From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology

Gerd Gigerenzer
Center for Advanced Study in the Behavioral Sciences
Stanford, California

The study of scientific discovery—where do new ideas come from?—has long been denigrated by philosophers as irrelevant to analyzing the growth of scientific knowledge. In particular, little is known about how cognitive theories are discovered, and neither the classical accounts of discovery as either probabilistic induction (e.g., Reichenbach, 1938) or lucky guesses (e.g., Popper, 1959), nor the stock anecdotes about sudden "eureka" moments deepen the insight into discovery. A heuristics approach is taken in this review, where heuristics are understood as strategies of discovery less general than a supposed unique logic of discovery but more general than lucky guesses. This article deals with how scientists' tools shape theories of mind, in particular with how methods of statistical inference have turned into metaphors of mind. The tools-to-theories heuristic explains the emergence of a broad range of cognitive theories, from the cognitive revolution of the 1960s up to the present, and it can be used to detect both limitations and new lines of development in current cognitive theories that investigate the mind as an "intuitive statistician."

Scientific inquiry can be viewed as "an ocean, continuous everywhere and without a break or division" (Leibniz, 1690/1951, p. 73). Hans Reichenbach (1938) nonetheless divided this ocean into two great seas, the context of discovery and the context of justification. Philosophers, logicians, and mathematicians claimed justification as a part of their territory and dismissed the context of discovery as none of their business, or even as "irrelevant to the logical analysis of scientific knowledge" (Popper, 1959, p. 31). Their sun shines over one part of the ocean and has been enlightening about matters of justification, but the other part of the ocean still remains in a mystical darkness where imagination and intuition reign, or so it is claimed. Popper, Braithwaite, and others ceded the dark part of the ocean to psychology and, perhaps, sociology, but few psychologists have fished in these waters. Most did not dare or care.

The discovery versus justification distinction has oversimplified the understanding of scientific inquiry. For instance, in the recent debate over whether the context of discovery is relevant to understanding science, both sides in the controversy have construed the question as whether the *earlier* stage of discovery should be added to the *later* justification stage (Nickles, 1980). Conceiving the two-context distinction as a temporal distinction (first discovery, then justification), however, can be misleading because justification procedures (checking and testing) and discovery processes (having new ideas) take place during all temporal stages of inquiry. In fact, the original distinction drawn by Reichenbach in 1938 did not include this temporal simplification; his was not even a strict dichotomy (see Curd, 1980). I believe that the prevailing interpretation of the two contexts as conceptually distinct events that are in one and only one temporal sequence has misled many into trying to understand discovery without taking account of justification.

In this article, I argue that discovery can be understood by heuristics (not a logic) of discovery. I propose a heuristic of discovery that makes use of methods of justification, thereby attempting to bridge the artificial distinction between the two. Furthermore, I attempt to demonstrate that this discovery heuristic may be of interest not only for an *a posteriori* understanding of theory development, but also for understanding limitations of present-day theories and research programs and for the further development of alternatives and new possibilities. The discovery heuristic that I call the *tools-to-theories heuristic* (see Gigerenzer & Murray, 1987) postulates a close connection between the light and the dark parts of Leibniz's ocean: Scientists' tools for justification provide the metaphors and concepts for their theories.

The power of tools to shape, or even to become, theoretical concepts is an issue largely ignored in both the history and philosophy of science. Inductivist accounts of discovery, from Bacon to Reichenbach and the Vienna School, focus on the role of data but do not consider *how* the data are generated or processed. Nor do the numerous anecdotes about discoveries—Newton watching an apple fall in his mother's orchard while pondering the mystery of gravitation; Galton taking shelter from a rainstorm during a country outing when discovering correlation and regression toward mediocrity; and the stories about Fechner, Kekulé, Poincaré, and others, which link discovery to beds, bicycles, and bathrooms. What unites these anecdotes is the focus on the vivid but prosaic circumstances; they

---

report the setting in which a discovery occurs, rather than analyzing the process of discovery.

The question Is there a logic of discovery? and Popper's (1959) conjecture that there is none have misled many into assuming that the issue is whether there exists a logic of discovery or only idiosyncratic personal and accidental reasons that explain the "flash of insight" of a particular scientist (Nickles, 1980). I do not think that formal logic and individual personality are the only alternatives, nor do I believe that either of these is a central issue for understanding discovery.

The process of discovery can be shown, according to my argument, to possess more structure than thunderbolt guesses but less definite structure than a monolithic logic of discovery, of the sort Hanson (1958) searched for, or a general inductive hypothesis-generation logic (e.g., Reichenbach, 1938). The present approach lies between these two extremes; it looks for structure beyond the insight of a genius but does not claim that the tools-to-theories heuristic is (or should be) the only account of scientific discovery. The tools-to-theories heuristic applies neither to all theories in science nor to all cognitive theories; it applies to a specific group of cognitive theories developed during the last three or four decades, after the so-called cognitive revolution.

Nevertheless, similar heuristics have promoted discovery in physics, physiology, and other areas. For instance, it has been argued that once the mechanical clock became the indispensable tool for astronomical research, the universe itself came to be understood as a kind of mechanical clock, and God as a divine watchmaker. Lenoir (1986) showed how Faraday's instruments for recording electric currents shaped the understanding of electrophysiological processes by promoting concepts such as "muscle current" and "nerve current."

Thus, this discovery heuristic boasts some generality both within cognitive psychology and within science, but this generality is not unrestricted. Because there has been little research in how tools of justification influence theory development, the tools-to-theories heuristic may be more broadly applicable than I am able to show in this article. If my view of heuristics of discovery as a heterogeneous bundle of search strategies is correct, however, this implies that generalizability is, in principle, bounded.

What follows has been inspired by Herbert Simon's notion of heuristics of discovery but goes beyond his attempt to model discovery with programs such as BACON that attempt to induce scientific laws from *data* (discussed later). My focus is on the role of the tools that process and produce data, not the data themselves, in the discovery and acceptance of theories.

How Methods of Justification Shape Theoretical Concepts

My general thesis is twofold: (a) Scientific tools (both methods and instruments) suggest new theoretical metaphors and theoretical concepts once they are entrenched in scientific practice. (b) Familiarity with the tools within a scientific community also lays the foundation for the general acceptance of the theoretical concepts and metaphors inspired by the tools.

By *tools* I mean both analytical and physical methods that are used to evaluate given theories. Analytical tools can be either empirical or nonempirical. Examples of analytical methods of the empirical kind are tools for data processing, such as statistics; examples of the nonempirical kind are normative criteria for the evaluation of hypotheses, such as logical consistency. Examples of physical tools of justification are measurement instruments, such as clocks. In this article, I focus on analytical rather than physical tools of justification, and among these, on techniques of statistical inference and hypothesis testing. My topic is theories of mind and how social scientists discovered them after the emergence of new tools for data analysis, rather than of new data.

In this context, the tools-to-theories heuristic consists in the discovery of new theories by changing the conception of the mind through the analogy of the statistical tool. The result can vary in depth from opening new general perspectives, albeit mainly metaphorical, to sharp discontinuity in specific cognitive theories caused by the direct transfer of scientists' tools into theories of mind.

A brief history follows. In American psychology, the study of cognitive processes was suppressed in the early 20th century by the allied forces of operationalism and behaviorism. The operationalism and the inductivism of the Vienna School, as well as the replacement of the Wundtian experiment by experimentation with treatment groups (Danziger, 1990), paved the way for the institutionalization of inferential statistics in American experimental psychology between 1940 and 1955 (Gigerenzer, 1987a; Toulmin & Leary, 1985). In experimental psychology, inferential statistics became almost synonymous with scientific method. Inferential statistics, in turn, provided a large part of the new concepts for mental processes that have fueled the so-called cognitive revolution since the 1960s. Theories of cognition were cleansed of terms such as restructuring and insight, and the new mind has come to be portrayed as drawing random samples from nervous fibers, computing probabilities, calculating analyses of variance (ANOVA), setting decision criteria, and performing utility analyses.

After the institutionalization of inferential statistics, a broad range of cognitive processes, conscious and unconscious, elementary and complex, were reinterpreted as involving "intuitive statistics." For instance, Tanner and Swets (1954) assumed in their theory of signal detectability that the mind "decides" whether there is a stimulus or only noise, just as a statistician of the Neyman-Pearson school decides between two hypotheses. In his causal attribution theory, Harold H. Kelley (1967) postulated that the mind attributes a cause to an effect in the same way as behavioral scientists have come to do, namely by performing an ANOVA and testing null hypotheses. These two influential theories show the breadth of the new conception of the "mind as an intuitive statistician" (Gigerenzer, 1988; Gigerenzer & Murray, 1987). They also exemplify cognitive theories that were suggested not by new data, but by new tools of data analysis.

