In K. Fiedler & P. Juslin (Eds.), *Information Sampling and Adaptive Cognition* (pp. 239–260). New York: Cambridge University Press. © 2006.

11

What's in a Sample? A Manual for Building Cognitive Theories

Gerd Gigerenzer

PREVIEW

How do you build a model of mind? I discuss this question from the point of view of sampling. The idea that the mind samples information – from memory or from the environment – became prominent only after researchers began to emphasize sampling methods. This chapter provides a toolbox of potential uses of sampling, each of which can form a building block in a cognitive theory. In it I ask four questions: *who* samples, *why*, *what*, and *how*.

Who: In the social sciences (in contrast to the natural sciences), not only researchers sample, but so do the minds they study. *Why*: I distinguish two goals of sampling, hypotheses testing and measurement. *What*: Researchers can sample participants, objects, and variables to get information about psychological hypotheses, and minds may sample objects and variables to get information about their world. *How*: I distinguish four ways to sample: (i) no sampling, (ii) convenience sampling, (iii) random sampling from a defined population, and (iv) sequential sampling. These uses of sampling have received unequal attention. The prime source of our thinking about sampling seems to be R. A. Fisher's small-sample statistics, as opposed to the use of random sampling in quality control, the use of sequential sampling, and the use of sampling for measurement and parameter estimation. I use this legacy to identify potentials of sampling in adaptive cognition that have received little attention.

In his *Opticks*, Isaac Newton (1952/1704) reported experiments with prisms to demonstrate that white light consists of spectral colors. Newton did not sample, nor was he interested in means or variances. In his view, good experimentation had nothing to do with sampling. Newton was not antagonistic to sampling, but he used it only when he thought it was appropriate, as in quality control. In his role as the master of the London Royal Mint, Newton conducted routine sampling inspections to determine whether the amount of gold in the coins was too little or too large. Just as in Newton's physics, experimentation and statistics were hostile rivals in nineteenth-century physiology and medicine. The great experimenter Claude Bernard used to ridicule the use of samples; his favorite example was that it is silly to collect the urine of one person, or of even a group of persons, over a twenty-four-hour period, because it is not the same before and after digestion, and averages are reifications of unreal conditions (Gigerenzer et al., 1989, p. 129). When B. F. Skinner demonstrated the effects of reinforcement schedules, he used one pigeon at a time, not two dozen. Although Skinner did not sample pigeons, William Estes (1959) pointed out that his theory assumed that his pigeons sampled information about the consequences of their behavior.

These cases illustrate some of the perplexing faces of sampling. What is in a sample? Why coins but not prisms or urine? Why did we come to believe that sampling and experimentation are two sides of the same coin, whereas Newton, Bernard, and Skinner did not? Why did Skinner not sample pigeons but assume that pigeons sample information? In this chapter, I try to put some order into the puzzling uses and nonuses of sampling. In the introduction to this volume, Fiedler and Juslin distinguished various forms of cognitive sampling, such as internal versus external sampling (e.g., memory versus the Internet), and the unit size of the objects of sampling. In contrast, I will focus on the evolution of the ideas of sampling – from the statistical toolbox to theories of mind. The selection of tools that ended up in cognitive theories and those that did not is in part historical accident. For instance, the tools that researchers happened to be familiar with had the best chances of being selected. What I hope to achieve with this chapter is not a complete taxonomy of sampling tools, but rather a view into the toolbox that can help in rethinking the possibilities of sampling and in using the toolbox creatively when building theories of mind. I will proceed to answer the question "What's in a sample?" by asking *who* samples, *what*, *why*, and *how*?

WHO SAMPLES?

I begin with the observation that the answer to the question of *who* samples information is different in the cognitive sciences than in the fields from which statistical sampling theory actually emerged – astronomy, agriculture, demographics, genetics, and quality control. In these noncognitive sciences, the researcher alone may sample (Figure 11.1). For instance, an astronomer may repeatedly measure the position of a star, or an agricultural researcher may fertilize a sample of plots and measure the average number of potatoes grown. Sampling concerns objects that are measured on some variable. Why would that be different in the cognitive sciences?

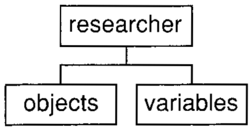

FIGURE 11.1. The structure of the potential uses of sampling in the noncognitive sciences. Researchers may sample objects (such as electrons) to measure these on variables (such as location and mass).

In the cognitive sciences (in contrast to the natural sciences), there are two "classes" of people who can engage in sampling: researchers and the participants of their studies (Figure 11.2). Whether and how researchers draw samples is generally seen as a methodological question. Whether and how researchers think that the minds of their participants engage in sampling of information is treated as a theoretical question. The labels "methodological" and "theoretical" suggest that both questions are unrelated and should be answered independently. After all, what do theories of cognitive processes have to do with the methods to test these theories?

I do not believe that these two issues are generally independent of each other. My hypothesis is that there is a significant correlation (not a one-to-one relation) in cognitive psychology between researchers' sampling practices and the role of sampling in their theories of mind. This hypothesis is an extension of my work on the tools-to-theories heuristic. The general thesis is twofold (Gigerenzer, 1991, 2000):

1. *Discovery*: New scientific tools, once entrenched in a scientist's daily practice, suggest new theoretical metaphors and concepts.
2. *Acceptance*: Once proposed by an individual scientist (or a group), the new theoretical concepts are more likely to be accepted by the scientific community if their members are also users of the new tool.

Note that Sigmund Freud, I. P. Pavlov, and the Gestalt psychologists, as well as the "father" of experimental psychology, Wilhelm Wundt, did not sample participants, and sampling played no role in their theories of mind.

