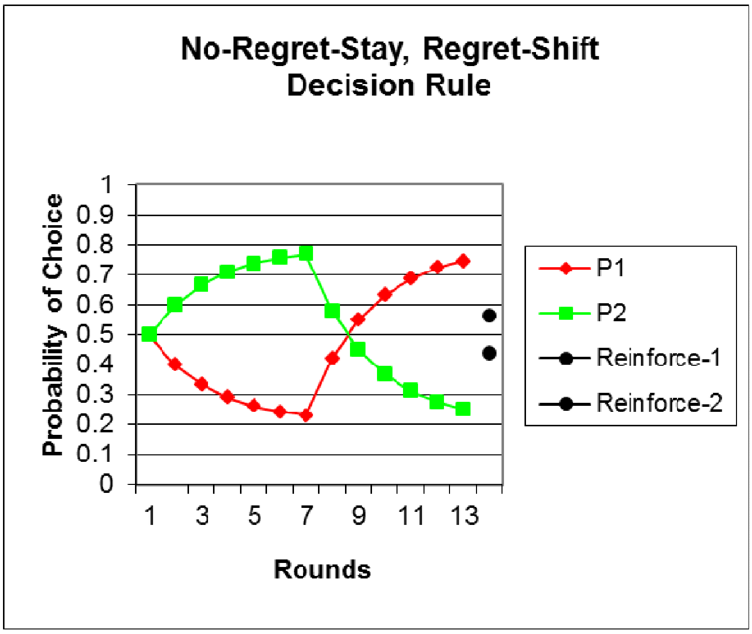

Figure 2a

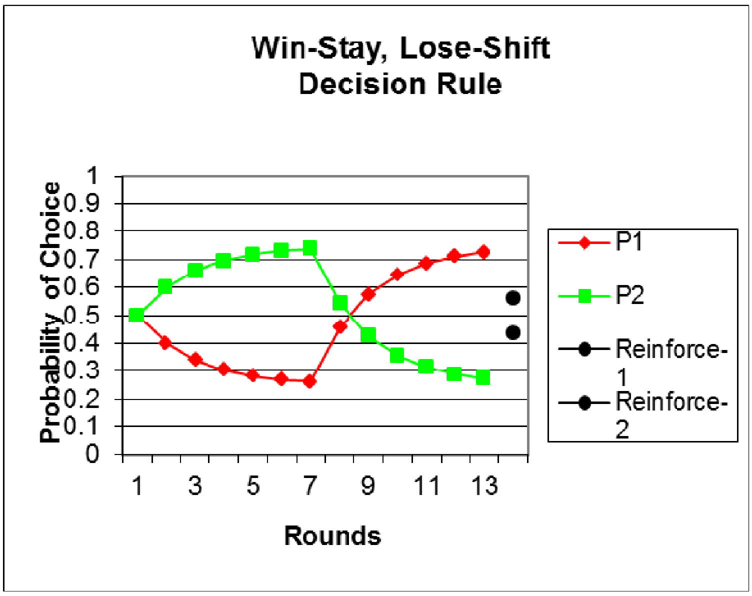

Figure 2b

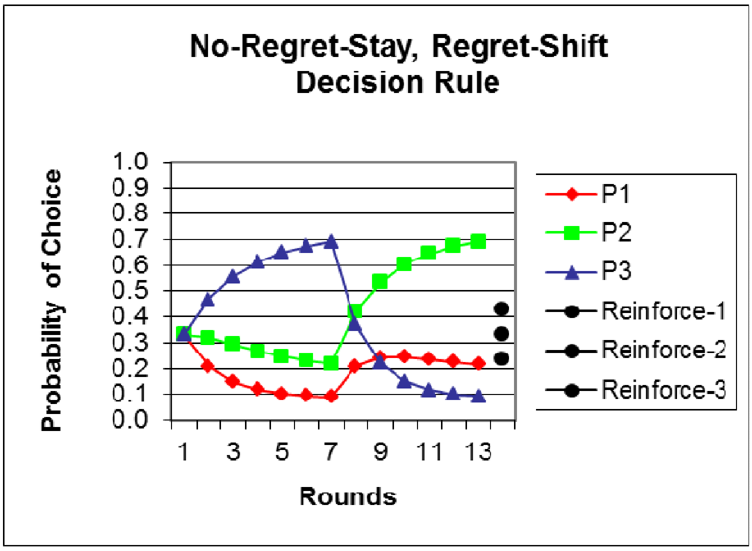

Figure 2c

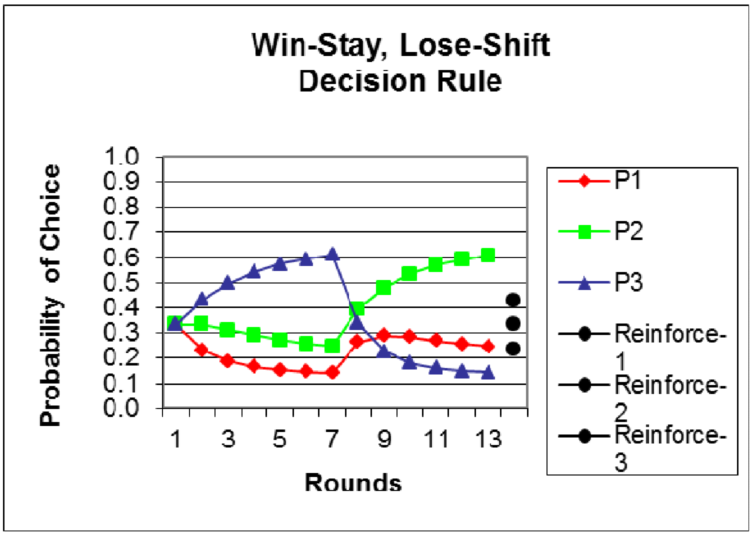

Figure 2d

Figure 10.2 Choice probabilities over time, as predicted by the "No-Regret—Stay, Regret—Shift," and the "Win—Stay, Lose—Shift" models. Choices are either between two options, whose probabilities of providing a reward are .7 and .9 (a and b), or between three options, whose probabilities of reward are .5, .7, and .9 (c and d). In all cases, initial choices are assumed random (no prior knowledge). The probabilities of reward change on trial 8. The reinforcement values, represented by black dots on the right, are asymptotic values following the matching law. Results are analytical, not simulations.

Thus, with respect to the detection of change, relying on a small sample of recent events confers a real, uncontestable advantage. One may even speculate that a species' STM capacity has been shaped, at least in part, in response to the amount of change that the species faces in its environment, with more change resulting in smaller capacity.

Conclusion

In conclusion, the argument put forward here is that, far from the being the cognitive limitation it may seem at first, the limited capacity of Short-Term Memory is in fact the result of multiple tradeoffs, and represents a compromise between the values that would be optimal for competing needs.