In what follows, I present evidence for three points. First, the discovery of theories based on the conception of the mind as an intuitive statistician caused discontinuity in theory rather than being merely a new, fashionable language: It radically changed the kind of phenomena reported, the kind of explanations looked for, and even the kind of data that were generated. This first point illustrates the profound power of the tools-to-theor-

ies heuristic to generate quite innovative theories. Second, I provide evidence for the "blindness" or inability of researchers to discover and accept the conception of the mind as an intuitive statistician before they became familiar with inferential statistics as part of their daily routine. The discontinuity in cognitive theory is closely linked to the preceding discontinuity in method, that is, to the institutionalization of inferential statistics in psychology. Third, I show how the tools-to-theories heuristic can help to define the limits and possibilities of current cognitive theories that investigate the mind as an intuitive statistician.

Discontinuity in Cognitive Theory Development

What has been called the "cognitive revolution" (Gardner, 1985) is more than the overthrow of behaviorism by mentalist concepts. These concepts have been continuously part of scientific psychology since its emergence in the late 19th century, even coexisting with American behaviorism during its heyday (Lovie, 1983). The cognitive revolution did more than revive the mental; it has changed what the mental means, often dramatically. One source of this change is the tools-to-theories heuristic, with its new analogy of the mind as an intuitive statistician. To show the discontinuity within cognitive theories, I briefly discuss two areas in which an entire statistical technique, not only a few statistical concepts, became a model of mental processes: (a) stimulus detection and discrimination and (b) causal attribution.

What intensity must a 440-Hz tone have to be perceived? How much heavier than a standard stimulus of 100 g must a comparison stimulus be in order for a perceiver to notice a difference? How can the elementary cognitive processes involved in those tasks, known today as *stimulus detection* and *stimulus discrimination,* be explained? Since Herbart (1834), such processes have been explained by using a threshold metaphor: Detection occurs only if the effect an object has on the nervous system exceeds an absolute threshold, and discrimination between two objects occurs if the excitation from one exceeds that from another by an amount greater than a differential threshold. E. H. Weber and G. T. Fechner's laws refer to the concept of fixed thresholds; Titchener (1896) saw in differential thresholds the long sought-after elements of mind (he counted approximately 44,000); and classic textbooks, such as Brown and Thomson's (1921) and Guilford's (1954), document methods and research.

Around 1955, the psychophysics of absolute and differential thresholds was revolutionized by the new analogy between the mind and the statistician. W. P. Tanner and others proposed a "theory of signal detectability" (TSD), which assumes that the Neyman-Pearson technique of hypothesis testing describes the processes involved in detection and discrimination. Recall that in Neyman-Pearson statistics, two sampling distributions (hypotheses H_0 and H_1) and a decision criterion (which is a likelihood ratio) are defined, and then the data observed are transformed into a likelihood ratio and compared with the decision criterion. Depending on which side of the criterion the data fall, the decision "reject H_0 and accept H_1" or "accept H_0 and reject H_1" is made. In straight analogy, TSD assumes that the mind calculates two sampling distributions for noise and signal plus noise (in the detection situation) and sets a decision criterion after weighing the cost of the two possible decision errors (Type I and Type II errors in Neyman-Pearson theory, now called *false alarms* and *misses*). The sensory input is transduced into a form that allows the brain to calculate its likelihood ratio, and depending on whether this ratio is smaller or larger than the criterion, the subject says "no, there is no signal" or "yes, there is a signal." Tanner (1965) explicitly referred to his new model of the mind as a "Neyman-Pearson" detector, and, in unpublished work, his flowcharts included a drawing of a homunculus statistician performing the unconscious statistics in the brain (Gigerenzer & Murray, 1987, pp. 49–53).

The new analogy between mind and statistician replaced the century-old concept of a fixed threshold by the twin notions of observer's attitudes and observer's sensitivity. Just as Neyman-Pearson technique distinguishes between a subjective part (e.g., selection of a criterion dependent on cost–benefit considerations) and a mathematical part, detection and discrimination became understood as involving both subjective processes, such as attitudes and cost–benefit considerations, and sensory processes. Swets, Tanner, and Birdsall (1964, p. 52) considered this link between attitudes and sensory processes to be the main thrust of their theory. The analogy between technique and mind made new research questions thinkable, such as How can the mind's decision criterion be manipulated? A new kind of data even emerged: Two types of error were generated in the experiments, false alarms and misses, just as the statistical theory distinguishes two types of error.

As far as I can tell, the idea of generating these two kinds of data was not common before the institutionalization of inferential statistics. The discovery of TSD was not motivated by new data; rather, the new theory motivated a new kind of data. In fact, in their seminal article, Tanner and Swets (1954, p. 401) explicitly admitted that their theory "appears to be inconsistent with the large quantity of existing data on this subject" and proceeded to criticize the "form of these data."

The Neyman-Pearsonian technique of hypothesis testing was subsequently transformed into a theory of a broad range of cognitive processes, ranging from recognition in memory (e.g., Murdock, 1982; Wickelgren & Norman, 1966) to eyewitness testimony (e.g., Birnbaum, 1983) to discrimination between random and nonrandom patterns (e.g., Lopes, 1982).

My second example concerns theories of causal reasoning. In Europe, Albert Michotte (1946/1963), Jean Piaget (1930), the gestalt psychologists, and others had investigated how certain temporospatial relationships between two or more visual objects, such as moving dots, produced phenomenal causality. For instance, the subjects were made to perceive that one dot launches, pushes, or chases another. After the institutionalization of inferential statistics, Harold H. Kelley (1967) proposed in his "attribution theory" that the long sought laws of causal reasoning are in fact the tools of the behavioral scientist: R. A. Fisher's ANOVA. Just as the experimenter has come to infer a causal relationship between two variables from calculating an ANOVA and performing an F test, the person-in-the-street infers the cause of an effect by unconsciously doing the same calculations. By the time Kelley discovered the new metaphor for causal inference, about 70% of all experimental articles already used ANOVA (Edgington, 1974).

The theory was accepted quickly in social psychology; Kelley and Michaela (1980) reported there were more than 900 references in one decade. The vision of the Fisherian mind radically changed the understanding of causal reasoning, the problems posed to the subjects, and the explanations looked for. I list a few discontinuities that reveal the "fingerprints" of the tool. (a) ANOVA needs repetitions or numbers as data in order to estimate variances and covariances. Consequently, the information presented to the subjects in studies of causal attribution consists of information about the frequency of events (e.g., McArthur, 1972), which played no role in either Michotte's or Piaget's work. (b) Whereas Michotte's work still reflects the broad Aristotelian conception of four causes (see Gavin, 1972), and Piaget (1930) distinguished 17 kinds of causality in children's minds, the Fisherian mind concentrates on the one kind of causes for which ANOVA is used as a tool (similar to Aristotle's "material cause"). (c) In Michotte's view, causal perception is direct and spontaneous and needs no inference, as a consequence of largely innate laws that determine the organization of the perceptual field. ANOVA, in contrast, is used in psychology as a technique for inductive inferences from data to hypotheses, and the focus in Kelley's attribution theory is consequently on the data-driven, inductive side of causal perception.

The latter point illustrates that the specific use of a tool, that is, its practical context rather than its mathematical structure, can also shape theoretical conceptions of mind. To elaborate on this point, assume that Harold Kelley had lived one and a half centuries earlier than he did. In the early 19th century, significance tests (similar to those in ANOVA) were already being used by astronomers (Swijtink, 1987), but they used their tests to reject data, so-called outliers, and not to reject hypotheses. At least provisionally, the astronomers assumed that the theory was correct and mistrusted the data, whereas the ANOVA mind, following the current statistical textbooks, assumes the data to be correct and mistrusts the theories. So, to a 19th-century Kelley, the mind's causal attribution would have seemed expectation driven rather than data driven: The statistician homunculus in the mind would have tested the data and not the hypothesis.

As is well documented, most of causal attribution research after Kelley took the theoretical stand that attribution is a "lay version of experimental design and analysis" (Jones & McGillis, 1976, p. 411), and elaboration of the theory was in part concerned with the kind of intuitive statistics in the brain. For instance, Ajzen and Fishbein (1975) argued that the homunculus statistician is Bayesian rather than Fisherian.

