

## Discovery in Cognitive Psychology: New Tools Inspire New Theories

Gerd Gigerenzer\*

### *The Argument*

Scientific tools—measurement and calculation instruments, techniques of inference—straddle the line between the context of discovery and the context of justification. In discovery, new scientific tools suggest new theoretical metaphors and concepts; and in justification, these tool-derived theoretical metaphors and concepts are more likely to be accepted by the scientific community if the tools are already entrenched in scientific practice.

Techniques of statistical inference and hypothesis testing entered American psychology first as tools in the 1940s and 1950s and then as cognitive theories in the 1960s and 1970s. Not only did psychologists resist statistical metaphors of mind prior to the institutionalization of inference techniques in their own practice; the cognitive theories they ultimately developed about “the mind as intuitive statistician” still bear the telltale marks of the practical laboratory context in which the tool was used.

### Introduction

#### *The Problem*

How do we arrive at new ideas? Gottfried Wilhelm Leibniz once had a vision of how to generate new knowledge mechanically. “Now since all human knowledge can be expressed by the letters of the alphabet, and since we may say that whoever understands the use of the alphabet knows everything, it follows that we can calculate the number of truths which men are able to express.” (1690/1951, p. 75) And calculate he did: The number was in the order of 1 followed by 73 trillion zeros. It included all truths, all falsehoods, and, for the most part, letter combinations that signify nothing at all. Swift parodied Leibniz’ “art of combinations” in *Gulliver’s Travels*. The academicians of Lagoda had written down all the words of their language and randomly combined them. If a sequence of words looked intelligible, the Lagodians wrote it down and stored it in their library of knowledge. Present-day computers can practice Leibniz’ art of combinations much faster than the Lagodians could. However, since Gulliver visited the academy of Lagoda long ago, the mechanical art of discovery has made little progress. Where do new ideas come from, if not from Lagoda?

#### *The Demise of the Problem*

Most working scientists—and in this regard social scientists are indistinguishable from natural scientists—have either been indifferent to the source of their new ideas, or wax mystical on the subject, relating dreams and thunderbolt inspirations. Many philosophers, especially neopositiv-

---

\* This article was supported by a grant, P8842-MED, from the Fonds zur Förderung der wissenschaftlichen Forschung (FWF), Austria.

ists, have generally followed the scientists' lead, drawing a sharp distinction between the context of discovery and the context of justification, and concentrating mainly on justification: "The philosopher of science is not much interested in the thought processes which lead to scientific discoveries ... that is, he is not interested in the context of discovery, but in the context of justification." (Reichenbach, 1949, p. 292)

Simply put, the message is: We neither know nor care where theories come from; we just want to know whether they are right or wrong. In the writings of many, discovery and justification became two distinct, temporally ordered and largely unconnected, entities—so distinct as to mark the lines between disciplines. For instance, the philosopher R. B. Braithwaite remarked on problems of discovery: "These are historical problems, both as to what causes the individual scientist to discover a new idea, and as to what causes the general acceptance of scientific ideas. The solution of these historical problems involves the individual psychology of thinking and the sociology of thought. None of these questions is our business here." (1953, pp. 20–21)

Karl Popper, in *The Logic of Scientific Discovery* (1935/1959), denied the very existence of the object named in his title (the original German version, however, is entitled *Logik der Forschung*, which means "logic of research"). Popper, like Braithwaite, handed down the study of discovery to psychology:

The question how it happens that a new idea occurs to a man—whether it is a musical theme, a dramatic conflict, or a scientific theory—may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge. ... Accordingly, I shall distinguish sharply between the process of conceiving a new idea, and the methods and results of examining it logically. ... My view may be expressed by saying that every discovery contains "an irrational element," or a "creative intuition," in Bergson's sense. (p. 31)

Discovery thus became associated with irrationality, or Popper's lucky guesses, and romanticized as fundamentally mysterious. Justification, in contrast, became associated with logic, mathematics, and statistics.

Contrary to Popper's suggestion, psychologists have not paid much attention to the processes of discovery. There are exceptions, such as Max Wertheimer (1959) on Einstein, Howard Gruber (1981) on Darwin, Ryan Tweney (1985) on Faraday, and Herbert Simon (1973) reworking a logic of discovery. But by and large, psychologists, like other social scientists, have been strongly concerned with the logic and the tools for justification—in particular, with Popper's falsification logic and, to a much greater extent, with the statistical methods of R. A. Fisher, Jerzy Neyman and Egon Pearson, and others, for drawing inferences from data to hypotheses.

### *A New Look at Discovery: The Theoretical Power of Tools*

The conduct of science is often discussed using the twin terms of theory and data. Hypothetico-deductivist methodology in general and fallibilism in particular hold that theory comes first, then data are obtained in an attempt to falsify the theory. In Popper's view, for instance, data hardly have lives of their own. Only through the lens of theory do we know what we observe (e.g., Popper, 1978, p. 46). On inductivist accounts, however, data play the primary role, and theories emerge as generalizations from data. Much of the debate surrounding the "theory-ladenness" of data has centered on the precise relationship of these twins. Until recently, however, few philosophers of science have looked beyond theory and data, to what actually happens in the laboratory. Even those philosophers who called themselves logical empiricists, ironically, did not include laboratory routines and tools in their analysis of scientific knowledge.

Scientific methods—from physical tools such as microscopes to analytical tools such as statistical techniques—are, however, not “neutral” with respect to either data or theory. First, data are tool-laden. For instance, quantitative judgments—ranging from estimates of the brightness of light to the number of hours per week one spends watching television—vary systematically with the particular measurement method used (e.g., Birnbaum, 1982; Schwarz et al., 1985). This poses a challenge to one fundamental assumption about the mind: the assumption that there exist, prior to measurement, definite subjective values for opinions, attitudes, or probabilities, which are to be “recorded” by some neutral method. The counterclaim is that such subjective values often do not exist but are created in the interaction between a measurement method and a mind (Gigerenzer, 1981). Accepting the challenge would lead us to analyzing the structure of a method (such as experimental tasks, instructions, and response scales) in terms of which information, probability cues, and demand characteristics act on the mind, rather than looking for a “neutral” method. This would revise the traditional questions of the type “What is in the mind?” into “How does this experimental setup (or environment) act on the mind?” However, much of psychology, like much of everyday thinking, still attempts to answer questions of the first type and looks for “neutral” tools of data generation that do not “distort” the subjective values that are assumed to have an independent prior existence (Gigerenzer, 1987a).