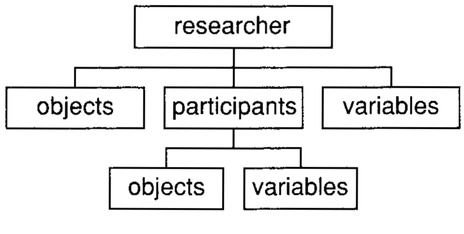

FIGURE 11.2. The structure of the potential uses of sampling in the cognitive sciences. Researchers may sample stimulus objects, participants, or variables, and their participants may themselves sample objects and variables.

All this changed after the unit of investigation ceased to be the individual person, and instead became the group mean – a process that started in the applied fields such as educational psychology (Danziger, 1990). Harold Kelley (1967), for instance, who used sampling and Fisher's analysis of variance to analyze his data, proposed that the mind attributes a cause to an effect in the same way, by sampling information and an intuitive version of analysis of variance. The community of social psychologists who also used analysis of variance as a routine tool accepted the theory quickly, and for a decade it virtually defined what social psychology was about. In contrast, R. Duncan Luce (1988) rejected routine use of analysis of variance as "mindless hypothesis testing in lieu of doing good research" (p. 582), and his theories of mind differed as a consequence. For instance, being familiar with the statistical tools of Jerzy Neyman and Egon S. Pearson and their doctrine of random sampling, Luce (1977) proposed that the mind might draw random samples and make decisions just as Neyman–Pearson theory does.

In summary, I propose that if researchers sample, they are likely to assume in their theories that the mind samples as well. If they do not, their view of cognitive processes typically also does not involve sampling. Moreover, the specific kind of sampling process that researchers use is likely to become part of their cognitive theories.

WHAT'S IN A SAMPLE?

In the cognitive sciences, the object of sampling can be threefold: participants, objects, and variables (Gigerenzer, 1981). Researchers can sample participants, stimulus objects, or variables. Today, participants are sampled habitually, objects rarely, and variables almost never. In addition, the minds under study can sample objects and variables. In cognitive theories, minds mostly sample objects, but rarely variables. This results in five possible uses of sampling in psychology (Figure 11.2).

My strict distinction between the cognitive and noncognitive sciences is an idealization; in reality there are bridges. The astronomers' concern with the "personal equation" of an observer illustrates such a link. Astronomers realized that researchers had systematically different response times when they determined the time a star travels through a certain point. This led to the study of astronomers' personal equations, that is, the time that needed to be subtracted to correct for their individual reaction times. In this situation, the object of sampling was both the researchers and their objects, such as stars (Gigerenzer et al., 1989).

WHY SAMPLING?

I distinguish two goals of sampling: hypotheses testing and measurement. Take significance tests as an example, where a sample statistic – such as t or

F – is calculated. In the early nineteenth century, significance tests were already being used by astronomers (Swijtink, 1987). Unlike present-day psychologists, astronomers used the tests to reject data (so-called outliers), not to reject hypotheses. At least provisionally, the astronomers assumed that a hypothesis (such as normal distribution of observational errors around the true position of a star) was correct and mistrusted the data. In astronomy, the goal was precise measurement, and this called for methods to identify bad data. In psychology, researchers trusted the data and mistrusted the hypotheses; that is, the goal became hypothesis testing, not measurement, following the influence of Fisher.

Hypothesis testing and measurement are concepts taken from statistical theory, and the obvious question is whether they are also good candidates for understanding how the mind works. Whatever the right answer may be, hypothesis testing has been widely assumed to be an adaptive goal of cognition, including in numerous studies that tried to show that people make systematic errors when testing hypotheses. Note that measurement has not been as extensively considered and studied as a goal of cognition (with some exceptions, such as the work of Brunswik, 1955), which is consistent with the fact that researchers tend to use their sampling tools for hypothesis testing rather than measurement.

HOW TO SAMPLE?

Sampling is not sampling. I distinguish four ways of how to sample, beginning with the nonuse of sampling.

Study Ideal Types, Not Samples

Newton thought that the science of optics was close to mathematics, where truth can be demonstrated in one single case, and he loathed researchers who replicated his experiments. Similarly, the most influential psychologists made their fame by studying one individual at a time. Freud's Anna O., Wundt's Wundt (the "father" of experimental psychology served as experimental subject), Pavlov's dog, Luria's mnemonist Shereshevski, and Simon's chess masters are illustrations. They represent ideal types, not averages. They may also represent distinct individual types, such as brain patients with specific lesions. Note that the ideal type approach does not mean that only one individual is studied. There may be several individuals, such as Freud's patients or Skinner's pigeons. The point is that the fundamental unit of analysis is $N = 1$, the singular case.

It is of a certain irony that Fisher's only psychological example in his influential *Design of Experiments* (1935) concerns the analysis of a Lady who claimed that she could tell whether the tea fusion or the milk was poured first into a cup of tea. This single-case study of extraordinary sensory abilities did not become the model for experimental research. Fisher sampled objects, not participants, as in Figure 11.1. Psychologists generally interpreted his methodology to be about sampling participants, not objects.

In his seminal book *Constructing the Subject*, Danziger (1990) argues that the reason why American psychologists turned away from studying individuals in the 1930s and 1940s and embraced means as their new "subject" had little to do with the scientific goals of our discipline. In contrast, this move was largely a reaction to university administrators' pressure on professors of psychology to show that their research was useful for applied fields, specifically educational research, which offered large sources of funding. The educational administrator was interested in questions such as whether a new curriculum would improve the *average* performance of pupils, and not in the study of the laws of the individual mind. Danziger provides detailed evidence that sampling of participants started in the applied fields but not in the core areas of psychology, and in the United States rather than in Germany, where professors of psychology were not, at that time, under pressure to legitimize their existence by proving their practical usefulness. Some of these differences still prevail today: Social psychologists tend to sample dozens or hundreds of undergraduates for five to ten minutes, whereas perceptual psychologists tend to study one or a few participants, each individually and for an extended time.