These two areas—detection and discrimination, and causal reasoning—may be sufficient to illustrate some of the fundamental innovations in the explanatory framework, in the research questions posed, and in the kind of data generated. The spectrum of theories that model cognition after statistical inference ranges from auditive and visual perception to recognition in memory and from speech perception to thinking and reasoning. It reaches from the elementary, physiological end to the global, conscious end of the continuum called cognitive. I give one example for each end. (a) Luce (1977) viewed the central nervous system (CNS) as a statistician who draws a random sample from all activated fibers, estimates parameters of the pulse rate, aggregates this estimate into a single number, and

uses a decision criterion to arrive at the final perception. This conception has led to new and interesting questions; for instance, How does the CNS aggregate numbers? and What is the shape of the internal distributions? (b) The 18th-century mathematicians Laplace and Condorcet used their "probability of causes" to model how scientists reason (Daston, 1988). Recently, Massaro (1987) proposed the same statistical formula as an algorithm of pattern recognition, as "a general algorithm, regardless of the modality and particular nature of the patterns" (p. 16).

The degree to which cognitive theories were shaped by the statistical tool varies from theory to theory. On the one hand, there is largely metaphorical use of statistical inference. An example is Gregory's (1974) hypothesis-testing view of perception, in which he reconceptualized Helmholtz's "unconscious inferences" as Fisherian significance testing: "We may account for the stability of perceptual forms by suggesting that there is something akin to statistical significance which must be exceeded by the rival interpretation and the rival hypothesis before they are allowed to supersede the present perceptual hypothesis" (p. 528). In his theory of how perception works, Gregory also explained other perceptual phenomena, using Bayesian and Neyman-Pearsonian statistics as analogies, thus reflecting the actual heterogeneous practice in the social sciences (Gigerenzer & Murray, 1987). Here, a new perspective, but no quantitative model, is generated. On the other hand, there are cognitive theories that propose quantitative models of statistical inference that profoundly transform qualitative concepts and research practice. Examples are the various TSDs of cognition mentioned earlier and the recent theory of adaptive memory as statistical optimization by Anderson and Milson (1989).

To summarize: The tools-to-theories heuristic can account for the discovery and acceptance of a group of cognitive theories in apparently unrelated subfields of psychology, all of them sharing the view that cognitive processes can be modeled by statistical hypothesis testing. Among these are several highly innovative and influential theories that have radically changed psychologists' understanding of what cognitive means.

Before the Institutionalization of Inferential Statistics

There is an important test case for the present hypotheses (a) that familiarity with the statistical tool is crucial to the discovery of corresponding theories of mind and (b) that the institutionalization of the tool within a scientific community is crucial for the broad acceptance of those theories. That test case is the era before the institutionalization of inferential statistics. Theories that conceive of the mind as an intuitive statistician should have a very small likelihood of being discovered and even less likelihood of being accepted. The two strongest tests are cases where (a) someone proposed a similar conceptual analogy and (b) someone proposed a similar probabilistic (formal) model. The chances of theories of the first kind being accepted should be small, and the chances of a probabilistic model being interpreted as "intuitive statistics" should be similarly small. I know of only one case each, which I analyze after defining first what I mean by the phrase "institutionalization of inferential statistics."

Statistical inference has been known for a long time but not used as theories of mind. In 1710, John Arbuthnot proved the existence of God using a kind of significance test; as mentioned earlier, astronomers used significance tests in the 19th century; G. T. Fechner's (1897) statistical text *Kollektivmasslehre* included tests of hypotheses; W. S. Gosset (using the pseudonym Student) published the *t* test in 1908; and Fisher's significance testing techniques, such as ANOVA, as well as Neyman-Pearsonian hypothesis-testing methods have been available since the 1920s (see Gigerenzer et al., 1989). Bayes's theorem has been known since 1763. Nonetheless, there was little interest in these techniques in experimental psychology before 1940 (Rucci & Tweney, 1980).

The statisticians' conquest of new territory in psychology started in the 1940s. By 1942, Maurice Kendall could comment on the statisticians' expansion: "They have already overrun every branch of science with a rapidity of conquest rivalled only by Attila, Mohammed, and the Colorado beetle" (p. 69). By the early 1950s, half of the psychology departments in leading American universities offered courses on Fisherian methods and had made inferential statistics a graduate program requirement. By 1955, more than 80% of the experimental articles in leading journals used inferential statistics to justify conclusions from the data (Sterling, 1959). Editors of major journals made significance testing a requirement for articles submitted and used the level of significance as a yardstick for evaluating the quality of an article (e.g., Melton, 1962).

I therefore use 1955 as a rough date for the institutionalization of the tool in curricula, textbooks, and editorials. What became institutionalized as the logic of statistical inference was a mixture of ideas from two opposing camps, those of R. A. Fisher on the one hand, and Jerzy Neyman and Egon S. Pearson (the son of Karl Pearson) on the other (Gigerenzer & Murray, 1987, chap. 1).

Discovery and Rejection of the Analogy

The analogy between the mind and the statistician was first proposed before the institutionalization of inferential statistics, in the early 1940s, by Egon Brunswik at Berkeley (e.g., Brunswik, 1943). As Leary (1987) has shown, Brunswik's probabilistic functionalism was based on a very unusual blending of scientific traditions, including the probabilistic world view of Hans Reichenbach and members of the Vienna School and Karl Pearson's correlational statistics.

The important point here is that in the late 1930s, Brunswik changed his techniques for measuring perceptual constancies, from calculating (nonstatistical) "Brunswik ratios" to calculating Pearson correlations, such as functional and ecological validities. In the 1940s, he also began to think of the organism as "an intuitive statistician," but it took him several years to spell out the analogy in a clear and consistent way (Gigerenzer, 1987b).

The analogy is this: The perceptual system infers its environment from uncertain cues by (unconsciously) calculating correlation and regression statistics, just as the Brunswikian researcher does when (consciously) calculating the degree of adaptation of a perceptual system to a given environment. Brunswik's intuitive statistician was a statistician of the Karl Pearson school, like the Brunswikian researcher. Brunswik's intuitive statistician was not well adapted to the psychological science of the time, however, and the analogy was poorly understood and generally rejected (Leary, 1987).

Brunswik's analogy came too early to be understood and accepted by his colleagues of the experimental community; it came before the institutionalization of statistics as the indispensable method of scientific inference, and it came with the "wrong" statistical model, correlational statistics. Correlation was an indispensable method not in experimental psychology, but rather in its rival discipline, known as the Galton-Pearson program, or, as Lee Cronbach (1957) put it, the "Holy Roman Empire" of "correlational psychology" (p. 671).

The schism between the two scientific communities had been repeatedly taken up in presidential addresses before the American Psychological Association (Cronbach, 1957; Dashiell, 1939) and had deeply affected the values and the mutual esteem of psychologists (Thorndike, 1954). Brunswik could not persuade his colleagues from the experimental community to consider the statistical tool of the competing community as a model of how the mind works. Ernest Hilgard (1955), in his rejection of Brunswik's perspective, did not mince words: "Correlation is an instrument of the devil" (p. 228).

Brunswik, who coined the metaphor of "man as intuitive statistician," did not survive to see the success of his analogy. It was accepted only after statistical inference became institutionalized in experimental psychology and with the new institutionalized tools rather than (Karl) Pearsonian statistics serving as models of mind. Only in the mid-1960s, however, did interest in Brunswikian models of mind emerge (e.g., Brehmer & Joyce, 1988; Hammond, Stewart, Brehmer, & Steinmann, 1975).

The tendency to accept the statistical tools of one's own scientific community (here, the experimental psychologists) rather than those of a competing community as models of mind is not restricted to Brunswik's case. For example, Fritz Heider (1958, pp. 123, 297), whom Harold Kelley credited for having inspired his ANOVA theory, had repeatedly suggested factor analysis—another indispensable tool of the correlational discipline—as a model of causal reasoning. Heider's proposal met with the same neglect by the American experimental community as did Brunswik's correlational model. Kelley replaced the statistical tool that Heider suggested by ANOVA, the tool of the experimental community. It seems to be more than a mere accident that both Brunswik and Heider came from a similar, German-speaking tradition, where no comparable division into two communities with competing methodological imperatives existed (Leary, 1987).