Second, and this is the argument I will develop here, theories are tool-laden, too. How do tools for justification shape theoretical concepts? Little seems to be known about this from the history of science, although there are a few suggestive examples. It has been argued that after the mechanical clock became an indispensable tool for astronomical measurement, the universe itself became seen as a kind of mechanical clock, and God as a divine watchmaker. Lenoir (1986) showed how Faraday’s instruments for recording electric currents shaped electrophysiological theory by promoting such concepts as “muscle current” and “nerve current.” Hackmann (1979) made a similar point about the role of instruments in eighteenth-century electrostatics. More generally, Gooding, Pinch, and Schaffer (1989) and Hacking (1983) have argued that experimental practice has a life of its own, and Galison (1987) and Lenoir (1986) have emphasized the role of the familiarity experimenters have with their tools, and the importance of experimental practice in the quest for knowledge. Nonetheless, despite the recent move toward emphasizing experimenters’ practice in addition to theory and data, not much is known about how that practice shapes new theoretical concepts.

What follows is a case study using theories of mind. I will (i) state the tools-to-theories heuristic, (ii) show how new statistical tools inspired new cognitive theories, and (iii) conclude with the argument that looking at the origins of theories is not only of historical interest but can help to evaluate limitations and possibilities in current cognitive theories.

### The Tools-to-Theories Heuristic

My general argument is that there is an understanding of discovery and acceptance of new ideas in terms of *heuristics* that goes beyond the accidental, mystical, and idiosyncratic. Heuristics are strategies that can be, but need not be, consciously applied by researchers. Heuristics are more general than thunderbolt guesses but less general than a monolithic logic of discovery, of the sort Hanson (1958) was looking for. The heuristics approach to discovery has been promoted by Herbert Simon and his co-workers (e.g., Langley et al., 1987; Simon & Kulkarni, 1988), and I will discuss below in what respects my approach differs from Simon’s data-driven concept of

heuristics. I will try to identify here only one of a potential bundle of heuristics that generate new ideas: the tools-to-theories heuristic (Gigerenzer, 1991a; Gigerenzer & Murray, 1987).

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

By "tools" I mean tools of justification—analytical or physical. Analytical tools may or may not rely on data. Examples of analytical methods of justification that use data are tools for data processing such as statistics; examples of analytical methods that do not use data are logical criteria for theory evaluation such as logical consistency. Examples of physical tools of justification include measurement instruments such as clocks. Thus the present thesis is more specific than the general assertion that scientists may use any new technology—such as the steam engine or the telephone switchboard—as metaphors or even as models.

In what follows I will deal with one specific analytical tool: techniques of statistical inference and hypothesis testing.

### The Institutionalization of the Tool: The "Inference Revolution"

The birthday of the experimental study of the mind is usually dated toward the end of the year 1879, when Wilhelm Wundt devoted some space at the University of Leipzig for conducting experiments. Wundt followed the German ideal for linking teaching and research in the form of an institute where students could perform experiments. Like G. T. Fechner's psychophysics, Wundt's psychology retained intimate links with philosophy and the *Geisteswissenschaften*. In the United States, in contrast, an intellectual structure comparable to German philosophy did not exist, and psychology developed disciplinary autonomy more quickly—given a university structure that favored specialization and was controlled by businessmen and politicians engaged in practical professional activity (Danziger, 1990). Thus proponents of the new project of a scientific psychology had to legitimate their project before a very different tribunal, depending on which side of the Atlantic they worked. In the United States, the primary legitimation was practical, not philosophical, and the primary market for marketable methods was education (Danziger, 1990; see also Rose, 1985, on how practical problems shaped scientific psychology in England).

To offer methods that would be able to solve the practical problems of educational administrators (e.g., to decide whether a new teaching method is more effective than an established one), the Wundtian model of experimentation had to be radically revised. In a Wundtian experiment, one or a few individuals were studied—usually Ph.D.s or professors—, each of them a trained observer of his own mental processes; the goal was knowledge, not practical application. During the 1920s and 1930s, a fundamental change in experimental practice occurred in the United States: from studying single individuals to aggregates; from studying professional psychologists to studying children, recruits, and undergraduates; and from studying natural groups such as boys versus girls (as in the Galtonian, not the Wundtian, tradition) to the study of treatment groups. Treatment groups are artificial, in the sense that their members happen to get the same treatment—such as a new method of instruction. Treatment groups represent neither a preexisting social or biological category nor an individual. Danziger (1987, 1990) has documented this shift in experimental practice. For instance, in 1914–1916, 70% of empirical studies in the *American Journal of Psychology* reported individual data only; by 1934–1936 this had dropped

to 31%, and by 1949–1951 to 17%. Group data only were reported by 80% of the studies in 1949–1951.

The triumph of the aggregate over Wundt's individual, and of the treatment group over Galton's natural groups, prepared the field for the institutionalization of statistical inference in the mid-1950s. Statistical methods for testing the significance of the difference between the means of two (or more) randomized treatment groups—such as Student's  $t$ -test and Fisher's analysis of variance—seemed to be tailor-made for the new subject matter of psychology: the aggregate.

Experimental studies before 1940 did not use any single standardized method to make inferences from data to hypotheses. On the contrary, they often employed a multitude of descriptive statistics—such as means, medians, modes, ranges, standard deviations, sums, ratios, and percentages. Inferences to hypotheses were sometimes made informally by eyeballing; sometimes, more formally, by critical ratios. Description and inference were not distinguished at all in other studies, or no inferences beyond the data were made (Gigerenzer & Murray, 1987, chap. 1). Experimentation before 1940 was not of one kind: Different ideas of experimental practice—ranging from Wundt to the treatment group and from the Gestalt psychologists' "demonstration experiment" to Egon Brunswik's "representative design"—coexisted.

The statistician R. A. Fisher (1935), however, emphasized that experimental design and statistical analysis are two sides of the same coin. To make Fisher's methods of significance testing fit psychological experiment, experimental design first had to show *repetition*, *randomization*, and *independence*—the mathematical ingredients of Fisher's significance testing. Wundtian-type experiments, for instance, did not do this, but the randomized treatment group experiment did. The Fisherian link between experiment and statistics effectively ruled out other ideas of experimental practice from experimental psychology, once the "inference revolution" had occurred.