Convenience Samples

In the 1920s, Ronald A. Fisher (1890–1962) was chief statistician at the agricultural station in Rothamsted. Before Fisher, agricultural researchers had little sense for sampling. For instance, in the mid-nineteenth century, the British agriculturist James F. W. Johnston tried to determine which fertilizer was the best for the growth of turnips. He fertilized one plot, which yielded 24 bushels, and compared this result with those from three plots without fertilizer, which respectively yielded 18, 21, and 24 bushels of grain. Johnston understood that turnips naturally show up to 25% variation from plot to plot and that the average difference of about 10% that he observed was therefore not indicative of a real improvement. What Johnston did not understand was the importance of sample size – that this variability becomes less and less important as the number of plots on which the average is based increases (Gigerenzer et al., 1989, chapter 3).

Fisher's major contribution was to unite the rival practices of scientific experimentation and statistics. From Newton to Bernard to Skinner, this connection, as mentioned, had not existed. Fisher turned the two rival practices into two sides of the same coin, introducing randomized trials to agriculture, genetics, and medicine. By way of parapsychology and education, his ideas also conquered experimental psychology. The marriage between statistics and experimentation also changed statistics, from the general emphasis on large samples to Fisher's small-sample statistics. The idea of basing inferences on small samples – as in the typical experiment – was highly controversial. The statistician Richard von Mises (1957, p. 159) predicted that "the heyday of small sample theory ... is already past." It was not past, however; Fisher prevailed.

Fisher's position emphasized some aspects of sampling – sample size, significance, and random assignment – and left out others. Most importantly, the concept of random sampling from a defined population had no place in Fisher's (1955) theory. Fisher's samples were not randomly drawn from a defined population. There was no such population in the first place. A sample whose population is not known is called a *convenience sample*. Fisher's liberal interpretation of how to sample became entrenched in psychology: The participants in psychological experiments are seldom randomly sampled, nor is a population defined.

Fisher did not think that convenience samples are a weakness. He held that in science there is no known population from which repeated sampling can be done. In a brilliant move, Fisher proposed to view any sample as a random sample from an *unknown hypothetical infinite population*. This solution has puzzled many statisticians: "This is, to me at all events, a most baffling conception"(Kendall, 1943, p. 17). However, Fisher's ideas about sampling were not the last word. Fisher had two powerful rivals, both of whom he detested.

Random Samples

The earliest quasi-random sampling procedure I know of is the trial of the Pyx (Stigler, 1999). The trial is a ceremony that goes back to the Middle Ages, the final stage of a sampling inspection scheme for the control of the quality of the London Royal Mint's production. The word Pyx refers to the box in which the sample of coins was collected, to determine whether the coins were too heavy or too light and contained too much or too little gold. As mentioned before, Newton served as master at the Royal Mint from 1699 until his death in 1727. The same Newton who did not use sampling for scientific experimentation supervised sampling for the purpose of quality control. The trial of the Pyx employed a form of sampling that is different from a convenience sample. It used a random sample drawn from a defined population, the total production of the Mint in one or a few years.

In the twentieth century, hypotheses testing that used random sampling from a defined population was formalized by the Polish statistician Jerzy Neyman and the British statistician Egon S. Pearson, the son of Karl Pearson. In their theory of hypotheses testing, one starts with two hypotheses (rather than one null hypothesis) and the probabilities of the two possible errors, Type I and Type II, from which the necessary sample size is calculated. A random sample is then drawn, after which one of the two hypotheses is accepted, and the other is rejected (in Fisher's scheme, the null can only be rejected, not accepted). Neyman and Pearson believed that they had improved the logic of Fisher's null hypothesis testing. Fisher (1955) did not think so. He thought that those who propose sampling randomly from a defined population and calculating sample size on the basis of cost–benefit trade-offs mistake science for quality control. He compared the Neyman–Pearsonians to Stalin's five-year plans, that is, to Russians who confuse technology with producing knowledge.

Sequential Sampling

A third line of sampling is sequential sampling, which had the status of a military secret during World War II and was later made public by Abraham Wald (1950). In comparison to Fisher's and Neyman and Pearsons's theories, sampling is sequential, not simultaneous. Whereas the sample size in Neyman–Pearson tests is fixed, calculated from a desired probability of Type I and Type II error, there is no fixed sample size in sequential sampling. Rather, a stopping criterion is calculated on the basis of the desired probabilities of Type I and Type II errors, and one continues to sample until it is reached. Sequential sampling has an advantage: It generally results in smaller sample sizes for the same alpha and power. Fisher was not fond of sequential sampling, for the same reasons that he despised Neyman–Pearson's theory. Although sequential sampling can save time and money, researchers in psychology rarely know and use it.

Which of these ideas of sampling have shaped psychological methods and theories of mind? I will now discuss each of the five possibilities for sampling in Figure 11.2.

DO RESEARCHERS SAMPLE PARTICIPANTS?

Do psychologists use individuals or samples as the unit of analysis? In the nineteenth and early twentieth centuries, the unit of analysis was clearly the individual. This changed in the United States during the 1920s, 1930s, and 1940s, when experimental studies of individuals were replaced by the treatment group experiment (Danziger, 1990). The use of samples of individuals began in the applied fields, such as education, and spread from there to the laboratories. The strongest resistance to this change in research practice came from the core of psychological science, perceptual research, where to the present day one can find reports of individual data rather than averages. Nonetheless, sampling participants has largely become the rule in psychology, and its purpose is almost exclusively hypothesis testing, or more precisely, null hypothesis testing. The use of samples for the measurement of parameters is comparatively rare.

How do researchers determine the size of the sample? Psychologists generally use rules of thumb ("25 in each group might be good enough") rather than the cost–benefit calculation prescribed by Neyman and Pearson. For instance, Cohen (1962) analyzed a volume of a major journal and found no calculation of sample size depending on the desired probabilities of Type I and Type II errors. When Sedlmeier and Gigerenzer (1989) analyzed the same journal twenty-four years later, nothing had changed: Sample size was still a matter of convenience, and as a consequence, the statistical power was embarrassingly low – a fact that went unnoticed.