Probabilistic Models Without the Intuitive Statistician

My preceding point is that the statistical tool was accepted as a plausible analogy of cognitive processes only after its institutionalization in experimental psychology. My second point is that although some probabilistic models of cognitive processes were advanced before the institutionalization of inferential statistics, they were not interpreted using the metaphor of the mind as intuitive statistician. The distinction I draw is between probabilistic models that use the metaphor and ones that do not. The latter kind is illustrated by models that use probability

distributions for perceptual judgment, assuming that variability is caused by lack of experimental control, measurement error, or other factors that can be summarized as experimenter's ignorance. Ideally, if the experimenter had complete control and knowledge (such as Laplace's demon), all probabilistic terms could be eliminated from the theory. This does not hold for a probabilistic model that is based on the metaphor. Here, the probabilistic terms model the ignorance of the mind rather than that of the experimenter. That is, they model how the homunculus statistician in the brain comes to terms with a fundamental uncertain world. Even if the experimenter had complete knowledge, the theories would remain probabilistic because it is the mind that is ignorant and needs statistics.

The key example is represented in L. L. Thurstone, who in 1927 formulated a model for perceptual judgment that was formally equivalent to the present-day TSD. But neither Thurstone nor his followers recognized the possibility of interpreting the formal structure of their model in terms of the intuitive statistician. Like TSD, Thurstone's model had two overlapping normal distributions, which represented the internal values of two stimuli and which specified the corresponding likelihood ratios, but it never occurred to Thurstone to include in his model the conscious activities of a statistician, such as the weighing of the costs of the two errors and the setting of a decision criterion. Thus, neither Thurstone nor his followers took the—with hindsight—small step to develop the "law of comparative judgment" into TSD. When Duncan Luce (1977) reviewed Thurstone's model 50 years later, he found it hard to believe that nothing in Thurstone's writings showed the least awareness of this small but crucial step. Thurstone's perceptual model remained a mechanical, albeit probabilistic, stimulus–response theory without a homunculus statistician in the brain. The small conceptual step was never taken, and TSD entered psychology by an independent route.

To summarize: There are several kinds of evidence for a close link between the institutionalization of inferential statistics in the 1950s and the subsequent broad acceptance of the metaphor of the mind as an intuitive statistician: (a) the general failure to accept, and even to understand, Brunswik's intuitive statistician before the institutionalization of the tool and (b) the case of Thurstone, who proposed a probabilistic model that was formally equivalent to one important present-day theory of intuitive statistics but was never interpreted in this way; the analogy was not yet seen. Brunswik's case illustrates that tools may act on two levels: First, new tools may suggest new cognitive theories to a scientist. Second, the degree to which these tools are institutionalized within the scientific community to which the scientist belongs can prepare (or hinder) the acceptance of the new theory. This close link between tools for justification on the one hand and discovery and acceptance on the other reveals the artificiality of the discovery–justification distinction. Discovery does not come first and justification afterward. Discovery is *inspired by* justification.

How Heuristics of Discovery May Help in Understanding Limitations and Possibilities of Current Research Programs

In this section I argue that the preceding analysis of discovery is of interest not only for a psychology of scientific discovery

and creativity (e.g., Gardner, 1988; Gruber, 1981; Tweney, Dotherty, & Mynatt, 1981) but also for the evaluation and further development of current cognitive theories. The general point is that institutionalized tools like statistics do not come as pure mathematical (or physical) systems, but with a practical context attached. Features of this context in which a tool has been used may be smuggled Trojan horse fashion into the new cognitive theories and research programs. One example was mentioned earlier: The formal tools of significance testing have been used in psychology as tools for rejecting *hypotheses,* with the assumption that the data are correct, whereas in other fields and at other times the same tools were used as tools for rejecting *data* (outliers), with the assumption that the hypotheses were correct. The latter use of statistics is practically extinct in experimental psychology (although the problem of outliers routinely emerges), and therefore also absent in theories that liken cognitive processes to significance testing. In cases like these, analysis of discovery may help to reveal blind spots associated with the tool and as a consequence, new possibilities for cognitive theorizing.

I illustrate this potential in more detail using examples from the "judgment under uncertainty" program of Daniel Kahneman, Amos Tversky, and others (see Kahneman, Slovic, & Tversky, 1982). This stimulating research program emerged from the earlier research on human information processing by Ward Edwards and his coworkers. In Edwards's work, the dual role of statistics as a tool and a model of mind is again evident: Edwards, Lindman, and Savage (1963) proposed Bayesian statistics for scientific hypothesis evaluation and considered the mind as a reasonably good, albeit conservative, Bayesian statistician (e.g., Edwards, 1966). The judgment-under-uncertainty program that also investigates reasoning as intuitive statistics but focuses on so-called errors in probabilistic reasoning. In most of the theories based on the metaphor of the intuitive statistician, statistics or probability theory is used both as normative and as descriptive of a cognitive process (e.g., both as the optimal and the actual mechanism for speech perception and human memory; see Massaro, 1987, and Anderson & Milson, 1989, respectively). This is not the case in the judgment-under-uncertainty program; here, statistics and probability theory are used only in the normative function, whereas actual human reasoning has been described as "biased," "fallacious," or "indefensible" (on the rhetoric, see Lopes, 1991).

In the following, I first point out three features of the practical *use* of the statistical tool (as opposed to the mathematics). Then I show that these features reemerge in the judgment-under-uncertainty program, resulting in severe limitations on that program. Finally, I suggest how this hidden legacy of the tool could be eliminated to provide new impulses and possibilities for the research program.

The first feature is an assumption that can be called "There is only one statistics." Textbooks on statistics for psychologists (usually written by nonmathematicians) generally teach statistical inference as if there existed only one logic of inference. Since the 1950s and 1960s, almost all texts teach a mishmash of R. A. Fisher's ideas tangled with those of Jerzy Neyman and Egon S. Pearson, but without acknowledgment. The fact that Fisherians and Neyman-Pearsonians could never agree on a logic of statistical inference is not mentioned in the textbooks, nor are the

controversial issues that divide them. Even alternative statistical logics for scientific inference are rarely discussed (Gigerenzer, 1987a). For instance, Fisher (1955) argued that concepts like Type II error, power, the setting of a level of significance before the experiment, and its interpretation as a long-run frequency of errors in repeated experiments, are concepts inappropriate for scientific inference—at best they could be applied to technology (his pejorative example was Stalin's). Neyman, for his part, declared that some of Fisher's significance tests are "worse than useless" (because their power is less than their size; see Hacking, 1965, p. 99). I know of no textbook written by psychologists for psychologists that mentions and explains this and other controversies about the logic of inference. Instead, readers are presented with an intellectually incoherent mix of Fisherian and Neyman-Pearsonian ideas, but a mix presented as a seamless, uncontroversial whole: *the* logic of scientific inference (for more details, see Gigerenzer et al., 1989, chap. 3 and 6; Gigerenzer & Murray, 1987, chap. 1).

The second assumption that became associated with the tool during its institutionalization is "There is only one meaning of probability." For instance, Fisher and Neyman-Pearson had different interpretations of what a level of significance means. Fisher's was an epistemic interpretation, that is, that the level of significance indicates the confidence that can be placed in the *particular* hypothesis under test, whereas Neyman's was a strictly frequentist and behavioristic interpretation, which claimed that a level of significance does not refer to a particular hypothesis, but to the relative frequency of wrongly rejecting the null hypothesis if it is true in the long run. Although the textbooks teach both Fisherian and Neyman-Pearsonian ideas, these alternative views of what a probability (such as a level of significance) could mean are generally neglected—not to speak of the many other meanings that have been proposed for the formal concept of probability.

Third and last, the daily practice of psychologists assumes that statistical inference can be applied mechanically without checking the underlying assumptions of the model. The importance of checking whether the assumptions of a particular statistical model hold in a given application has been repeatedly emphasized, particularly by statisticians. The general tendency in psychological practice (and other social sciences) has been to apply the test anyhow (Oakes, 1986), as a kind of ritual of justification required by journals, but poorly understood by authors and readers alike (Sedlmeier & Gigerenzer, 1989).

These features of the practical context, in which the statistical tool has been used, reemerge at the theoretical level in current cognitive psychology, just as the tools-to-theories heuristic would lead one to expect.

Example 1: There Is Only One Statistics, Which Is Normative

Tversky and Kahneman (1974) described their judgment-under-uncertainty program as a two-step procedure. First, subjects are confronted with a reasoning problem, and their answers are compared with the so-called normative or correct answer, supplied by statistics and probability theory. Second, the *deviation* between the subject's answer and the so-called

normative answer, also called a *bias of reasoning,* is explained by some heuristic of reasoning.