Rucci and Tweney (1980) found only seventeen articles using Fisher's analysis of variance (ANOVA) before 1940, mainly from educational psychology and parapsychology. By 1955, more than 80% of articles in four leading journals used ANOVA and related methods of significance testing for evaluating hypotheses (Sterling, 1959). And today, the number is close to 100%. I will take 1955 to be an approximate date for the firm institutionalization of inferential statistics in curricula, textbooks, and editorials—in short, the *inference revolution* (Gigerenzer & Murray, 1987). By the early 1950s, half of the psychology departments of leading American universities had made statistical inference a graduate program requirement (Rucci & Tweney, 1980). Editors of major journals made statistical inference a requirement for publication. Editors and reviewers alike looked for objective criteria to evaluate the quality of research submitted, independent of its content. For instance, after editing the *Journal of Experimental Psychology* for twelve years, A. W. Melton said in an editorial that he used the level of significance (.05, .01, or .001) as a yardstick for the quality of the studies submitted (1962).

A flood of statistical textbooks for psychologists appeared, typically written by other psychologists—that is, nonmathematicians. Early textbook writers, such as J. P. Guilford in his influential *Fundamental Statistics in Psychology and Education* (1942), taught Fisher's significance testing, but in a version deeply confused with a "Bayesian" interpretation of levels of significance. That is, probabilities of data given some hypothesis (levels of significance) were confused with Bayesian probabilities of hypotheses given some data; significance testing cannot provide the latter, but probabilities of hypotheses are what researchers want, after all. After World War II, textbook writers became aware also of Jerzy Neyman and Egon S. Pearson's hypothesis-testing theory and started to add concepts from the Neyman-Pearson theory, such as "power," to their texts. The resulting statistical tool was presented as a single, monolithic body of knowledge, simply entitled "statistics." Even today, experimenters and students are not informed about the raging contro-

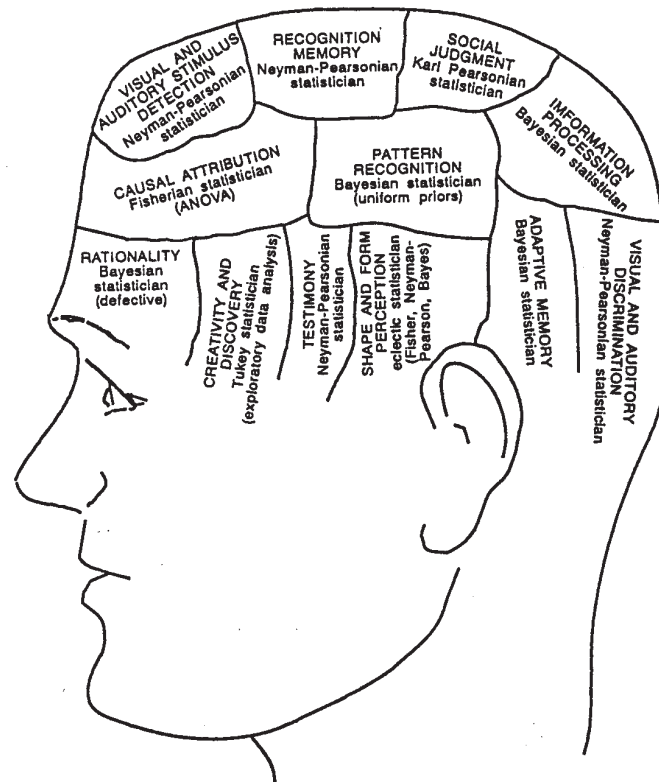
sies between Fisher on the one hand and Neyman and Pearson on the other, and between these frequentists and the Bayesians, or about the controversial issues surrounding statistical inference. Nor is it pointed out that physicists, chemists, and molecular biologists get along well without all these techniques of inference. Instead, the institutionalized tool has been presented as the *sine qua non* of scientific inference.

The German situation provides an instructive comparison. As mentioned above, German experimental psychology could legitimate itself independently of marketing methods. Moreover, there existed a well-established educational system, which provided only a limited market for the products of applied psychology (Danziger, 1990). These reasons, among others, seem to be responsible for the fact that in Germany the various experimental practices developed since Wundt were not as easily dominated by the treatment group method. For instance, in the period 1924–1925, the proportion of studies reporting group data only was as low as 0% in *Psychologische Forschung* and 13% in *Archiv für die gesamte Psychologie*, compared to 35% in the *American Journal of Psychology* and 44% in the *Journal of Experimental Psychology*. Ten years later that proportion was still low in the two German journals: 13% and 23%, respectively. In the *American Journal of Psychology* and the *Journal of Experimental Psychology*, however, the reporting of group data only had already reached 55% and 61%, respectively (Danziger, 1990). In Germany psychology, consequently, inferential statistics was institutionalized only when experimental methods imported from the United States reshaped it after World War II.

### From Institutionalized Statistics to Intuitive Statistics: The “Cognitive Revolution”

Shortly after the inference revolution, American behaviorism was eroded and ultimately overthrown by mentalist concepts. This is commonly referred to as the “cognitive revolution” of the 1960s (Gardner, 1985). The cognitive revolution was, however, not a “re-volution” in the original sense of the term; it was not a return to the earlier mentalist concepts. The cognitive revolution did more than revive the mental; it changed the very meaning of what the mental is. Tools such as computers and statistics turned into theories of mind.

Techniques of statistical inference provided a large part of the new mentalist concepts that fueled the cognitive revolution. Cognitive processes became understood as statistical calculations, and the mind became seen as a statistician that draws random samples from nervous fibers, computes probabilities, calculates analyses of variances, and sets decision criteria dependent on cost-benefit analyses. Figure 1 summarizes the most influential theories. For each theory, I have labeled the cognitive process it deals with and the kind of homunculus statistician in charge of that mental activity. The cognitive processes range from elementary to complex, and from conscious to unconscious. For the most part these topics were not new; they were studied intensively before the inference revolution (e.g., in Fechnerian psychophysics and in the Gestalt tradition). Using two topics from the forehead in Figure 1, I will illustrate how old questions were reshaped and found new answers through the analogy between tool and mind.