Do researchers draw random samples of participants from a defined population? Experimental studies in which first a population is defined, then a random sample is drawn, and then the members of the sample are randomly assigned to the treatment conditions are extremely rare (e.g., Gigerenzer, 1984). When is sequential sampling of participants used? Virtually never. In summary, when researchers sample participants, they have perfectly internalized Fisher's ideas about sampling – except that, as already mentioned, Fisher sampled objects, not participants.

This almost exclusive reliance on convenience samples and Fisher's analysis of variance creates many of the problems that other uses of sampling tried to avoid. Researchers do not know the power of their tests; measuring constants and curves does not seem to be an issue; they waste time and money by never considering sequential sampling; and when they conclude that there is a true difference in the population means, nobody knows what this population is.

Why sample participants and analyze means if there is no population in the first place? Why not analyze a few individuals? In 1988, I spent a sabbatical at Harvard and had my office next to B. F. Skinner's. I asked him over tea why he continued to report one pigeon rather than averaging across pigeons. Skinner confessed that he once tried to run two dozen pigeons and feed the data into an analysis of variance, but he found that the results were less reliable than with one pigeon. You can keep one pigeon at a constant level of deprivation, he said, but you lose experimental control with twenty-four. Skinner had a point, which W. Gosset, the inventor of the *t* test, made before: "Obviously the important thing... is to have a low real error, not to have a 'significant' result at a particular station. The latter seems to me to be nearly valueless in itself" (quoted in Pearson, 1939, p. 247). The real error can be measured by the standard deviation of the measurements, whereas a *p* value reflects sample size. One can get small real errors by increasing experimental control, rather than by increasing sample size. Experimental control can reveal individual differences in cognitive strategies that get lost in aggregate analyses of variance (e.g., Gigerenzer & Richter, 1990; Robinson, 1950).

In summary, psychologists' sampling of participants follows Fisher's convenience samples. Alternative sampling procedures are practically nonexistent. I believe that it is bad scientific practice to routinely use convenience samples and their averages as units of analysis. Rather, the default should be to analyze each individual on its own. This allows researchers to minimize the real error, to recognize systematic individual differences, and – last but not least – to know one's data.

DO RESEARCHERS SAMPLE OBJECTS?

Fisher made no distinction between the analysis of participants and objects. Do researchers sample stimulus objects in the same way they sample participants? The answer is "no": The classic use of random sampling for measurement in psychophysics has declined, and concern with sampling of objects is rare compared with sampling of participants.

In the astronomer's tradition, the use of random sampling for measurement is the first major use of sampling in psychophysics. In Fechner's work, samples were used to measure absolute and relative thresholds. In Thurstone's (1927) law of comparative judgment, an external stimulus corresponds to an internal normal distribution of subjective values, and a particular encounter with the stimulus corresponds to a randomly drawn subjective value from this distribution. The goal of repeated presentation of the same stimuli is to obtain psychological scales for subjective quantities. As Luce (1977) noted, there is a close similarity between the mathematics in Thurstone's law of comparative judgment and that in signal detection theory, but a striking difference in the interpretation. Thurstone used random variability for measurement, whereas in signal detection theory the mind is seen as an intuitive statistician who actively samples objects (Gigerenzer & Murray, 1987). The use of sampling for measurement has strongly declined since then, owing to the influence of Stevens and Likert, who promoted simple techniques such as magnitude estimation and rating scales that dispensed with the repeated presentation of the same stimulus. A tone, a stimulus person, or an attitude question is presented only once, and the participant is expected to rate it on a scale from, say, one to seven. Aside from research in perception and measurement theory, sampling of objects for the purpose of measuring subjective values and attitudes has been largely driven out of cognitive psychology (see, e.g., Wells & Windschitl, 1999).

As a consequence, Egon Brunswik (e.g., 1955) accused his colleagues of practicing a double standard by being concerned with the sampling of participants but not of stimulus objects. He argued that "representative" sampling of stimuli in natural environments is indispensable for studying vicarious functioning and the adaptation of cognition to its environment. For Brunswik, representative sampling meant random sampling from a defined population. In a classic experiment on size constancy, he walked with the participant through her natural environment and asked her at random intervals to estimate the size of objects she was looking at.

Like Fechner and Thurstone, Brunswik was concerned with measurement, but not with the construction of subjective scales. He understood cognition as an adaptive system and measured its performance in terms of "Brunswik ratios" (during his Vienna period, e.g., for measuring size constancy) and later (while at Berkeley) by means of correlations. He was not concerned with repeated presentations of the same object, nor with random sampling from any population, but with random sampling of objects from a natural population. Brunswik was influenced by the large-sample statistics of Karl Pearson. Pearson, who invented correlation statistics together with Galton, was involved in an intense intellectual and personal feud with Fisher. The clash between these two towering statisticians replicated itself in the division of psychology into two methodologically opposed camps: the large-sample correlational study of intelligence and personality, using the methods of Galton, Pearson, and Spearman, and the small-sample experimental study of cognition, using the methods of Fisher. The schism between these two scientific communities has been repeatedly discussed by the American Psychological Association (e.g., Cronbach, 1957) and still exists in full force today. Intelligence is studied with large samples; thinking is studied with small samples. The members of each community tend not to read and cite what the others write. Brunswik could not persuade his colleagues from the experimental community to take the correlational statistics of the rival discipline seriously. His concept of representative sampling died in the no-man's land between the hostile siblings. Even after the Brunswikian program was revived a decade after Brunswik died (Hammond, 1966; Hammond & Stuart, 2001), the one thing that is hard to find in neo-Brunswikian research is representative sampling.