References

Aktipis, C. A. 2004. Know when to walk away: Contingent movement and the evolution of cooperation. *J. Theor. Biol.* **231**:249–260. [Kareev]

Alloy, L. B., and N. Tabachnik. 1984. Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychol. Rev.* **91**:112–149. [Kareev]

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

Aumann, R. J., and S. Sorin. 1989. Cooperation and bounded recall. *Games Econ. Behav.* **1**:5–39. [Brown, Kareev]

Bacon, F. 1620/1905. Novum Organum, trans. R. L. Ellis and J. Spedding, ed. J. M. Robertson. London: George Routledge and Sons. [Kareev]

Case, R., D. M. Kurland, and J. Goldberg. 1982. Operational efficiency and the growth of short-term memory span. *J. Exp. Child Psychol.* **33**:386–404. [Kareev]

Chase, W. G., and H. A. Simon. 1973. Perception in chess. *Cogn. Psychol.* **4**:55–81. [Kareev]

Cohen, J. D., S. M. McClure, and A. J. Yu. 2007. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Phil. Trans. R. Soc. B* **362(1481)**:933. [Adams et al., Kareev]

Cowan, N. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* **24**:87–114. [Kareev]

Cowen, T., and A. Glazer. 1996. More monitoring can induce less effort. *J. Econ. Behav. Org.* **30**:113–123. [Kareev]