*Examples:* Visual and auditory stimulus detection: J. A. Swets (1964); recognition memory: B. B. Murdock (1982); social judgment: B. Brehmer and C. R. B. Joyce (1988); information processing: W. Edwards (1966); causal attribution: H. H. Kelley (1967); pattern recognition: D. W. Massaro (1987); rationality (defective): A. Tversky and D. Kahneman (1974); creativity and discovery: P. Langley, H. A. Simon, G. L. Bradshaw, and J. M. Zytkow (1987); testimony: M. H. Birnbaum (1983); shape and form perception: R. L. Gregory (1974); visual and auditory discrimination: W. P. Tanner (1965); adaptive memory: J. R. Anderson and R. Milson (1989).

*Figure 1.* Cognition as intuitive statistics.

### *Causal Attribution as Intuitive Statistics*

What makes us perceive a causal link between two events? What different kinds of causal relations does the human mind distinguish? Jean Piaget (1930), Albert Michotte (1946/1963), Karl Duncker (1935/1945), and others studied these and related questions in Europe. Michotte, for instance, investigated how certain temporal and spatial relationships between visual objects, such as moving dots, produced phenomenal causality. Subjects were made to “perceive” that one dot pushes or launches another. After the inference revolution, the American psychologist Harold H. Kelley and his colleagues made causal attribution a main topic in the cognitive social psychology of the 1970s and 1980s. His new idea was this: In his “causal attribution theory,” Kelley (1967, 1973) postulated that the mind attributes a cause to an effect in the same way

as American behavioral scientists have come to do—namely, by calculating (unconsciously) an analysis of variance, computing  $F$ -ratios, and testing null hypotheses.

The assumption is that the man in the street, the naive psychologist, uses a naive version of the method used in science. Undoubtedly, his naive version is a poor replica of the scientific one—incomplete, subject to bias, ready to proceed on incomplete evidence, and so on. Nevertheless, it has certain properties in common with the analysis of variance as we behavioral scientists use it. (Kelley, 1973, p. 109)

Kelley assumed that the mind calculates an analysis of variance (ANOVA) with three independent variables, which he called person, entity, and circumstances (time and modality). These were the potential causes for an observed behavior. For instance, in a study on causal attribution by McArthur, subjects were given the following information: “Paul is enthralled by a painting he sees at the art museum. Hardly anyone who sees the painting is enthralled by it. Paul is also enthralled by almost every other painting. In the past, Paul has almost always been enthralled by the same painting.” (1972, p. 110)

Subjects were asked what caused the effect (being enthralled by the painting): Paul (person), the painting (entity), the particular circumstances (time), or some interaction of these factors? The information supplied to the mind’s hypothesized ANOVA system specified that there is small variance (in being enthralled) across paintings, high variance across persons, and small variance over time. Therefore, the  $F$ -value (in ANOVA,  $F$  after Fisher) is low for entity and circumstance but high for person, and the mind should attribute the cause to the person—that is, to Paul.

The analogy between the statistical tool and causal inference was more than a new fashionable language, for it radically changed the kind of questions posed and the kind of research undertaken. Here is a shortlist of discontinuities in cognitive theory that bear the telltale fingerprints of the new tool:

(1) *Kinds of causal inference.* Michotte’s work reflected the broadly Aristotelian conception of four kinds of causes (see Gavin, 1972). Piaget (1930) even distinguished seventeen kinds of causes in children’s minds. Through the analogy with ANOVA, however, the new Fisherian mind focused only on the one kind of cause for which ANOVA is used as a tool. This kind of causality is similar to Aristotle’s material cause.

(2) *Nature of causal inference.* In Michotte’s view, and also in the view of the Gestalt psychologists, causal perception is direct and spontaneous and needs no inference—a consequence of laws inherent in the perceptual field. ANOVA, in contrast, is used in psychology as a technique for inductive inferences from data to hypotheses; the focus in Kelley’s attribution theory is, consequently, on the inductive, data-driven side of causal perception.

(3) *Material to study causal inference.* In Michotte’s and Heider’s work, the experimental setup for studying causal perception consisted, typically, of moving stimuli such as dots. ANOVA, in contrast, needs repetitions or numbers as data in order to estimate variances and covariances. Consequently, the material presented to subjects in order to study the new Fisherian mind consisted of stories with information about the frequency of events (e.g., McArthur, 1972)—material that had played no role in either Michotte’s or Piaget’s work.

These three fingerprints of the tool may suffice here to illustrate the sharp discontinuity in theory before and after the inference revolution. I turn now to some details important for the present thesis on discovery and acceptance.

Kelley repeatedly credited Heider for having inspired his ANOVA theory of causal attribution. Heider indeed suggested that causal reasoning might be analogous to experimental methods, but he metaphorically talked about an implicit “factor analysis” (Heider, 1958, p. 123 and 297). Now factor analysis was not a method of the experimental community; rather it was an indispensable method of its rival community, the “correlational” discipline that investigates indi-

vidual differences in intelligence and personality. The schism between the two communities, the “Tight Little Island” of experimental psychology and the “Holy Roman Empire” of correlational psychology, as Cronbach (1957) called them, had been repeatedly taken up in presidential addresses before the American Psychological Association and had deeply affected the values and the mutual esteem of psychologists (Gigerenzer, 1987b). There is no a priori reason why factor analysis or analysis of variance should be a better model of intuitive causal reasoning—and in fact both tools have been used (largely uncritically) in their respective communities to justify causal claims. But Kelley did not follow Heider’s suggestion here. He chose the tool he and his colleagues of the American experimental community were familiar with: analysis of variance.

By the time Kelley proposed the analogy between ANOVA and causal reasoning, about 70% of the articles in major American journals using statistical tests already used ANOVA to process data and justify conclusions to hypotheses (Edgington, 1974). Thus when Kelley proposed his new vision of the mind, his colleagues in the experimental community were familiar with the tool from their laboratory practice and were prepared to accept the analogy as plausible. The long-sought laws of causal reasoning were the tools of the behavioral scientists. What else could they be?