But does it matter if researchers use random (representative) sampling or a convenience sample that is somehow selected? The answer depends on the goal of the study. If its goal is to measure the accuracy of perception or inaccuracy of judgment, then random sampling matters; if the goal is to test the predictions of competing models of cognitive mechanism, random sampling can be counterproductive, because tests will have higher power when critical items are selected (Rieskamp & Hoffrage, 1999). For claims about cognitive errors and illusions, the sampling of stimulus objects does matter (Gigerenzer & Fiedler, 2004). Research on the so-called overconfidence bias illustrates the point.

In a large number of experiments, participants were given a sample of general knowledge questions, such as "Which city has more inhabitants, Hyderabad or Islamabad?" Participants chose one alternative, such as "Islamabad," and then gave a confidence judgment, such as "70%," that their answer was correct. Average confidence was substantially higher than the proportion correct; this was termed "overconfidence bias" and attributed to a cognitive or motivational flaw. How and from what population the questions were sampled were not specified in these studies. As the story goes, one of the first researchers who conducted these studies went through almanacs and chose the questions with answers that surprised them. However, one can always demonstrate good or bad performance, depending on what items one selects. When we introduced random sampling from a defined population (cities in Germany), "overconfidence bias" largely or completely disappeared (Gigerenzer, Hoffrage, & Kleinbölting, 1991). The message that one of the most "stable" cognitive illusions could be largely due to researchers' sampling procedure was hard to accept, however, and was debated for years (e.g., by Griffin & Tversky, 1992). Finally, Juslin, Winman, & Olsson (2000) published a seminal review of more than 100 studies showing that "overconfidence bias" is practically zero with random sampling, but substantial with selected sampling (for a discussion of sampling in overconfidence research, see also Klayman et al., this volume; Hoffrage & Hertwig, this volume).

In summary, whereas sampling of participants has become institutionalized in experimental psychology, sampling of stimulus objects has not. Except for a few theories of measurement, which include psychophysics and Brunswik's representative design, it is not even an issue of general concern.

DO RESEARCHERS SAMPLE VARIABLES?

Now we enter no-man's land. Why would a researcher sample variables, and what would that entail? Few theories in psychology are concerned with how the experimenter samples the variables on which participants judge objects. One is personal construct theory (Kelly, 1955). The goal of the theory is to analyze the "personal constructs" people use to understand themselves and their world. George Kelly's emphasis on the subjective construction of the world precludes the use of a fixed set of variables, such as a semantic differential, and imposition of it on all participants. Instead, Kelly describes methods that elicit the constructs relevant for each person. One is to present triples of objects (such as mother, sister, and yourself), and to ask the participant first which of the two are most similar, then what it is that makes them so similar, and finally what makes the two different from the third one.

Unlike when sampling participants and objects, situations in which a population of variables can be defined are extremely rare. In Kelly's attempts to probe individual constructs, for instance, the distinction between convenience samples and random or representative samples appears blurred. If the goal of the research is to obtain statements about the categories or dimensions in which people see their world, then the researcher needs to think of how to sample the relevant individual variables.

I turn now to theories of how minds sample. According to our scheme, minds can sample along two dimensions: objects and cues (variables).

DO MINDS SAMPLE OBJECTS?

The idea that the mind samples objects to compute averages, variances, or test hypotheses only emerged after the institutionalization of inferential statistics in psychology (1940–1955), consistent with the tools-to-theories heuristic (Gigerenzer & Murray, 1987). From Fechner to Thurstone, probability was linked with the measurement of thresholds and the construction of scales of sensation, but not with the image of the mind as an intuitive statistician who draws samples for *cognitive inferences* or *hypothesis testing*. One of the first and most influential theories of intuitive statistics was signal detection theory (Tanner & Swets, 1954), which transformed Neyman–Pearson theory into a theory of mind.

There seem to be two main reasons for this late emergence of the view that the mind actively engages in sampling, The first is described by tools-to-theories: Only after a combination of Fisher's and Neyman–Pearson's statistical tools became entrenched in the methodological practices of psychologists around 1950 did researchers begin to propose and accept the idea that the mind might also be an intuitive statistician who uses similar tools (Gigerenzer, 1991). The second reason is the influence of Stanley S. Stevens, who rejected inferential statistics (Gigerenzer & Murray, 1987, chapter 2). For instance, in the first chapter of his *Handbook of Experimental Psychology*, Stevens (1951) included a section entitled "probability" (pp. 44–47), whose only purpose seems to be to warn the reader of the confusion that might result from applying probability theory to anything, including psychology. He was deeply suspicious of probabilistic models on the grounds that they can never be definitely disproved. Like David Krech and Edwin G. Boring, Stevens stands in a long tradition of psychologists who are determinists at heart.

Yet many current theories in cognitive and social psychology still do not incorporate any models of sampling. Consistent with this omission, most experimental tasks lay out all objects in front of the participants and thereby exclude information search in the first place. This tends to create cognitive theories with a blind spot for how people sample information and when they stop. This in turn creates a blind spot for the situations in which the mind does and does not sample, including when there might be evolutionary reasons to rely only on a single observation.

When Is It Adaptive Not to Sample?

Although Skinner did not sample pigeons, as mentioned before, his view about operant conditioning can be seen as a theory of information sampling. Specifically, this interpretation is invited by his variable reinforcement schedules, where an individual repeatedly exhibits a behavior (such as pecking in pigeons and begging in children) and samples information about consequences (such as food). Skinner's laws of operant behavior were designed to be general-purpose, that is, to hold true for all stimuli and responses. This assumption is known as the *equipotentiality* hypothesis. Similarly, when Thorndike found that cats were slow in learning to pull strings to escape from puzzle boxes, he concluded that learning occurs by trial and error, and he hoped that this would be a general law of learning. If all stimuli were equal, minds should always sample information to be able to learn from experience. The assumption that all stimuli are equal is also implicit in many recent versions of reinforcement learning (e.g., Erev & Roth, 2001). Consider William Estes, who was one of the first to formulate Skinner's ideas in the language of sampling:

> All stimulus elements are equally likely to be sampled and the probability of a response at any time is equal to the proportion of elements... that are connected to it.... On any acquisition trial all stimulus elements sampled by the organism become connected to the response reinforced on that trial. (Estes, 1959, p. 399)

Is the assumption of the equivalence of stimulus objects in sampling correct? Are there stimulus objects that an organism does not and should not sample? John Garcia is best known for his challenge of the equipotentiality hypothesis. For instance, he showed that in a single trial, a rat can learn to avoid flavored water when it is followed by experimentally induced nausea, even when the nausea occurs two hours later. However, the same rat has great difficulty learning to avoid the flavored water when it is repeatedly paired with an electric shock immediately after the tasting:

> From the evolutionary view, the rat is a biased learning machine designed by natural selection to form certain CS-US [conditional stimulus – unconditional stimulus] associations rapidly but not others. From a traditional learning viewpoint, the rat was an unbiased learner able to make any association in accordance with the general principles of contiguity, effect, and similarity. (Garcia y Robertson & Garcia, 1985, p. 25)

The evolutionary rationale for one-trial learning as opposed to sampling stimulus objects is transparent. Learning by sampling and proportionally increasing the probability of response can be dangerous or deadly when it comes to food, diet, and health. To avoid food poisoning, an organism can have a genetically inherited aversion against a food, or a genetically coded preparedness to learn a certain class of associations in one or a few instances.

Genetically coded preparedness shows that sampling cannot and should not be an element of all cognitive processes. Rather, whether an organism samples (a so-called bottom-up process) or does not (a top-down process) largely depends on the past and present environmental contingencies. A mind can afford to learn some contingencies, but not all – sampling can be overly dangerous. One-trial learning amply illustrates the adaptive nature of cognitive processes, which codes what will be sampled and what will not.

Convenience Sampling

One class of models developed after the inference revolution assumes that the mind samples information to test hypotheses, just as researchers came to do. Consider the question of how the mind attributes a cause to an event, which has been investigated in the work of Piaget and Michotte. In Michotte's (1963/1946) view, for instance, causal attribution was a consequence of certain spatio-temporal relationships; that is, it was determined "outside" the mind and did not involve inductive inference based on samples of information (see also Chater & Oaksford, this volume). After analysis of variance became institutionalized in experimental psychology, Harold Kelley (1967) proposed that the mind attributes a cause to an event just as researchers test causal hypotheses: by analyzing samples of covariation information and calculating F ratios (F for Fisher) in an intuitive analogy to analysis of variance. Note that the new ANOVA mind used the tests for rejecting hypotheses while trusting the data, parallel to the way researchers in psychology use ANOVA. If Kelley had lived a century and a half earlier, he might have instead looked to the astronomers' significance tests. As pointed out earlier, the astronomers assumed (at least, provisionally) that the hypothesis was correct but mistrusted the data. If this use of sampling had been taken as an analogy, the mind would have appeared to be expectation-driven rather than data-driven.

Kelley's causal attribution theory illustrates how Fisher's ANOVA was used to model the mind's causal thinking, assuming that the mind uses convenience samples for making inductive inferences about causal hypotheses.

As clear as the distinction between convenience and random sampling is in statistical theory, it is less so in theories that assume that the mind samples objects. Is the sample of people a tourist encounters on a trip to Beijing a random sample or a convenience sample? It may depend on whether the tour guide has planned all encounters ahead, or whether the tourist strolls through the city alone, or whether the tour guide has picked a random sample of Beijing tenth-graders to meet with.

Random Sampling

Psychophysics has been strongly influenced by Neyman–Pearson theory. Under the name of signal detection theory, it became a model of how the mind detects a stimulus against noise or a difference between two stimuli, and it replaced the concepts of absolute and relative thresholds.

Neyman's emphasis on random sampling from a defined population, as in quality control, became part of the cognitive mechanisms. For instance, Luce (1977; Luce & Green, 1972) assumed that a transducer (such as the human ear) transforms the intensity of a signal into neural pulse trains in parallel nerve fibers, and the central nervous system (CNS) draws a random sample of all activated fibers. The size of the sample is assumed to depend on whether or not the signal activates fibers to which the CNS is attending. From each fiber in the sample, the CNS estimates the pulse rate by either counting or timing, and these numbers are then aggregated into a single internal representation of the signal intensity. In Luce's theory, the mind was pictured as a statistician of the Neyman and Pearson school, and the processes of random sampling, inference, decision, and hypotheses testing were freed of their conscious connections and seen as unconscious mechanisms of the brain.

Sequential Sampling

The former First Lady, Barbara Bush, is reported to have said, "I married the first man I ever kissed. When I tell this to my children they just about throw up" (quoted in Todd & Miller, 1999). Is one enough, just as in Garcia's experiments, or should Barbara Bush have sampled more potential husbands? After Johannes Kepler's first wife died of cholera, he immediately began a methodological search for a replacement. Within two years, he investigated eleven candidates and finally married Number 5, a woman who was well educated but not endowed with the highest rank or dowry. Are eleven women a large enough sample? Perhaps too large, because the candidate Number 4, a woman of high social status and with a tempting dowry, whom friends urged Kepler to choose, rejected him for having toyed with her too long (Todd & Miller, 1999). Swiss economists Frey and Eichenberger (1996) asserted that people do not sample enough when seeking a mate, taking the high incidence of divorce and marital misery as evidence. In contrast, Todd and Miller (1999) argued that given the degree of uncertainty – one never can know how a prospective spouse will turn out – the goal of mate search can only be to find a fairly good partner, and they showed that under certain assumptions, Kepler's sample was large enough.

Mate search is essentially sequential for humans, although there are female birds that can inspect an entire sample of males lined up simultaneously. Since sequential sampling has never become part of the statistical tools used by researchers in psychology, one might expect from the tools-to-theories heuristic that minds are not pictured as performing sequential sampling either. This is mostly, but not entirely, true.