One implicit assumption at the heart of this research program says that statistical theory provides exactly one answer to the real-world problems presented to the subjects. If this were not true, the deviation between subjects' judgments and the "normative" answer would be an inappropriate *explanandum,* because there are as many different deviations as there are statistical answers. Consider the following problem:

> A cab was involved in a hit-and-run accident at night. Two companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (i) 85% of the cabs in the city are Green and 15% are Blue. (ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green), the witness made correct identifications in 80% of the cases and erred in 20% of the cases.
>
> Question: What is the probability that the cab involved in the accident was Blue rather than Green? (Tversky & Kahneman, 1980, p. 62)

The authors inserted the values specified in this problem into Bayes's formula and calculated a probability of .41 as the "correct" answer; and, despite criticism, they have never retreated from that claim. They saw in the difference between this value and the subjects' median answer of .80 an instance of a reasoning error, known as *neglect of base rates.* But alternative statistical solutions to the problem exist.

Tversky and Kahneman's reasoning is based on one among many possible Bayesian views—which the statistician I. J. Good (1971), not all too seriously, once counted up to 46,656. For instance, using the classical principle of indifference to determine the Bayesian prior probabilities can be as defensible as Tversky and Kahneman's use of base rates of "cabs in the city" for the relevant priors, but it leads to a probability of .80 instead of .41 (Levi, 1983). Or, if Neyman-Pearson theory is applied to the cab problem, solutions range between .28 and .82, depending on the psychological theory about the witness's criterion shift—the shift from witness testimony at the time of the accident to witness testimony at the time of the court's test (Birnbaum, 1983; Gigerenzer & Murray, 1987, pp. 167–174).

There may be more arguable answers to the cab problem, depending on what statistical or philosophical theory of inference one uses. Indeed, the range of possible statistical solutions is about the range of subjects' actual answers. The point is that none of these statistical solutions is the *only* correct answer to the problem, and therefore it makes little sense to use the deviation between a subject's judgment and *one* of these statistical answers as the psychological explanandum.

Statistics is an indispensable tool for scientific inference, but, as Neyman and Pearson (1928, p. 176) pointed out, in "many cases there is probably no single best method of solution." Rather, several such theories are legitimate, just as "Euclidean and non-Euclidean geometrics are equally legitimate" (Neyman, 1937, p. 336). My point is this: The idée fixe that statistics speaks with one voice has reappeared in research on intuitive statistics. The highly interesting judgment-under-uncertainty program could progress beyond the present point if (a) subjects' judgments rather than deviations between judgments and a so-

called normative solution are considered as the data to be explained and if (b) various statistical models are proposed as competing hypotheses of problem-solving strategies rather than one model being proposed as *the* general norm for rational reasoning. The willingness of many researchers to accept the claim that statistics speaks with one voice is the legacy of the institutionalized tool, not of statistics per se.

Note the resulting double standard: Many researchers on intuitive statistics argue that their subjects should draw inferences from data to hypotheses by using Bayes's formula, although they themselves do not. Rather, the researchers use the institutionalized mixture of Fisherian and Neyman-Pearsonian statistics to draw their inferences from data to hypotheses.

Example 2: There Is Only One Interpretation of Probability

Just as there are alternative logics of inference, there are alternative interpretations of probability that have been part of the mathematical theory since its inception in the mid-17th century (Daston, 1988; Hacking, 1975). Again, both the institutionalized tool and the recent cognitive research on probabilistic reasoning exhibit the same blind spot concerning the existence of alternative interpretations of probability. For instance, Lichtenstein, Fischhoff, and Phillips (1982) have reported and summarized research on a phenomenon called *overconfidence.* Briefly, subjects were given questions such as "Absinthe is (a) a precious stone or (b) a liqueur"; they chose what they believed was the correct answer and then were asked for a confidence rating in their answer, for example, 90% certain. When people said they were 100% certain about individual answers, they had in the long run only about 80% correct answers; when they were 90% certain, they had in the long run only 75% correct answers; and so on. This discrepancy was called overconfidence bias and was explained by general heuristics in memory search, such as confirmation biases, or general motivational tendencies, such as a so-called illusion of validity.

My point is that two different interpretations of probability are compared: degrees of belief in single events (i.e., that *this* answer is correct) and relative frequencies of correct answers in the long run. Although 18th-century mathematicians, like many of today's cognitive psychologists, would have had no problem in equating the two, most mathematicians and philosophers since then have. For instance, according to the frequentist point of view, the term *probability,* when it refers to a single event, "has no meaning at all" (Mises, 1957, p. 11) because probability theory is about relative frequencies in the long run. Thus, for a frequentist, probability theory does not apply to single-event confidences, and therefore no confidence judgment can violate probability theory. To call a discrepancy between confidence and relative frequency a *bias* in probabilistic reasoning would mean comparing apples and oranges. Moreover, even subjectivists would not generally think of a discrepancy between confidence and relative frequency as a bias (see Kadane & Lichtenstein, 1982, for a discussion of conditions). For a subjectivist such as Bruno de Finetti, probability *is* about single events, but rationality is identified with the internal consistency of probability judgments. As de Finetti (1931/1989, p. 174) emphasized: "However an individual evaluates the proba-

bility of a particular event, no experience can prove him right, or wrong; nor in general, could any conceivable criterion give any objective sense to the distinction one would like to draw, here, between right and wrong."

Nonetheless, the literature on overconfidence is largely silent on even the possibility of this conceptual problem (but see Keren, 1987). The question about research strategy is whether to use the deviation between degrees of belief and relative frequencies (again considered as a bias) as the *explanandum* or to accept the existence of several meanings of probability and to investigate the kind of conceptual distinctions that untutored subjects make. Almost all research has been done within the former research strategy. And, indeed, if the issue were a general tendency to overestimate one's knowledge, as the term *overconfidence* suggests—for instance, as a result of general strategies of memory search or motivational tendencies—then asking the subject for degrees of belief or for frequencies should not matter.

But it does. In a series of experiments (Gigerenzer, Hoffrage, & Kleinbölting, in press; see also May, 1987) subjects were given several hundred questions of the absinthe type and were asked for confidence judgments after every question was answered (as usual). In addition, after each 50 (or 10, 5, and 2) questions, the subjects were asked *how many* of those questions they believed they had answered correctly; that is, frequency judgments were requested. This design allowed comparison both between their confidence in their individual answers and true relative frequencies of correct answers, and between judgments of relative frequencies and true relative frequencies. Comparing frequency judgments with the true frequency of correct answers showed that overestimation or overconfidence disappeared in 80% to 90% of the subjects, depending on experimental conditions. Frequency judgments were precise or even showed underestimation. Ironically, after each frequency judgment, subjects went on to give confidence judgments (degrees of belief) that exhibited what has been called overconfidence.

As in the preceding example, a so-called bias of reasoning disappears if a controversial norm is dropped and replaced by several descriptive alternatives, statistical models, and meanings of probability, respectively. Thus, probabilities for single events and relative frequencies seem to refer to different meanings of confidence in the minds of the subjects. This result is inconsistent with previous explanations of the alleged bias by deeper cognitive deficiencies (e.g., confirmation biases) and has led to the theory of probabilistic mental models, which describes mechanisms that generate different confidence and frequency judgments (Gigerenzer et al., in press). Untutored intuition seems to be capable of making conceptual distinctions of the sort statisticians and philosophers make (e.g., Cohen, 1986; Lopes, 1981; Teigen, 1983). And it suggests that the important research questions to be investigated are How are different meanings of probability cued in everyday language? and How does this affect judgment?, rather than How can the alleged bias of overconfidence be explained by some general deficits in memory, cognition, or personality?

The same conceptual distinction can help to explain other kinds of judgments under uncertainty. For instance, Tversky and Kahneman (1982, 1983) used a personality sketch of a character named Linda that suggested she was a feminist. Subjects

were asked which is more probable: (a) that Linda is a bank teller or (b) that Linda is a bank teller and active in the feminist movement. Most subjects chose Alternative b, which Tversky and Kahneman (1982) called a "fallacious" belief, to be explained by their hypothesis that people use a limited number of heuristics—in the present case, representativeness (the similarity between the description of Linda and the alternatives a and b). Subjects' judgments were called a *conjunction fallacy* because the probability of a conjunction of events (bank teller and active in the feminist movement) cannot be greater than the probability of one of its components.

As in the example just given, this normative interpretation neglects two facts. First, in everyday language, words like *probable* legitimately have several meanings, just as "if. . . then" and "or" constructions do. The particular meaning seems to be automatically cued by content and context. Second, statisticians similarly have alternative views of what probability is about. In the context of some subjectivist theories, choosing Alternative b truly violates the rules of probability; but for a frequentist, judgments of single events such as in the Linda problem have nothing to do with probability theory: As the statistician G. A. Barnard (1979, p. 171) objected, they should be treated in the context of psychoanalysis, not probability.