*Practical context.* I want to emphasize that the heuristic of discovery and acceptance I am dealing with here is *not* simply the application of a new kind of mathematics to an old topic. My focus is on the practical context of justification a scientist is familiar with. This context is much richer than the pure mathematics or statistics used. For instance, nothing in the mathematics of significance tests (as used in ANOVA) tells the scientists whether they should *use* the test to reject hypotheses or reject data (so-called “outliers”). Both are tasks in experimental science: get rid of bad hypotheses and of bad data.

In the experimental community within psychology, however, significance tests have been used only for rejecting hypotheses, not as a tool for rejecting data. In the laboratory the problem of outliers, or potentially bad data, has been handled informally, by judgment. In contrast, significance tests in other disciplines or periods were applied to reject data, not hypotheses. For instance, in the early nineteenth century it was common practice for astronomers to use significance tests (similar to those in ANOVA) for rejecting outliers (Swijtink, 1987). At least provisionally, the astronomers assumed their theory to be correct and mistrusted the data, whereas the ANOVA mind, following the laboratory practice in experimental psychology, assumes the data to be correct and mistrusts the theory. Assume that Harold Kelley had lived one and a half centuries earlier than he did and was trained in the use of significance tests to reject data. According to the present thesis, it is likely that our nineteenth-century Kelley would have seen causal attribution as hypothesis-driven, rather than data-driven. The homunculus statistician in the brain would have checked the data—that is, the information presented—and not the hypothesis.

The general point is that the new tool is not just an application of new mathematics; it also bears the marks of the practical laboratory context in which it is used. In the case of Kelley’s new theory, that context is an inductive, data-driven view of justification.

### *Stimulus Detection as Intuitive Statistics*

Detection of a stimulus and discrimination between two stimuli are seen as the elementary building blocks of cognition. How intense must a signal on a radar screen be to be detected against a background of white noise? How much heavier than a standard stimulus of 100 grams must a comparison stimulus be for a difference to be perceived? Since Herbart (1816), detection and dis-

crimination have been explained using a threshold metaphor. Detection occurs only if the effect a stimulus has on the nervous system exceeds a certain threshold value, the “absolute threshold.” Detecting a difference (discrimination) between two stimuli occurs if the excitation from one exceeds that of the other by an amount greater than a “differential threshold.” A long psychophysical tradition, from E. H. Weber to G. T. Fechner to E. B. Titchener, saw in the differential thresholds the elements of mind (Titchener counted about 44,000).

After the inference revolution, the psychophysics of thresholds was revolutionized by the new metaphor of the mind as a statistician. W. P. Tanner and J. A. Swets proposed in their theory of signal detectability (TSD) 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 (Swets, 1964; Tanner & Swets, 1954). In Neyman-Pearson statistics, two sampling distributions (hypotheses  $H_0$  and  $H_1$ ) and a decision criterion are defined. The latter balances the probabilities of the two possible decision errors, Type I and Type II errors. Depending on which side of the criterion the data fall, the decision “reject  $H_0$  and accept  $H_1$ ,” or vice versa, is made. In straight analogy, TSD assumes that the mind calculates two sampling distributions for “noise” and “signal plus noise” and sets a decision criterion. The latter balances the costs of the two possible perceptual errors, now called “false alarms” and “misses.” Depending on which side of the criterion an observation falls (both the criterion and the observation are represented as likelihood ratios), the subject says “No, there is no signal” or “Yes, there is a signal.” Tanner and his colleagues were, like Kelley, explicit about the analogy between the human mind and a statistician. The subject’s “task is, in fact, the task of testing a statistical hypothesis” (Tanner & Swets, 1954, p. 403). Tanner (1965) referred to the mind as a “Neyman-Pearson” detector and, in unpublished work, included in flow charts the drawing of a homunculus statistician performing the unconscious statistics of the mind (see Gigerenzer & Murray, 1987, pp. 49–53).

Again, the new analogy radically changed the understanding of the nature of stimulus detection and created new research questions and even a new kind of data. Here are some fingerprints of the statistical tool:

(1) *Nature of stimulus detection.* The century-old concept of a fixed threshold was replaced by the twin notions of observer’s attitudes and observer’s sensitivity. Just as in hypotheses testing the Neyman-Pearson technique emphasizes both a subjective part (e.g., selection of a decision criterion depending on cost-benefit considerations) and a mathematical part, stimulus detection became understood as involving both subjective processes—such as attitudes and cost-benefit considerations—and sensory processes (Swets et al., 1964).

(2) *New questions.* The analogy between stimulus detection and intuitive statistics made new questions thinkable. For instance, “How can the mind’s decision criterion be manipulated?” A substantial number of experiments on visual and acoustic detection and discrimination were performed to answer this question (see Green & Swets, 1966; Swets, 1964).

(3) *New data.* The analogy made new kinds of data visible. Two types of errors—false alarms and misses—were generated and investigated in the new experiments, just as the statistical theory distinguishes between two types of errors. As far as I can tell, the practice of generating these two kinds of data was not common before.

For the present argument it is important that the new analogy between stimulus detection and intuitive statistics was not suggested or even forced by new data. Quite the contrary. In their seminal paper, Tanner and Swets (1954, p. 401) explicitly admit that their new theory “appears to be inconsistent with the large quantity of existing data on this subject.” What happened was that a new tool inspired a new theory, which changed the kind of data generated.

Following Tanner's lead, the Neyman-Pearson technique of hypothesis testing was subsequently transformed into theories for a broad range of cognitive processes (see Figure 1), ranging from recognition memory (e.g., Murdock 1982; Wickelgreen & Norman, 1966) to eyewitness testimony (e.g., Birnbaum, 1983) and perception of randomness (e.g., Lopes, 1982).

### *The Mind as an Intuitive Statistician*

I will briefly discuss some of the remaining homunculi statisticians appearing in Figure 1. How does the mind perceive shape and form? Why is there no blank in the viewed world that corresponds to the blind spot in the eye? How is this blank filled in? To answer such questions, Gregory (1974, 1980) developed a hypothesis-testing version of Helmholtz's "unconscious inferences." The mind is a "betting machine" or "induction machine," and the best model of its cognitive strategies is the scientific method: "Perception is similar to science itself." (Gregory, 1980, p. 63) To perceive an object means to test and accept a hypothesis, and Gregory takes the notion of perceptions as hypotheses quite literally. For instance, significance levels explain why we perceive an object as having only one shape or form, despite the presence of competing hypotheses: "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 hypotheses." (Gregory, 1974, p. 528) Gregory's "betting machine" is an eclectic statistician who does Fisherian significance testing, Neyman-Pearson hypothesis testing, and Bayesian statistics without clearly distinguishing between them—a procedure similar to the eclectic teaching of statistics and the corresponding practice in psychological laboratories mentioned above.