David, F. N. 1954. Tables of the Ordinates and Probabiliy Integral of the Distribution of the Correlation Coefficient in Small Samples Issued by the Biometrika Office University College, London. Cambridge: Cambridge Univ. Press. [Kareev]

De Groot, A. D. 1965. Thought and Choice in Chess. The Hague: Mouton Publishers. [Kareev] de Jong, P. F., and E. A. Das-Small. 1995. Attention and intelligence: The validity of the Star Counting Test. *J. Educ. Psychol.* **87**:80–92. [Kareev]

Diamond, D. M., M. Fleshner, N. Ingersoll, and G. Rose. 1996. Psychological stress impairs spatial working memory: Relevance to electrophysiological studies of hippocampal function. *Behav. Neurosci.* **110**:661–672. [Kareev]

Driver, P. M., and D. A. Humphries. 1988. Protean Behavior: The Biology of Unpredictability. Oxford: Oxford Univ. Press. [Kareev]

Dubey, P., and O. Haimanko. 2003. Optimal scrutiny in multi-period promotion tournaments. *Games Econ. Behav.* **42**:1–24. [Kareev]

Dubey, P., and C. Wu. 2001. Competitive prizes: When less scrutiny induces more effort. *J. Math. Econ.* **36**:311–336. [Kareev]

Elman, J. L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* **48**:71–99. [Kareev]

Engle, R. W. 2002. Working memory capacity as executive attention. *Curr. Dir. Psychol. Sci.* **11**:19–23. [Kareev]

Engle, R. W., S. W. Tuholski, J. E. Laughlin, and A. R. A. Conway. 1999. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *J. Exp. Psychol. Gen.* **128**:309–331. [Kareev]

Gaissmaier, W., L. Schooler, and J. Rieskamp. 2006. Simple predictions fueled by capacity limitations: When are they successful? *J. Exp. Psychol. Learn. Mem. Cogn.* **32**:966–982. [Kareev]

Gallistel, C. R., T. A. Mark, A. P. King, and P. Latham. 2001. The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *J. Exp. Psychol. Anim. Behav. Proc.* **27(4)**:354–372. [Gallistel, Kareev]

Gigerenzer, G., P. M. Todd, and the ABC Research Group, eds. 1999. Simple Heuristics That Make Us Smart. New York: Oxford Univ. Press. [2, 4, 10, 15]

Gilbert, D. T., B. W. Pelham, and D. S. Krull. 1988. On cognitive business: When person perceivers meet persons perceived. *J. Pers. Soc. Psychol.* **54**:733–740. [Kareev]

Gilovich, T., D. Griffin, and D. Kahneman, eds. 2002. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge: Cambridge Univ. Press. [Kareev]

Gross, R., A. I. Houston, E. J. Collins, et al. 2008. Simple learning rules to cope with changing environments. *J. R. Soc. Interface* **5**:1193–1202. [Kareev]

Haselton, M. G., and D. M. Buss. 2000. Error management theory: A new perspective on biases in cross-sex mind-reading. *J. Pers. Soc. Psychol.* **78**:81–91. [Kareev]

Haselton, M. G., and D. Nettle. 2006. The paranoid optimist: An integrative evolutionary model of cognitive biases. *Pers. Soc. Psychol. Bull.* **10**:47–66. [Kareev]

Hart, S. 2005. Adaptive heuristics. *Econometrica* **73**:1401–1430. [Kareev]

Hart, S., and A. Mas-Colell. 2000. A simple adaptive procedure leading to correlated equilibrium. *Econometrica* **68**:1127–1150. [Kareev]

Hays, W. L. 1963. Statistics for psychologists. New York: Holt Rinehart & Winston. [Kareev]

Hertwig, R., and T. J. Pleskac. 2010. Decisions from experience: Why small samples. *Cognition* **115**:225–237. [Kareev]

Hertwig, R., and P. M. Todd. 2003. More is not always better: The benefits of cognitive limits. In: Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making, ed. D. Hardman and L. Macchi, pp. 213–231. New York: John Wiley & Sons. [Kareev]

Hogarth, R. M., and H. J. Einhorn. 1992. Order effects in belief updating: The belief-adjustment model. *Cogn. Psychol.* **24**:1–55. [Kareev]

