Psychological Bulletin
2004, Vol. 130, No. 6, 959–988

0033-2909/04/$12.00 DOI: 10.1037/0033-2909.130.6.959

The Role of Representative Design in an Ecological Approach to Cognition

Mandeep K. Dhami
City University

Ralph Hertwig
University of Basel

Ulrich Hoffrage
Max Planck Institute for Human Development

Egon Brunswik argued that psychological processes are adapted to environmental properties. He proposed the method of representative design to capture these processes and advocated that psychology be a science of organism–environment relations. Representative design involves randomly sampling stimuli from the environment or creating stimuli in which environmental properties are preserved. This departs from systematic design. The authors review the development of representative design, examine its use in judgment and decision-making research, and demonstrate the effect of design on research findings. They suggest that some of the practical difficulties associated with representative design may be overcome with modern technologies. The importance of representative design in psychology and the implications of this method for ecological approaches to cognition are discussed.

> There is little technical basis for telling whether a given experiment is an ecological normal, located in the midst of a crowd of natural instances, or whether it is more like a bearded lady at the fringes of reality, or perhaps like a mere homunculus of the laboratory out in the blank. (Brunswik, 1955c, p. 204)

If a student of psychology or philosophy had a time machine and could choose when and where to study, early 20th-century Vienna would be a good choice. Its sparkling intellectual atmosphere produced and brought together great thinkers such as Rudolf Carnap, Herbert Feigl, Kurt Gödel, Otto Neurath, Moritz Schlick, and Ludwig Wittgenstein—all united in their interest in questions of epistemology and the philosophy of science. One individual who was fortunate enough to have been there at the time was Egon Brunswik (Leary, 1987). Brunswik was trained in Karl Bühler's Vienna Psychological Institute, and he participated in discussions with members of the Vienna Circle, who promoted the ideas of logical positivism. It was to Bühler and Schlick whom Brunswik submitted his doctoral thesis in 1927. In the minds of numerous people, Brunswik was to become one of the most outstanding and creative psychologists of the 20th century.

Mandeep K. Dhami, Department of Psychology, City University, London, England; Ralph Hertwig, Institute for Psychology, University of Basel, Basel, Switzerland; Ulrich Hoffrage, Max Planck Institute for Human Development, Berlin, Germany.

We are grateful to Michael Birnbaum, James Cutting, Gerd Gigerenzer, Ken Hammond, David Mandel, Gary McClelland, James Shanteau, and Tom Stewart for many constructive comments on ideas presented in this article. We also thank Tina-Sarah Auer, Lydia Lange, and Carsten Kissler for their assistance in the literature searches and Anita Todd for editing a draft of this article.

Correspondence concerning this article should be addressed to Mandeep K. Dhami, Department of Psychology, City University, Northampton Square, London EC1V 0HB, England. E-mail: m.k.dhami@city.ac.uk

During his career, Brunswik developed a comprehensive theoretical framework referred to as *probabilistic functionalism* and a corresponding methodological innovation called *representative design* (for reviews, see Hammond, 1966; Hammond & Stewart, 2001a; Postman & Tolman, 1959). For Brunswik (1952), an organism's behavior is organized with reference to some end or objective that it has to achieve in an inherently probabilistic world. To study these *organism–environment relations*, stimuli should be sampled from the organism's natural environment so as to be representative of the population of stimuli to which it has adapted and to which the experimenter wishes to generalize (Brunswik, 1956). As the opening quote of this article illustrates, Brunswik stressed that experimenters should first avoid oversampling highly improbable stimuli in a population and, second, avoid using stimuli that do not exist in the population.

Whereas initially Brunswik's ideas were largely ignored (Gigerenzer, 2001), they were later silently—insofar as it happened without acknowledgment of their source—absorbed into mainstream psychology (Hammond & Stewart, 2001b). Today, for instance, the notion that cognitive functioning is adapted to the structure of the environment or the demands of the task is commonplace. However, there appears to have been little acceptance of the method of representative design. Of this method, James J. Gibson (1957/2001) wrote that

> [Brunswik] realized more than his contemporaries that the problems of perception, as of behavior, cannot be solved by setting up situations in the laboratory which are convenient for the experimenter but atypical for the individual. He asks us, the experimenters in psychology, to revamp our fundamental thinking. . . . It is an onerous demand. Brunswik imposed it first on his own thinking and showed us how burdensome it can be. (p. 246)

Gibson expressed this view almost 50 years ago, shortly after Brunswik's death. Since then, psychology has developed considerably, and its methodology has evolved. The main goal of the present article is to explore whether the notion of representative stimulus sampling is of relevance to present-day research practice in cognitive psychology. We focus on research conducted within the field of judgment and decision making, because it is here that Brunswik's ideas have had the greatest impact.