Again, the normative evaluation explicit in the term *conjunction fallacy* is far from being uncontroversial, and progress in understanding reasoning may be expected by focusing on subjects' judgments as *explanandum* rather than on their deviations from a so-called norm. As in the previous example, if problems of the Linda type are rephrased as involving frequency judgments (e.g., "How many out of 100 cases that fit the description of Linda are [a] bank tellers and [b] bank tellers and active in the feminist movement?"), then the so-called conjunction fallacy decreases from 77% to 27%, as Fiedler (1988) showed. "Which alternative is more probable?" is not the same as "Which alternative is more frequent?" in the Linda context. Tversky and Kahneman (1983) found similar results, but they maintained their normative claims and treated the disappearance of the phenomenon merely as an exception to the rule (p. 293).

Example 3: Commitment to Assumptions Versus Neglect of Them

It is a commonplace that the validity of a statistical inference is to be measured against the validity of the assumptions of the statistical model for a given situation. In the actual context of justification, however, in psychology and probably beyond, there is little emphasis on pointing out and checking crucial assumptions. The same neglect is a drawback in some Bayesian-type probability revision studies. Kahneman and Tversky's (1973) famous engineer–lawyer study is a case in point. In the study, a group of students was told that a panel of psychologists had made personality descriptions of 30 engineers and 70 lawyers, that they (the students) would be given 5 of these descriptions, chosen at random, and that their task was to estimate for each description the probability that the person described was an engineer. A second group received the same instruction and the same descriptions, but was given inverted base rates, that is, 70 engineers and 30 lawyers. Kahneman and Tversky found

that the mean probabilities were about the same in the two groups and concluded that base rates were ignored. They explained this alleged bias in reasoning by postulating that people use a general heuristic, called *representativeness*, which means that people generally judge the posterior probability simply by the similarity between a description and their stereotype of an engineer.

Neither Kahneman and Tversky's (1973) study nor any of the follow-up studies checked whether the subjects were committed to or were aware of a crucial assumption that must hold in order to make the given base rates relevant: the assumption that the descriptions have been randomly drawn from the population. If not, the base rates are irrelevant. There have been studies, like Kahneman and Tversky's (1973) "Tom W." study, where subjects were not even told whether the descriptions were randomly sampled. In the engineer–lawyer study, subjects were so informed (in only one word), but the information was false. Whether a single word is sufficient to direct the attention of subjects toward this crucial information is an important question in itself, because researchers cannot assume that in everyday life, people are familiar with situations in which profession guessing is about randomly selected people. Thus, many of the subjects may not have been committed to the crucial assumption of random selection.

In a controlled replication (Gigerenzer, Hell, & Blank, 1988), a simple method was used to make subjects aware of this crucial assumption: Subjects themselves drew each description (blindly) out of an urn and gave their probability judgments. This condition made base rate neglect disappear; once the subjects were committed to the crucial assumption of random sampling, their judgments were closer to Bayesian predictions than to base rate neglect. This finding indicates that theories of intuitive statistics have to deal with how the mind analyzes the structure of a problem (or environment) and how it infers the presence or absence of crucial statistical assumptions—just as the practicing statistician has to first check the structure of a problem in order to decide whether a particular statistical model can be applied. Checking structural assumptions precedes statistical calculations (see also Cohen, 1982, 1986; Einhorn & Hogarth, 1981; Ginossar & Trope, 1987).

My intention here is not to criticize this or that specific experiment, but rather to draw attention to the hidden legacy that tools bequeath to theories. The general theme is that some features of the practical context in which a tool has been used (to be distinguished from its mathematics) have reemerged and been accepted in a research program that investigates intuitive statistics, impeding progress. Specifically, the key problem is a simplistic conception of normativeness that confounds one view about probability with the criterion for rationality.

Although I have dwelt on the dangerous legacy that tools hand on to theories, I do not mean to imply that a theory that originates in a tool is ipso facto a bad theory. The history of science, not just the history of psychology, is replete with examples to the contrary. Good ideas are hard to come by, and one should be grateful for those few that one has, whatever their lineage. But knowing that lineage can help to refine and criticize the new ideas. In those cases where the tools-to-theories heuristic operates, this means taking a long, hard look at the

tools—and the statistical tools of psychologists are overdue for such a skeptical inspection.

Discussion

New technologies have been a steady source of metaphors of mind: "In my childhood we were always assured that the brain was a telephone switchboard. ('What else could it be?')," recalled John Searle (1984, p. 44). The tools-to-theories heuristic is more specific than general technology metaphors. Scientists' tools, not just any tools, are used to understand the mind. Holograms are not social scientists' tools, but computers are, and part of their differential acceptance as metaphors of mind by the psychological community may be a result of psychologists' differential familiarity with these devices in research practice.

The computer, serial and parallel, would be another case study for the tools-to-theories heuristic—a case study that is in some aspects different. For instance, John von Neumann (1958) and others explicitly suggested the analogy between the serial computer and the brain. But the main use of computers in psychological science was first in the context of justification: for processing data; making statistical calculations; and as an ideal, endlessly patient experimental subject. Recently, the computer metaphor and the statistics metaphors of mind have converged, both in artificial intelligence and in the shift toward massively parallel computers simulating the interaction between neurons.

Herbert A. Simon's Heuristics of Discovery and the Tools-to-Theories Heuristic

Recently, in the work of Herbert A. Simon (1973) and his co-workers (e.g., Langley, Simon, Bradshaw, & Zytkow, 1987), the possibility of a logic of discovery has been explicitly reconsidered. For example, a series of programs called BACON has "rediscovered" quantitative empirical laws, such as Kepler's third law of planetary motion. How does BACON discover a law? Basically, BACON starts from data and analyzes them by applying a group of heuristics until a simple quantitative law can be fitted to the data. Kepler's law, for instance, can be rediscovered by using heuristics such as "If the values of two numerical terms increase together, then consider their ratio" (Langley et al., 1987, p. 66). Such heuristics are implemented as production rules.

What is the relation between heuristics used in programs like BACON and the tools-to-theories heuristics? First, the research on BACON was concerned mainly with the ways in which laws could be induced from data. BACON's heuristics work on extant data, whereas the tools-to-theories heuristic works on extant tools for data generation and processing and describes an aspect of discovery (and acceptance) that goes beyond data. As I argued earlier, new data can be a consequence of the tools-to-theories heuristic, rather than the starting point to which it is applied. Second, what can be discovered seems to have little overlap. For Langley et al. (1987), discoveries are of two major kinds: quantitative laws such as Kepler's law and qualitative laws such as taxonomies using clustering methods. In fact, the heuristics of discovery proposed in that work are similar to the statistical methods of exploratory data analysis (Tukey, 1977). It

is this kind of intuitive statistics that serves as the analogy to discovery in Simon's approach. In contrast, the tools-to-theories heuristic can discover new conceptual analogies, new research programs, and new data. It cannot—at least not directly—derive quantitative laws by summarizing data, as BACON's heuristics can.

The second issue, What can be discovered?, is related to the first, that is, to Simon's approach to discovery as induction from *data*, as "recording in a parsimonious fashion, sets of empirical data" (Simon, 1973, p. 475). More recently, Simon and Kulkarni (1988) went beyond that data-centered view of discovery and made a first step toward characterizing the heuristics used by scientists for planning and guiding experimental research. Although Simon and Kulkarni did not explore the potential of scientists' tools for suggesting theoretical concepts (and their particular case study may not invite this), the tools-to-theories heuristic can complement this recent, broader program to understand discovery. Both Simon's heuristics and the tools-to-theories heuristic go beyond the inductive probability approach to discovery (such as Reichenbach's). The approaches are complementary in their focus on aspects of discovery, but both emphasize the possibility of understanding discovery by reference to heuristics of creative reasoning, which go beyond the merely personal and accidental.

The Tools-to-Theories Heuristic Beyond Cognitive Psychology

The examples of discovery I give in this article are modest instances compared with the classical literature in the history of science treating the contribution of a Copernicus or a Darwin. But in the narrower context of recent cognitive psychology, the theories I have discussed count as among the most influential. In this more prosaic context of discovery, the tools-to-theories heuristic can account for a group of significant theoretical innovations. And, as I have argued, this discovery heuristic can both open and foreclose new avenues of research, depending on the interpretations attached to the statistical tool. My focus is on analytical tools of justification, and I have not dealt with physical tools of experimentation and data processing. Physical tools, once familiar and considered indispensable, also may become the stuff of theories. This holds not only for the hardware (like the software) of the computer, but also for theory innovation beyond recent cognitive psychology. Smith (1986) argued that Edward C. Tolman's use of the maze as an experimental apparatus transformed Tolman's conception of purpose and cognition into spatial characteristics, such as cognitive maps. Similarly, he argued that Clark L. Hull's fascination with conditioning machines has shaped Hull's thinking of behavior as if it were machine design. With the exception of Danziger's (1985, 1987) work on changing methodological practices in psychology and their impact on the kind of knowledge produced, however, there seems to exist no systematic research program on the power of familiar tools to shape new theories in psychology.