Ward Edwards and his co-workers were among the first to test whether the homunculus statistician is a Bayesian statistician. In Edwards' work, the dual function of statistics as a tool and a model of mind is again evident. On the one hand, Edwards, Lindman, and Savage (1963) proposed Bayesian statistics for scientific hypothesis evaluation. On the other hand, in his research on information processing, Edwards confronted untutored people with probability revision problems and studied how good a Bayesian statistician the mind actually is. He found the mind to be a reasonably good, albeit conservative Bayesian (e.g., Edwards, 1966). His program was taken up by Daniel Kahneman, Amos Tversky, and others, who claimed the opposite result: The mind is not a Bayesian at all. In the 1970s and 1980s, Edwards' Bayesian statistician in the mind came to be seen as defective—a claim that stimulated controversies about human rationality (e.g., Cohen, 1982; Tversky & Kahneman, 1974). However, the "defective" Bayesian statistician is rather the exception. Most recent theories that interpret pattern recognition, speech perception, adaptive memory, and other cognitive processes as Bayesian statistics (e.g., Anderson, 1991) consider this analogy to be both normatively and descriptively appropriate. For instance, Massaro (1987) and Anderson and Milson (1989) propose Bayes' rule as both the optimal and the actual mechanism for speech perception and human memory, respectively.

The examples of new ideas and new cognitive theories I have given count among the most influential and innovative ideas in cognitive psychology since the cognitive revolution. Smith (1986) discussed related cases with physical tools. For instance, he 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."

The impact of tools on new ideas may be particularly great in such disciplines as psychology, where experimenter and theoretician are but two roles of the same person. Unlike physics, psychology has not yet established a division of labor between a theoretical psychology and an experimental psychology.

### Discovery Without Acceptance

One important test for the present thesis is to look at the period before the institutionalization of inferential statistics in experimental psychology. Theories that conceive of the mind as an intuitive statistician (of whatever school) should have a very small likelihood of being discovered during this earlier period, and an even smaller likelihood of being accepted. I know of only a single case, that of Egon Brunswik, in which an analogy between mind and statistician was proposed during this period. How was his analogy received?

Brunswik's psychology was based on a quite unusual blending of the European functionalist tradition, represented by his doctoral adviser Karl Bühler at Vienna; the logical positivist philosophy of the Vienna Circle, represented by his second doctoral adviser, Moritz Schlick; and American neo-behaviorism, as set forth by Edward Tolman (Leary, 1987). The personal and professional ties with Tolman were the main reason why Brunswik left Vienna for Berkeley in 1937. Once in the United States, he became influenced by a fourth tradition, the Anglo-American statistical tradition of correlation and regression analyses, as founded by Karl Pearson and Francis Galton. Under the influence of the latter tradition, Brunswik changed his methods for measuring perceptual constancies, from calculating (nonstatistical) "Brunswik ratios" (see Brunswik, 1934) to calculating Pearson correlations (Brunswik, 1940), such as "functional" and "ecological validities." After adopting the new tools in the late 1930s, he began to think of the perceptual system as an "intuitive statistician." Like the Brunswikian researcher, the perceptual system was supposed to calculate correlations and regressions to infer the structure of the environment from ambiguous or incomplete perceptual cues. Brunswik seems to have been the first to propose the metaphor of the mind as intuitive statistician.

Brunswik's analogy came too early, around 1940—which is about fifteen years before the institutionalization of inferential statistics in American experimental psychology. Moreover, it rested on the "wrong" techniques: correlational statistics. Brunswik's intuitive statistician, like the Brunswikian researcher, was of the Karl Pearson school. As mentioned earlier, correlation was an indispensable method not in the experimental community but rather in its rival community of personality researchers, known as "correlational psychology" (Cronbach, 1957). Brunswik could not succeed in persuading his colleagues in the experimental community to consider the statistical tools of the competing community as a model of how the mind works. Ernest Hilgard (1955), in his rejection of Brunswik's ideas, did not mince words: "Correlation is an instrument of the devil." (p. 228)

Not only was Brunswik's analogy not accepted by the experimental community at the time; most of his colleagues did not even understand what Brunswik was saying. The misunderstanding and the resistance are well-documented in the two discussions of Brunswik's new ideas—at Chicago in 1941, and at Berkeley in 1953 (see Gigerenzer, 1987b). As Leary (1987, pp. 133–134) points out, even Brunswik's own students and colleagues at Berkeley admitted a considerable degree of incomprehension.

Two months after Hilgard's statement appeared in a series of comments on Brunswik's ideas that epitomize rejection and lack of understanding, Brunswik committed suicide. Brunswik 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 with (Karl) Pearsonian statistics as models of mind.

Brunswik's case illustrates the difference between discovery and acceptance. Tools familiar from one's own research practice can suggest new theoretical concepts and metaphors. Familiarity with the tool in the scientific community to which the scientist belongs can prepare or hinder the acceptance of the new ideas.

### The Double Legacy of the Tool: Mathematics and Practical Context

Where do theories come from, and does it matter? I have argued that the context of justification is a resource for discovery. But does knowledge about the origins of a theory also matter for its evaluation? I will argue that looking at the origins of ideas is not only of interest for understanding discovery and for acceptance of new theories, but that it can also help in evaluating theories.

Tools do not come as pure mathematical (or physical) systems; they come laden with the baggage of practical contexts of application. Features of the context in which a tool has been used may be smuggled Trojan-horse fashion into new theories. I have mentioned one such "birthmark" above. Because methods of statistical inference had been used in experimental psychology exclusively for rejecting hypotheses while trusting the data (and not vice versa), causal attribution was seen to be data-driven—that is, to be the product of a cognitive mechanism that tests hypotheses (potential causes) but not data (information).