Cognitive processes that involve sequential sampling have been modeled in two different ways: optimizing models and heuristic models. Optimizing models are based on Abraham Wald's (1950) statistical theory, which has a stopping rule that is optimal relative to given probabilities of Type I and Type II errors (e.g., Anderson, 1990; for a discussion of optimal stopping rules see Busemeyer & Rapoport, 1988). Many of these models have been applied to psychophysical tasks, such as judging which of two lines is longer. In the case of a binary hypothesis (such as line *A* or *B*; marry or not marry), the basic idea of most sequential models is the following: A threshold is calculated for accepting one of the two hypotheses, based on the costs of the two possible errors, such as wrongly judging line *A* as larger, or wrongly deciding that to marry is the better option. Each reason or observation is then weighted and sampling of objects is continued until the threshold for one hypothesis is met, at which point the search is stopped and the hypothesis is accepted. These models are often presented as *as-if* models, whose task is to predict the outcome rather than the process of decision making, although it has been suggested that the calculations might be performed unconsciously.

Heuristic models of sequential sampling assume an aspiration level rather than optimization. Their goal is to model the process and the outcome of judgment or decision making. For instance, in Herbert Simon's (1955, 1956) models of satisficing, a person sequentially samples objects (such as houses or potential spouses) until encountering the first one that meets an aspiration level. In Reinhard Selten's (2001) theories of satisficing, the aspiration level can change with the duration of the sampling process.

Can sequential sampling ever be random? In statistical theory, the answer is yes. One draws sequentially from a population, until the stopping rule applies. In the case of mental sampling, it is much harder to decide whether a sequential search process should count as random. Consider, for instance, a satisficer who sequentially encounters potential spouses or houses until finding one that exceeds her aspiration level. In most cases, the sequential sample will be a convenience sample rather than a random sample from a defined population.

The relative rarity of sequential sampling in models of the mind goes hand in hand with experimenters' preference for tasks that do not provide an opportunity for the participants to sample objects: All objects are already displayed in front of the participant. Few experiments address the following questions: (i) When does the mind sample simultaneously versus sequentially? (ii) Is there an order in sequential search, that is, is the search random or systematic? (iii) How is sequential search stopped, that is, what determines when a sample is large enough?

DOES THE MIND SAMPLE VARIABLES?

Satisficing refers to a class of heuristics that apply to situations in which an aspiration level is given and the objects or alternatives are sampled sequentially. Alternatively, the objects can be given and the variables (cues, reasons, or features) need to be searched. Examples include choosing between two job offers (paired comparison) and classifying patients as high-risk or low-risk (categorization). As I mentioned, cognitive theories that model how minds sample objects are few, but those that model how minds sample variables are even rarer. For instance, models of similarity generally assume that the variables (features, cues, etc.) are already given and then postulate some way in which individual features are combined to form a similarity judgment – city block distance and feature matching are illustrations of this. However, in everyday situations, the features are not always laid out in front of a person but need to be searched for, and since there is typically a large or infinite number of features or cues, cognition may involve sampling features. Sampling cues or features can occur inside or outside of memory (e.g., on the Internet).

Unlike for sequential sampling of objects, there seem to be no optimizing models but only heuristic models for sampling of variables. There are two possible reasons. First, it is hard to think of a realistic population of variables, in contrast to a population of objects. Two job candidates, for instance, can vary on many different cues, and it is hard to define a population of cues. Second, the large number of cues makes optimizing models such as Bayes's rule or full classification trees computationally intractable, because the number of decision nodes increases exponentially with the number of cues in a full tree. Thus, even optimizing models need to use heuristic simplifications, as in Bayesian trees (Martignon & Laskey, 1999).

Heuristic models of sequential sampling include two major classes: one-reason decision making and tallying. Each heuristic consists of a search rule that specifies the direction of sampling, a stopping rule that specifies when sampling is terminated, and a decision rule. "Take the Best" (Gigerenzer & Goldstein, 1996) is an example of a heuristic that employs ordered search and one-reason decision making.

Take the Best

1. Search by validity: Search through cues in order of their validity. Look up the cue values of the cue with the highest validity first.
2. One-reason stopping rule: If one object has a positive cue value (1) and the other does not (0 or unknown), then stop the search and proceed to Step 3. Otherwise exclude this cue and return to Step 1. If no more cues are found, guess.
3. One-reason decision making: Predict that the object with the positive cue value (1) has the higher value on the criterion.

The validity of a cue i is defined as $v_i = R_i/P_i$, where R_i = number of correct predictions by cue i, and P_i = number of pairs where the values of cue i differ between objects. "Take The Best" typically samples a small number of cues. Now consider an example for a tallying heuristic, which relies on adding but not on weighing (or order).

Tallying

1. Random search: Search through cues in random order. Look up the cue values.
2. Stopping rule: After $m(1 < m \leq M)$ cues, stop the search and determine which object has more positive cue values (1), and proceed to Step 3. If the number is equal, return to Step 1 and search for another cue. If no more cues are found, guess.
3. Tallying rule: Predict that the object with the higher number of positive cue values (1) has the higher value on the criterion.

Here M refers to the total number of cues and m refers to the number of cues searched for.

The literature discusses various versions of tallying, such as unit-weight models in which all cues ($m = M$) or the m significant cues are looked up (Dawes, 1979; Einhorn & Hogarth, 1975). Unlike as-if models, which predict outcomes only, these models of heuristics predict process and outcome and can be subjected to a stronger test. A discussion of these and similar heuristics and the situations in which they are accurate can be found in Gigerenzer et al. (1999) and Gigerenzer & Selten (2001).

In summary, cognitive sampling of cues or features is a process that has been given little attention. Just as for sampling of objects, heuristic models exist that formulate stopping rules to determine when such a sample is large enough.

WHAT'S IN A SAMPLE?