But the history of science beyond psychology provides some striking instances of scientists' tools, both analytical and physical, that ended up as theories of nature. Hackmann (1979), Lenoir (1986), and Wise (1988) have explored how scientific

instruments shaped the theoretical concepts of, among others, Emil DuBois-Reymond and William Thomson (Lord Kelvin).

The case of Adolphe Quetelet illustrates nicely how the tools-to-theories heuristic can combine with an interdisciplinary exchange of theories. The statistical error law (normal distribution) was used by astronomers to handle observational errors around the true position of a star. Quetelet (1942/1969), who began as an astronomer, transformed the astronomer's tool for taming error into a theory about society: The true position of a star turned into *l'homme moyen,* or the ideal average person within a society, and observational errors turned into the distribution of actual persons (with respect to any variable) around *l'homme moyen*—actual persons now being viewed as nature's errors. Quetelet's social error theory was in turn seminal in the development of statistical mechanics; Ludwig Boltzmann and James Clerk Maxwell in the 1860s and 1870s reasoned that gas molecules might behave as Quetelet's humans do; erratic and unpredictable as individuals, but regular and predictable when considered as a collective (Porter, 1986). By this strange route of discovery—from astronomer's tool to a theory of society, and from a theory of society to a theory of a collective of gas molecules—the deterministic Newtonian view of the world was finally overthrown and replaced by a statistical view of nature (see Gigerenzer et al., 1989). Thus, there seems to exist a broader, interdisciplinary framework for the tools-to-theories heuristic proposed here, which has yet to be explored.

Discovery Reconsidered

Let me conclude with some reflections on how the present view stands in relation to major themes in scientific discovery.

Data-to-theories reconsidered. Should psychologists continue to tell their students that new theories originate from new data? If only because "little is known about how theories come to be created," as J. R. Anderson introduced the reader to his *Cognitive Psychology* (1980, p. 17)? Holton (1988) noted the tendency among physicists to reconstruct discovery with hindsight as originating from new data, even if this is not the case. His most prominent example is Einstein's special theory of relativity, which was and still is celebrated as an empirical generalization from Michelson's experimental data by such eminent figures as R. A. Millikan and H. Reichenbach, as well as by the textbook writers. As Holton demonstrated with firsthand documents, the role of Michelson's data in the discovery of Einstein's theory was slight, a conclusion shared by Einstein himself.

Similarly, with respect to more modest discoveries, I argue that a group of recent cognitive theories did not originate from new data, but in fact often created new kinds of data. Tanner and Swets (1954) are even explicit that their theory was inconsistent with the extant data. Numerical probability judgments have become the stock-in-trade data of research on inductive thinking since Edwards's (1966) work, whereas this kind of dependent variable was still unknown in Humphrey's (1951) review of research on thinking.

The strongest claim for an inductive view of discovery came from the Vienna Circle's emphasis on sensory data (reduced to the concept of "pointer readings"). Carnap (1928/1969), Reichenbach (1938), and others focused on what they called the *rational reconstruction* of actual discovery rather than on actual

discovery itself, in order to screen out the merely irrational and psychological. For instance, Reichenbach reconstructed Einstein's special theory of relativity as being "suggested by closest adherence to experimental facts," a claim that Einstein rejected, as mentioned earlier (see Holton, 1988, p. 296). It seems fair to say that all attempts to logically reconstruct discovery in science have failed in practice (Blackwell, 1983, p. 111). The strongest theoretical disclaimer concerning the possibility of a logic of discovery came from Popper, Hempel, and other proponents of the hypotheticodeductive account, resulting in the judgment that discovery, not being logical, occurs irrationally. Theories are simply "guesses guided by the unscientific" (Popper, 1959, p. 278). In contrast, I have dealt with guesses that are guided by the scientific, by tools of justification. Induction from data and irrational guesses are not exhaustive of scientific discovery, and the tools-to-theories heuristic explores the field beyond.

Scientists' practice reconsidered. The tools-to-theories heuristic is about scientists' practice, that is, the analytical and physical tools used in the conduct of experiments. This practice has a long tradition of neglect. The very philosophers who called themselves logical empiricists had, ironically, little interest in the empirical practice of scientists. Against their reduction of observation to pointer reading, Kuhn (1970) emphasized the theory ladenness of observation. Referring to perceptual experiments and gestalt switches, he said: "Scientists see new and different things when looking with familiar instruments in places they have looked before." (p. 111). Both the logical empiricists and Kuhn were highly influential on psychology (see Toulmin & Leary, 1985), but neither's view has emphasized the role of tools and experimental conduct. Their role in the development of science has been grossly underestimated until recently (Danziger, 1985; Lenoir, 1988).

Through the lens of theory, it has been said, growth of knowledge can be understood. But there is a recent move away from a theory-dominated account of science that pays attention to what really happens in the laboratories. Hacking (1983) argued that experimentation has a life of its own and that not all observation is theory laden. Galison (1987) analyzed modern experimental practice, such as in high-energy physics, focusing on the role of the fine-grained web of instruments, beliefs, and practice that determine when a fact is considered to be established and when experiments end. Both Hacking and Galison emphasized the role of the familiarity experimenters have with their tools, and the importance and relative autonomy of experimental practice in the quest for knowledge. This is the broader context in which the present tools-to-theories heuristic stands: the conjecture that theory is inseparable from instrumental practices.

In conclusion, my argument is that discovery in recent cognitive psychology can be understood beyond mere inductive generalizations or lucky guesses. More than that, I argue that for a considerable group of cognitive theories, neither induction from data nor lucky guesses played an important role. Rather, these innovations in theory can be accounted for by the tools-to-theories heuristic. So can conceptual problems and possibilities in current theories. Scientists' tools are not neutral. In the present case, the mind has been recreated in their image.

References

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin, 82,* 261–277.

Anderson, J. R. (1980). *Cognitive psychology and its implications.* San Francisco: Freeman.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96,* 703–719.

Barnard, G. A. (1979). Discussion of the paper by Professors Lindley and Tversky and Dr. Brown. *Journal of the Royal Statistical Society, Series A, 142,* 171–172.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85–94.

Blackwell, R. J. (1983). Scientific discovery: The search for new categories. *New Ideas in Psychology, 1,* 111–115.

Brehmer, B., & Joyce, C. R. B. (Eds.). (1988). *Human judgment: The SJT view.* Amsterdam: North-Holland.

Brown, W., & Thomson, G. H. (1921). *The essentials of mental measurement.* Cambridge, England: Cambridge University Press.

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

Carnap, R. (1969). *The logical structure of the world* (R. A. George, Trans.). Berkeley: University of California Press. (Original work published 1928)

Cohen, L. J. (1982). Are people programmed to commit fallacies? Further thoughts about the interpretation of experimental data on probability judgment. *Journal for the Theory of Social Behaviour, 12,* 251–274.

Cohen, L. J. (1986). *The dialogue of reason.* Oxford, England: Clarendon Press.

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

Curd, M. (1980). The logic of discovery: An analysis of three approaches. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 201–219). Dordrecht, The Netherlands: Reidel.

Danziger, K. (1985). The methodological imperative in psychology. *Philosophy of the Social Sciences, 15,* 1–13.

Danziger, K. (1987). Statistical method and the historical development of research practice in American psychology. In L. Krüger, G. Gigerenzer, & M. S. Morgan (Eds.), *The probabilistic revolution: Vol. 2. Ideas in the sciences* (pp. 35–47). Cambridge, MA: MIT Press.

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

Dashiell, J. F. (1939). Some rapprochements in contemporary psychology. *Psychological Bulletin, 36,* 1–24.

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

de Finetti, B. (1989). Probabilism. *Erkenntnis, 31,* 169–223. (Original work published 1931)

Edgington, E. S. (1974). A new tabulation of statistical procedures used in APA journals. *American Psychologist, 29,* 25–26.