We believe the notion of representative design will eventually become an important instrument in our methodological toolbox. Our assertion lies partly in the impact that representatively designed studies have already had in shaping policy and, consequently, human lives; partly in the rising concern with the limited external validity and generalizability of research findings; and partly with the inclusion of ecological factors in contemporary cognitive theories. However, we do not advocate a dogmatic approach to methodology. The problem of how to conduct experiments such that they enable substantive inductive inferences has no universally accepted solution. We believe that informed judgment, instead of dogmatism, should aid decisions about when to use representative design.

In Part 1 of this article, we review the development of the method of representative design. In Part 2, we analyze a present-day research practice, namely, that of researchers who have declared their commitment to Brunswik's ideas and methods. In Part 3, we examine whether representative stimulus sampling matters for the results obtained in psychology experiments. Finally, in Part 4, we discuss the future of representative design.

Part 1: Theoretical Foundations and Evolution of Representative Design

Brunswik's methodological ambition was grand. He proposed an alternative to *systematic design*, which was the methodological dictate of the time. In this design, experimenters select and isolate one or a few independent variables that are varied systematically while holding extraneous variables constant or allowing them to vary randomly. Experimenters then observe the resulting changes in the dependent variable(s). This design places a strong emphasis on the notion of *internal validity*, that is, on the sound, defensible demonstration that a causal relationship exists between two or more variables. Internal validity is ensured sometimes at the expense of *external validity*, which refers to the generalizability of a causal relationship beyond the circumstances under which it was studied or observed. After all, so the logic goes, if internal validity is weak, then one cannot draw confident conclusions about the effects of the independent variable, and the findings should consequently not be generalized.

According to Gillis and Schneider (1966), the idea of searching for one-to-one, cause–effect relations by disentangling causes via experimentation has many historical roots dating back to the early Greeks. Systematic design became a popular method for disciplines such as medicine and physics. The notion of such experimentation in psychology was greatly influenced by the philosopher J. S. Mill and was adopted by the founding fathers of experimental psychology, Wilhelm Wundt and Hermann Ebbinghaus, who believed psychology could discover cause–effect relations very much like the nomothetic, law-finding tradition of the natural sciences. These early leading figures were trained in the natural sciences. They wished to establish psychology as a science and dissociate themselves from its philosophical roots. This may partly explain the acceptance of systematic design in psychology (Gillis & Schneider, 1966). It has since been featured in textbooks on experimental psychology and become synonymous with the experimental method per se. One example is Woodworth's (1938) classic textbook, *Experimental Psychology*, also known as the "Columbia Bible," which espoused systematic design as *the* experimental tool.

Brunswik opposed psychology's accepted experimental method (Kurz & Tweney, 1997). He questioned both the feasibility of disentangling variables and the realism of the stimuli created in doing so. In what he considered to be the simplest variant of systematic design, the one-variable design, Brunswik (1944, 1955c, 1956) pointed to the fact that variables were "tied," thus making it impossible to determine the exact cause of an observed effect. For example, in the Galton bar experiment, observers are presented with the task of estimating the physical size of differently sized objects. The distance between the objects and the observers is held constant; consequently, physical size and retinal size are artificially "tied," because a small object will project a smaller retinal size and a large object will project a larger retinal size. Brunswik (1944, 1955c, 1956) further argued that in the sophisticated variant of systematic design, namely, factorial design, variables are artificially "untied." Here, the range of the variables is arbitrary and is divided into k levels. The levels of one variable are combined with the levels of another variable, exhausting all possible combinations. All combinations have equal frequency, and the natural covariation among variables is eliminated. Factorial designs may, therefore, yield artificial combinations that are impossible in the real world. Such stimuli may inhibit a researcher's ability to examine how psychological processes function and thus limit the generalizability of research findings.