In this section I will briefly point to a second feature of the practical context, which has reappeared in recent cognitive research programs (for further details, see Gigerenzer, 1991a). Statistical inference is taught and used in experimental psychology as if there existed a single, monolithic method that gives the correct, or at least the best, solution to all problems of scientific inference (Gigerenzer, 1992). Kelley (1973, p. 109), for instance, even speaks of analysis of variance as "the method used in science." The assumption that "statistics speaks with one voice" explains why, after the inference revolution, almost all experimental psychologists started to do one kind of statistical calculation for all kinds of inductive inference. This practice is in sharp contrast to the actual situation in statistics, where several competing proposals for good statistical inference exist. The problem of scientific inference has not yet found a unique solution. As Neyman and Pearson (1928, p. 176) emphasized, in "many cases there is probably no single best method of solution."

The practice-governing assumption that "statistics speaks with one voice" has reemerged in, among others, Tversky and Kahneman's "heuristics and biases" program—a program studying intuitive statistical reasoning. Everyday problems of inference are posed, and subjects' answers are recorded. The experimenters claim that each of these problems has only one "correct" answer; if subjects deviate from this answer, their responses are labeled "fallacies" of reasoning. In this rhetoric the allegedly correct solution is directly derived from "the dictates of normative statistical theory" (Bar-Hillel, 1984, p. 99) or from "the normative theory of prediction" (Kahneman & Tversky, 1973, p. 234). But there is no single normative theory of prediction; consequently, the reasoning problems do not have only one "correct" answer. For instance, Neyman-Pearson theory leads to different answers for the famous "cab problem" than Tversky and Kahneman (1980) cal-

culated using Bayes' theorem (see Birnbaum, 1983; Gigerenzer & Murray, 1987, pp. 167–174). In the “cab problem,” the subjects' task is to estimate the probability that a cab involved in a hit-and-run accident at night was from the Blue (as opposed to the Green) company, given a witness' testimony “blue,” and information about the perceptual accuracy of the witness and about the base rates of blue and green cabs in the city. The fact that this and other reasoning problems have several good answers rather than a single “normative” one has important consequences for claims that people's answers reflect “fallacies” and “biases” of probabilistic reasoning, and for a revised research program on intuitive statistics (see Gigerenzer, 1991b; Lopes, 1991).

### Heuristics of Discovery: Data-Driven Versus Tool-Driven

In their computational explorations of scientific creativity, Herbert Simon and his co-workers (e.g., Langley et al., 1987) attempted to develop general-purpose discovery systems. These do not rely on domain-dependent heuristics, as did earlier systems such as DENDRAL—a program that identifies organic molecules from mass spectrograms and nuclear magnetic resonances. The creators of BACON, in contrast, claim that it is “a system of considerable generality that is capable of significant scientific discovery” (Langley et al., 1987, p. 63). Langley et al. (1987) deal with two kinds of discoveries: quantitative laws using BACON and related programs, and qualitative taxonomies using clustering methods. In both cases, the underlying metaphor for scientific discovery is *intuitive statistics*. Here, the homunculus statistician uses exploratory data analysis (Tukey, 1977), as indicated in Figure 1.

How does BACON make a discovery? What is the relation between BACON's heuristics of discovery and the tools-to-theories heuristic?

Consider how BACON rediscovers Kepler's third law of planetary motion. As the program's name indicates, discovery starts from data: two rows of measurements—for a planet's distance and its period, respectively. The heuristics of discovery are rules for transforming that data, such as “If the values of two numerical terms increase together, then consider their ratio.” The successive transformations of the original data by such heuristics finally produce Kepler's law (noisy data poses some problems).

There are basic differences between the concept of a heuristic in BACON and the tools-to-theories heuristic. First, BACON's heuristics work on data, whereas the tools-to-theories heuristic works on tools for data generation or processing. The latter does not depend on empirical data at all; as I have argued above, it can even produce new kinds of data. Second, BACON's heuristics discover descriptions of data—that is, numerical laws—but no explanations of phenomena. The tools-to-theories heuristic does not discover any numerical relationships, since it does not operate on data. It can, however, generate rich conceptual frameworks that not only have explanatory power but in fact revolutionize our understanding of an old phenomenon and create new phenomena. A good example is the theory of signal detectability (Gigerenzer & Murray, 1987, chap. 2). Thus there is little overlap between the two concepts of heuristics; their function is complementary.

More recently, Simon and Kulkarni (1988) went beyond the data-driven view of discovery in an attempt to include heuristics for planning and guiding experimental research. Still, the present approach differs from Simon's in the emphasis on experimental practice as a source of new ideas. Both approaches, however, emphasize that it is possible to understand discovery in terms of heuristics of creative thinking and thus to go beyond lucky guesses.

## Conclusion

The present case study of tool-laden theories gives a new twist to the relation between the context of discovery and the context of justification. In the debate on whether discovery is an important topic for understanding science, both sides in the debate have construed the issue to be whether an *earlier* stage of discovery should be added to a *later* stage of justification (Nickles, 1980). In contrast, I have described a situation where methods of justification come first and discovery follows.

Scientific discovery is still, by and large, *terra incognita*. As Gerald Holton in his seminal *Thematic Origins of Scientific Thought* (1988) puts it, “There has been no systematic development of the point” (p. 41). Holton analyzed “thematic” preconceptions in theories, whereas I have examined methodological preconceptions in the origins of new theories. One or the other seem to be indispensable for theorizing.

Imagine that finally a member of the academy of Lagoda found a nineteen-word sentence—which happened to be the same one Charles Darwin used in 1838 to express the three principles of heredity, variation, and superfecundity, from which natural selection and evolution followed inexorably (Gruber, 1977). If the Lagodans did not share Darwin’s framework of metaphors—based on, among other things, his practical experience with artificial selection as a pigeon breeder—they might not even understand the meaning of the sentence. Sharing familiar experiences with research practice and tools may be indispensable for understanding and accepting new theories.

Popper (1935/1959, p. 278) once declared that “scientific guesses are guided by the unscientific.” However, the tools-to-theories heuristic describes a process of discovery which emphasizes that science, in the form of tools of justification, does indeed guide guesswork. At least in some cases, the context of justification turns out to explain the context of discovery.

## Acknowledgements

I would like to thank Mitchell Ash, Kurt Danziger, Lorraine Daston, and two anonymous reviewers for many helpful comments.