Edwards, W. (1966). *Nonconservative information processing systems* (Rep. No. 5893-22-F). Ann Arbor: University of Michigan, Institute of Science and Technology.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32,* 53–88.

Fechner, G. T. (1897). *Kollektivmasslehre* [The measurement of collectivities]. Leipzig: W. Engelmann.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50,* 123–129.

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

Galison, P. (1987). *How experiments end.* Chicago: Chicago University Press.

Gardner, H. (1985). *The mind's new science.* New York: Basic Books.

Gardner, H. (1988). Creative lives and creative works: A synthetic scientific approach. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 298–321). Cambridge, England: Cambridge University Press.

Gavin, E. A. (1972). The causal issue in empirical psychology from Hume to the present with emphasis upon the work of Michotte. *Journal of the History of the Behavioral Sciences, 8,* 302–320.

Gigerenzer, G. (1987a). Probabilistic thinking and the fight against subjectivity. In L. Krüger, G. Gigerenzer, & M. S. Morgan (Eds.), *The probabilistic revolution: Vol. 2. Ideas in the sciences* (pp. 11–33). Cambridge, MA: MIT Press.

Gigerenzer, G. (1987b). Survival of the fittest probabilist: Brunswik, Thurstone, and the two disciplines of psychology. In L. Krüger, G. Gigerenzer, & M. S. Morgan (Eds.), *The probabilistic revolution: Vol. 2. Ideas in the sciences* (pp. 49–72). Cambridge, MA: MIT Press.

Gigerenzer, G. (1988). Woher kommen Theorien über kognitive Prozesse? [Where do cognitive theories come from?]. *Psychologische Rundschau, 39,* 91–100.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513–525.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (in press). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review.*

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

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: Cambridge University Press.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52,* 464–474.

Good, I. J. (1971). 46656 varieties of Bayesians. *The American Statistician, 25,* 62–63.

Gregory, R. L. (1974). *Concepts and mechanisms of perception.* New York: Scribner.

Gruber, H. E. (1981). *Darwin on man: A psychological study of scientific creativity* (2nd ed.). Chicago: University of Chicago Press.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Hacking, I. (1965). *Logic of statistical inference.* Cambridge, England: Cambridge University Press.

Hacking, I. (1975). *The emergence of probability.* Cambridge, England: Cambridge University Press.

Hacking, I. (1983). *Representing and intervening.* Cambridge, England: Cambridge University Press.

Hackmann, W. D. (1979). The relationship between concept and instrument design in eighteenth-century experimental science. *Annals of Science, 36,* 205–224.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. (1975). Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes* (pp. 271–312). New York: Academic Press.

Hanson, N. R. (1958). *Patterns of discovery.* Cambridge, England: Cambridge University Press.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Herbart, J. F. (1834). *A textbook in psychology* (M. K. Smith, Trans.). New York: Appleton.

Hilgard, E. R. (1955). Discussion of probabilistic functionalism. *Psychological Review, 62,* 226–228.

Holton, G. (1988). *Thematic origins of scientific thought* (2nd ed.). Cambridge, MA: Harvard University Press.

Humphrey, G. (1951). *Thinking.* New York: Wiley.

Jones, E. E., & McGillis, D. (1976). Correspondent inferences and the attribution cube: A comparative reappraisal. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 389–420). Hillsdale, NJ: Erlbaum.

Kadane, J. B., & Lichtenstein, S. (1982). *A subjectivist view of calibration* (Decision Research Report No. 82-6). Eugene, OR: Decision Research, A Branch of Perceptronics.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

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

Kelley, H. H., & Michaela, I. L. (1980). Attribution theory and research. *Annual Review of Psychology, 31,* 457–501.

Kendall, M. G. (1942). On the future of statistics. *Journal of the Royal Statistical Society, 105,* 69–80.

Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes, 39,* 98–114.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery.* Cambridge, MA: MIT Press.

Leary, D. E. (1987). From act psychology to probabilistic functionalism: The place of Egon Brunswik in the history of psychology. In M. G. Ash & W. R. Woodward (Eds.), *Psychology in twentieth-century thought and society* (pp. 115–142). Cambridge, England: Cambridge University Press.

Leibniz, G. W. von. (1951). The horizon of human doctrine. In P. P. Wiener (Ed.), *Selections* (pp. 73–77). New York: Scribner. (Original work published after 1690)

Lenoir, T. (1986). Models and instruments in the development of electrophysiology, 1845–1912. *Historical Studies in the Physical and Biological Sciences, 17,* 1–54.

Lenoir, T. (1988). Practice, reason, context: The dialogue between theory and experiment. *Science in Context, 2,* 3–22.

Levi, I. (1983). Who commits the base rate fallacy? *Behavioral and Brain Sciences, 6,* 502–506.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, England: Cambridge University Press.

Lopes, L. L. (1981). Decision making in the short run. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 377–385.

Lopes, L. L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 626–636.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory and Psychology, 1,* 65–82.

Lovie, A. D. (1983). Attention and behaviourism—Fact and fiction. *British Journal of Psychology, 74,* 301–310.

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

Massaro, D. W. (1987). *Speech perception by ear and eye.* Hillsdale, NJ: Erlbaum.

May, R. S. (1987). *Realismus von subjektiven Wahrscheinlichkeiten* [Accuracy of subjective probabilities]. Frankfurt, Federal Republic of Germany: Lang.

McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology, 22,* 171–193.

Melton, A. W. (1962). Editorial. *Journal of Experimental Psychology, 64,* 553–557.

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

Mises, R., von. (1957). *Probability, statistics, and truth.* London: Allen & Unwin.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Neumann, J. von. (1958). *The computer and the brain.* New Haven, CT: Yale University Press.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London, Series A, 236,* 333–380.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. Part I. *Biometrika, 20A,* 175–240.

Nickles, T. (1980). Introductory essay: Scientific discovery and the future of philosophy of science. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 1–59). Dordrecht, The Netherlands: Reidel.

Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioural sciences.* Chichester, England: Wiley.

Piaget, J. (1930). *The child's conception of causality.* London: Kegan Paul, Trench, & Trubner.

Popper, K. (1959). *The logic of scientific discovery.* New York: Basic Books.

Porter, T. (1986). *The rise of statistical thinking, 1820–1900.* Princeton, NJ: Princeton University Press.

Quetelet, L. A. J. (1969). *A treatise on man and the development of his faculties* (R. Knox, Trans.). Gainesville, FL: Scholars' Facsimiles and Reprints. (Original work published 1842)

Reichenbach, H. (1938). *Experience and prediction.* Chicago: University of Chicago Press.

Rucci, A. J., & Tweney, R. D. (1980). Analysis of variance and the "second discipline" of scientific psychology: A historical account. *Psychological Bulletin, 87,* 166–184.

Searle, J. (1984). *Minds, brains and science.* Cambridge, MA: Harvard University Press.

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

Simon, H. A. (1973). Does scientific discovery have a logic? *Philosophy of Science, 40,* 471–480.

Simon, H. A., & Kulkarni, D. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science, 12,* 139–175.

Smith, L. D. (1986). *Behaviorism and logical positivism.* Stanford, CA: Stanford University Press.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association, 54,* 30–34.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1964). Decision processes in perception. In J. A. Swets (Ed.), *Signal detection and recognition in human observers* (pp. 3–57). New York: Wiley.

Swijtink, Z. G. (1987). The objectification of observation: Measurement and statistical methods in the nineteenth century. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution: Vol. 1. Ideas in history* (pp. 261–285). Cambridge, MA: MIT Press.

Tanner, W. P., Jr. (1965). *Statistical decision processes in detection and recognition* (Technical Report). Ann Arbor: University of Michigan, Sensory Intelligence Laboratory, Department of Psychology.

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

Teigen, K. H. (1983). Studies in subjective probability IV: Probabilities, confidence, and luck. *Scandinavian Journal of Psychology, 24*, 175–191.

Thorndike, R. L. (1954). The psychological value systems of psychologists. *American Psychologist, 9*, 787–789.

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

Titchener, E. B. (1896). *An outline of psychology.* New York: Macmillan.

Toulmin, S., & Leary, D. E. (1985). The cult of empiricism in psychology, and beyond. In S. Koch & D. E. Leary (Eds.), *A century of psychology as science* (pp. 594–617). New York: McGraw-Hill.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 49–72). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). Cambridge, England: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.

Tweney, R. D., Dotherty, M. E., & Mynatt, C. R. (Eds.). (1981). *On scientific thinking.* New York: Columbia University Press.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3*, 316–347.

Wise, M. N. (1988). Mediating machines. *Science in Context, 2*, 77–113.