Shakespeare has Juliet ask; "What's in a name?" What is in a name uncovers what the name means to us, and by analogy, what is in a sample reveals what sampling means to us. The taxonomy proposed in this chapter distinguishes two subjects of sampling (experimenter versus participant), two purposes of sampling (measurement versus hypotheses testing), three targets of sampling (participants, objects, and variables), and four ways of how to sample ($N = 1$, i.e., no sampling; convenience sampling; random sampling; and sequential sampling). As in Brunswik's representative design, these dimensions do not form a complete factorial design; for instance, participants do not sample participants. Among the logically possible uses of sampling, some are realized in practice, whereas others are not or are realized only by a minority. Is the resulting picture of the actual uses and nonuses of sampling one of chaos, orderly chaos, or reasonable choice? Is the overreliance on Fisher's convenience sampling in methodology a good or bad thing, and is the relative neglect of sequential sampling in both methodology and cognitive theories reasonable or unreasonable? Why is so little attention paid to the mind's sampling of features?

Whatever the reader's evaluation is, a toolbox can open one's eyes to the missed opportunities or blind spots of sampling. There may be other paths to a toolbox of methods for sampling; the present one has a deliberate bias toward the evolution of the various ideas of sampling and the intellectual inheritance we owe to competing statistical schools. This historical window allows us to understand the current patchwork of sampling in both methodology and theory along with the possibilities of designing new theories of mind that overcome the historical biases we inherited.

References

Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Erlbaum.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193–217.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology, 32*, 91–134.

Chater, N., & Oaksford, M. (2005). Mental mechanisms: Speculations on human causal learning and reasoning. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition*. Cambridge, UK: Cambridge University Press.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145–153.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge, UK: Cambridge University Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571–582.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13*, 171–192.

Erev, I., & Roth, A. E. (2001). Simple reinforcement learning models and reciprocation in the prisoner's dilemma game. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 215–231). Cambridge, MA: MIT Press.

Estes, W. K. (1959). The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of science* (Vol. 2, pp. 380–491). New York: McGraw-Hill.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society, Series B, 17*, 69–78.

Frey, B. S., & Eichenberger, R. (1996). Marriage paradoxes. *Rationality and Society, 8*, 187–206.

Garcia y Robertson, R., & Garcia, J. (1985). X-rays and learned taste aversions: Historical and psychological ramifications. In T. G. Burish, S. M. Levy, & B. E. Meyerowitz (Eds.), *Cancer, nutrition and eating behavior: A biobehavioral perspective* (pp. 11–41). Hillsdale, NJ: Lawrence Erlbaum.

Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie* [Measurement and modeling in psychology]. Munich: Ernst Reinhardt (UTB).

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *American Journal of Psychology, 97*, 185–195.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review, 98*, 254–267.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world.* New York: Oxford University Press.

Gigerenzer, G., & Fiedler, K. (2004). *Minds in environments: The potential of an ecological approach to cognition.* Manuscript submitted for publication.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Lawrence Erlbaum.

Gigerenzer, G., & Richter, H. R. (1990). Context effects and their interaction with development: Area judgments. *Cognitive Development, 5*, 235–264.

Gigerenzer, G., & Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: MIT Press.

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, UK: Cambridge University Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

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

Hammond, K. R. (1966). *The psychology of Egon Brunswik.* New York: Holt, Rinehart & Winston.

Hammond, K. R., & Stewart, T. R. (Eds.) (2001). *The essential Brunswik: Beginnings, explications, applications.* New York: Oxford University Press.

Hoffrage, U., & Hertwig, R. (2005). Which world should be represented in representative design? In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition.* Cambridge, UK: Cambridge University Press.

Juslin, P., Winman, A., & Olssen, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107*, 384–396.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238). Lincoln: University of Nebraska Press.

Kelly, G. A. (1955). *The Psychology of personal constructs.* New York: Norton.

Kendall, M. G. (1943). *The advanced theory of statistics* (Vol. 1). New York: Lippincott.

Klayman, J., Soll, J., Juslin, P., & Winman, A. (2005). Subjective confidence and the sampling of knowledge. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition.* Cambridge, UK: Cambridge University Press.

Luce, R. D. (1977). Thurstone's discriminal processes fifty years later. *Psychometrika, 42*, 461–489.

Luce, R. D. (1988). The tools-to-theory hypothesis. Review of G. Gigerenzer and D. J. Murray, "Cognition as intuitive statistics". *Contemporary Psychology, 32*, 151–178.

Luce, R. D., & Green, D. M. (1972). A neural timing theory for response times and the psychophysics of intensity. *Psychological Review, 79*, 14–57.

Martignon, L., & Laskey, K. B. (1999). Bayesian benchmarks for fast and frugal heuristics. In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 169–188). New York: Oxford University Press.

Michotte, A. (1963). *The perception of causality*. London: Methuen. (Original work published 1946.)

Newton, I. (1952). *Opticks: Or a treatise of the reflections, refractions, inflections and colours of light*. New York: Dover. (Original work published 1704.)

Pearson, E. S. (1939). "Student" as statistician. *Biometrika, 30*, 210–250.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.

Robinson, W. (1950). Ecological correlation and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309–316.

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

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

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

Stevens, S. S. (Ed.) (1951). *Handbook of experimental psychology*. New York: Wiley.

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.

Swijtink, Z. G. (1987). The objectification of observation: Measurement and statistical methods in the nineteenth century. In L. Krüger, L. Daston, & M. Heidelberger (Eds.), *The probabalistic revolution, Vol. I: Ideas in history* (pp. 261–285). Cambridge, MA: MIT Press.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*, 401–409.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.

Todd, P. M., & Miller, G. F. (1999). From pride and prejudice to persuasion: Satisficing in mate search. In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 287–308). New York: Oxford University Press.

von Mises, R. (1957). *Probability, statistics, and truth*. London: Allen and Unwin. (Original work published 1928.)

Wald, A. (1950). *Statistical decision functions*. New York: Wiley.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115–1125.