Houston, A. I., A. Kacelnik, and J. M. McNamara. 1982. Some learning rules for acquiring information. In: Functional Ontogeny, ed. D. J. McFarland, pp. 140–191. London: Pitman. [Kareev]

Humphries, D. A., and P. M. Driver. 1967. Erratic display as a device against predators. *Science* **156**:1767–1768. [Kareev]

———. 1970. Protean defence by prey animals. *Oecologia* **5**:285–302. [Kareev]

Huttenlocher, J., and D. Burke. 1976. Why does memory span increase with age? *Cogn. Psychol.* **8**:1–31. [Kareev]

Johnson, T. R., D. V. Budescu, and T. S. Wallsten. 2001. Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *J. Behav. Decis. Mak.* **14**:123–140. [Kareev]

Jurden, F. H. 1995. Individual differences in working memory and complex cognition. *J. Educ. Psychol.* **87**:93–102. [Kareev]

Juslin, P., and H. Olsson. 2005. Capacity limitations and the detection of correlations: Comment on Kareev (2000). *Psychol. Rev.* **112**:256–267. [Kareev]

Just, M. A., and P. A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychol. Rev.* **99**:122–149. [Kareev]

Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. Judgment under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press. [Chater, Dall, Kareev]

Kandori, M., G. T. Mailath, and R. Rob. 1993. Learning, mutation, and long run equilibria in games. *Econometrica* **61**:29–56. [Kareev]

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

———. 2000. Seven (indeed, plus or minus two) and the perception of correlation. *Psychol. Rev.* **107**:397–402. [Kareev]

———. 2005. And yet the small-sample effect does hold: Reply to Juslin and Olsson (2005) and Anderson, Doherty, Berg, and Friedrich (2005). *Psychol. Rev.* **112**:280–285. [Kareev]

Kareev, Y., S. Arnon, and R. Horwitz-Zeliger. 2002. On the misperception of variability. *J. Exp. Psychol. Gen.* **131**:287–297. [Kareev]

Kareev, Y., and J. Avrahami. 2007. Choosing between adaptive agents: Some unexpected implications of level of scrutiny. *Psychol. Sci.* **18**:636–641. [Kareev]

Kareev, Y., Avrahami, J., & Fiedler, K. (in prepartion). No-regret - Stay, Regret - Shift. Manuscript in preparation.

Kareev, Y., and K. Fiedler. 2006. Non-proportional sampling and the amplification of correlations. *Psychol. Sci.* **17**:715–720. [Kareev]

Kareev, Y., I. Lieberman, and M. Lev. 1997. Through a narrow window: Sample size and the perception of correlation. *J. Exp. Psychol. Gen.* **126**:278–287. [Kareev]

Kareev, Y., and Y. Trope. 2011. Correct acceptance weighs more than correct rejection: A decision bias induced by question framing. *Psychon. Bull. Rev.* **18**:103–109. [Kareev]

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson. 1982. Rational cooperation in the finitely repeated Prisoner's Dilemma. *J. Econ. Theory* **27**:245–252. [Kareev]

Langer, E. J. 1975. The illusion of control. *J. Pers. Soc. Psychol.* **32**:311–328. [Kareev]

Linville, P. W., P. Salovey, and G. W. Fischer. 1989. Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *J. Pers. Soc. Psychol.* **57**:165–188. [Kareev]

Massey, C., and G. Wu. 2005. Detecting regime shifts: The causes of under- and over-reaction. *Manag. Sci.* **51**:932–947. [Kareev]

McNamara, J. M., and A. I. Houston. 1985. Optimal foraging and learning. *J. Theor. Biol.* **117**:231–249. [Kareev]

———. 1987. Memory and the efficient use of information. *J. Theor. Biol.* **125**:385–395. [Kareev]

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity to process information. *Psychol. Rev.* **63**:81–97. [Kareev, Dayan]

Morales, A. J. 2002. Absolutely expedient imitative behavior. *Intl. J. Game Theory* **31**:47–492. [Kareev]

Myerson, R. B. 1978. Refinement of the Nash Equilibrium Concept. *Intl. J. Game Theory* **7**:73–80. [Kareev]

Nettle, D. in press. *Error management*.This volume.