## References

- Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517.
- Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 703–719.
- Bar-Hillel, M. (1984). Representativeness and fallacies of probability judgment. *Acta Psychologica*, *55*, 91–107.
- Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysics*. Hillsdale, NJ: Erlbaum.
- Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, *96*, 85–94.
- Braithwaite, R. B. (1953). *Scientific explanation*. Cambridge, UK: Cambridge University Press.
- Brehmer, B., & Joyce, C. R. B. (Eds.). (1988). *Human judgments: The SyT view*. Amsterdam: North-Holland.
- Brunswik, E. (1934). *Wahrnehmung und Gegenstandswelt*. Leipzig: Deuticke.
- Brunswik, E. (1940). Thing constancy as measured by correlation coefficients. *Psychological Review*, *47*, 69–78.
- Cohen, L. J. (1982). Are people programmed to commit fallacies? Further thoughts about the interpretation of experimental data on probability judgment. *Journal of the Theory of Social Behavior*, *12*, 251–274.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671–684.

- 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, UK: Cambridge University Press.
- Duncker, K. (1935/1945). On problem solving. *Psychological Monographs*, 58 (5, Whole No. 270).
- Edgington, E. E. (1974). A new tabulation of statistical procedures used in APA journals. *American Psychologist*, 29, 25–26.
- Edwards, W. (1966). *Nonconservative information processing systems*. University of Michigan, Institute of Science and Technology (Report 5893-22-F).
- Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193–242.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh, UK: Oliver and Boyd.
- Galison, P. (1987). *How experiments end*. Chicago: Chicago University Press.
- Gardner, H. (1985). *The mind's new science*. New York: Basic Books.
- 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. (1981). *Messung und Modellbildung in der Psychologie*. Munich: Reinhardt (UTB).
- 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. (1991a). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254–267.
- Gigerenzer, G. (1991b). How to make cognitive illusions disappear: Beyond heuristics and biases. *European Review of Social Psychology*, 2, 83–115.
- Gigerenzer, G. (1992). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *Methodological and quantitative issues in the analysis of psychological data* (pp. 311–339). Hillsdale, NJ: Erlbaum.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gooding, D., Pinch, T., & Schaffer, S. (1989). Introduction: Some uses of experiment. In D. Gooding, T. Pinch, & S. Schaffer (Eds.), *The uses of experiment: Studies in the natural sciences* (pp. 1–27). Cambridge, UK: Cambridge University Press.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Gregory, R. L. (1974). *Concepts and mechanism of perception*. New York: Scribner.
- Gregory, R. L. (1980). The confounded eye. In R. L. Gregory & E. H. Gombrich (Eds.), *Illusion in nature and art* (pp. 49–96). New York: Scribner.
- Gruber, H. E. (1977). The fortunes of a basic Darwinian idea: Chance. In R. W. Rieber & K. Salzinger (Eds.), *The roots of American psychology: Historical influences and implications for the future* (pp. 233–245). New York: New York Academy of Sciences.
- Gruber, H. E. (1981). *Darwin on man: A psychological study of scientific creativity* (2nd ed.). Chicago: University of Chicago Press.
- Guilford, J. P. (1942). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.
- Hacking, I. (1983). *Representing and intervening*. Cambridge, UK: 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.
- Hanson, N. R. (1958). *Patterns of discovery*. Cambridge, UK: Cambridge University Press.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Herbart, J. F. (1816/1891). *A textbook in psychology* (M. K. Smith, Trans.). New York: Appleton. (Originally published as *Lehrbuch zur Psychologie*, Hamburg: G. Hartenstein)
- 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.
- 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, pp. 192–238). Lincoln: University of Nebraska Press.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 28, 107–128.
- Krüger, L., Gigerenzer, G., & Morgan, M. S. (Eds.). (1987). *The probabilistic revolution: Vol. 2. Ideas in the sciences*. Cambridge, MA: MIT 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, UK: Cambridge University Press.
- von Leibniz, G. W. (1690/1951). The horizon of human doctrine. In P. P. Wiener (Ed.), *Selections* (pp. 73–77). New York: Scribner.
- Lenoir, T. (1986). Models and instruments in the development of electrophysiology, 1845–1912. *Historical Studies in the Physical Sciences*, 17, 1–54.
- 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.
- McArthur, L. A. (1972). The how and what of why: Some determinants and consequents of causal attribution. *Journal of Personality and Social Psychology*, 22, 171–193.
- Massaro, D. W. (1987). *Speech perception by ear and eye*. Hillsdale, NJ: Erlbaum.
- Melton, A. W. (1962). Editorial. *Journal of Experimental Psychology*, 64, 553–557.
- Michotte, A. (1946/1963). *The perception of causality*. London: Methuen.
- Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626.
- Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part 1. *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: Reidel.
- Piaget, J. (1930). *The child's conception of causality*. London: Kegan Paul.
- Popper, K. (1935/1959). *The logic of scientific discovery*. New York: Basic Books.
- Popper, K. (1978). *Conjectures and refutations*. London: Routledge & Kegan Paul.
- Reichenbach, H. (1949). The philosophical significance of the theory of relativity. In P. A. Schilpp (Ed.), *Albert Einstein: Philosopher-scientist* (pp. 289–311). Evanston, IL: Library of Living Philosophers.
- Rose, N. (1985). *The psychological complex: Psychology, politics and society in England, 1869–1939*. London: Routledge & Kegan Paul.
- 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.
- Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and subsequent judgments. *Public Opinion Quarterly*, 49, 1460–1469.
- 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, 139–176.
- 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. (Ed.). (1964). *Signal detection and recognition by human observers*. New York: Wiley.
- Swets, J. A., Tanner, W. D., & 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. (1987). The objectification of observation. In L. Krüger, L. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution: Vol. 1. Ideas in history*. Cambridge, MA: MIT Press.
- Tanner, W. P., Jr. (1965). *Statistical decision processes in detection and recognition* (Technical Report). 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.
- 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.
- Tweney, R. D. (1985). Faraday's discovery of induction: A cognitive approach. In D. Gooding & F. James (Eds.), *Faraday rediscovered: Essays on the life and work of Michael Faraday, 1791–1867*. London: Macmillan.
- Wertheimer, M. (1959). *Productive thinking*. New York: Harper.
- Wickelgreen, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 3, 316–347.