Newport, E. L. 1988. Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Lang. Sci.* **10**:147–172. [Kareev]

———. 1990. Maturational constraints on language learning. *Cogn. Sci.* **14**:11–28. [Kareev]

Nickerson, R. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psych.* **2**:175–220. [Kareev]

Nowak, M., and K. Sigmund. 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner’s Dilemma game. *Nature***(364)**:56–58. [Kareev, Dayan]

Nowak, M. A., and R. M. May. 1992. Evolutionary games and spatial chaos. *Nature* **359**:826–829. [Kareev]

Posch, M. 1999. Win-stay, lose-shift strategies for repeated games—memory length, aspiration levels and noise. *J. Theor. Biol.* **198**:183–195. [Kareev]

Posch, M., A. Pichler, and K. Sigmund. 1999. The efficiency of adapting aspiration levels. *Proc. Roy. Soc. B* **266**:1427–1435. [Kareev]

Rakow, T., and K. Miler. 2009. Doomed to repeat the successes of the past: History is best forgotten for repeated choices with nonstationary payoffs. *Mem. Cogn.* **37**:985–1000. [Kareev]

Sandholm, W. H. 1998. Simple and clever decision rules for a model of evolution. *Econ. Lett.* **61**:165–170. [Kareev]

Schlag, K. H. 1999. Which one should I imitate? *J. Math. Econ.* **31**:493–522. [Kareev]

Seligman, M. E. P. 1975. Helplessness: On depression, development, and death. San Francisco: Freeman. [Kareev]

Selten, R. 1975. A reexamination of the perfectness concept for equilibrium points in extensive games. *Intl. J. Game Theory* **4**:25–55. [Brown, Dall, Kareev]

———. 2001. What is bounded rationality? In: Bounded Rationality: The Adaptive Toolbox, ed. G. Gigerenzer and R. Selten, pp. 13–36. Cambridge, MA: MIT Press. [Kareev]

Selten, R., and T. Chmura. 2008. Stationary concepts for experimental 2x2 games. *Am. Econ. Rev.* **98**:938–966. [Kareev]

Shafir, S., and J. Roughgarden. 1994. The effect of memory length on the foraging behavior of a lizard. In: From Animals to Animats: Proceedings of the Third International Conference on the Simulation of Adaptive Behavior, ed. D. Cliff et al., pp. 221–225. Cambridge, MA: MIT Press. [Kareev]

Shanks, D. R. 1995. Is human learning rational? *Q. J. Exp. Psychol. A* **48**:257–279. [Kareev]

Shettleworth, S. J., J. R. Krebs, D. W. Stephens, and J. Gibbon. 1988. Tracking a fluctuating environment: A study of sampling. *Anim. Behav.* **36**:87–105. [Kareev]

Simon, H. A. 1990. Invariants of human behavior. *Ann. Rev. Psychol.* **41**:1–19. [Kareev]

Soffer, S., and Y. Kareev. 2011. The effects of problem content and scientific background on information search and the assessment and valuation of correlations. *Mem. Cogn.* **39**:107–116. [Kareev]

Speekenbrink, M., and D. R. Shanks. 2010. Learning in a changing environment. *J. Exp. Psychol. Gen.* **139**:266–298. [Kareev]

Stewart, N., N. Chater, and G. D. A. Brown. 2006. Decision by sampling. *Cogn. Psychol.* **53(1)**:1–26. [Brown, Chater, Dall]

Turkewitz, G., and P. A. Kenny. 1982. Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Devel. Psychobiol.* **15**:357–368. [Kareev]

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**:1124–1131. [Kareev]

Wallsten, T. S., R. H. Bender, and Y. L. Li. 1999. Dissociating judgment from response processes in statement verification: The effects of experience on each component. *J. Exp. Psychol. Learn. Mem. Cogn.* **25**:96–115. [Kareev]

Wallsten, T. S., and C. González-Vallejo. 1994. Statement verification: A stochastic model of judgment and response. *Psychol. Rev.* **101**:490–504. [Kareev]

Wechsler, D. 1949. Manual for the Wechsler Intelligence Scale for Children. Oxford: The Psychological Corporation. [Kareev]