Cognitive Psychology as a Science of Organism–Environment Relations

Brunswik's methodological innovations were closely intertwined with his theoretical outlook (see Kurz & Hertwig, 2001). In his theory of *probabilistic functionalism*, Brunswik (1943, 1952) argued that the real world is an important consideration in experimental research because psychological processes are adapted in a Darwinian sense (Hammond, 1996b) to the environments in which they function. The main tenets of probabilistic functionalism are illustrated in his *lens model* as presented in Figure 1. The double convex lens shows a collection of proximal effects (cues) diverging from a distal stimulus (criterion or outcome) in the environment. The proximal effects may be used as proximal cues by the organism for achieving the distal variable and so converge at the point of a response in the organism. To illustrate the lens model as applied to a study of bail decision making (Dhami, 2001), one can assume that in order for a judge to decide whether to release a defendant on bail or to remand him or her in custody, the judge must predict the likelihood of the defendant absconding (i.e., the distal variable).

An organism's goal is to infer a distal variable in the environment by using proximal variables. The environment, which is defined as the "natural–cultural habitat of an individual or group" (Brunswik, 1955c, p. 198), consists of reference classes that define populations of stimuli (Brunswik, 1943). Stimuli may be considered as sets of "variate packages" because they comprise a number of cues whose values vary along their range (Brunswik, 1955c).

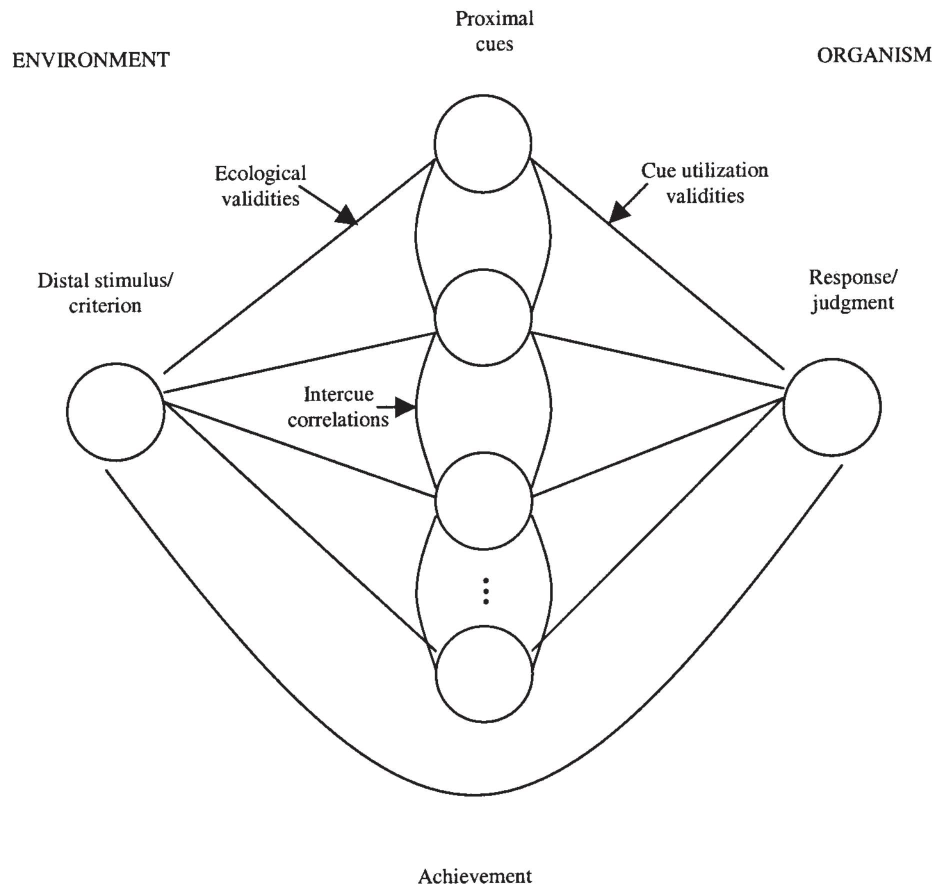

Figure 1. Adapted lens model. From "The conceptual framework of psychology." In *International Encyclopedia of Unified Science* (p. 678), by E. Brunswik, 1952, Chicago: University of Chicago Press. Copyright 1952 by the University of Chicago Press.

For example, in a bail hearing, criminal cases are stimuli that comprise cues such as the defendant's age, offense, and community ties. In one case, the defendant may be young, charged with a serious assault, and may not have any permanent residence. In another case, the defendant may also be young but may be charged with a minor burglary and be residing with his or her family.

The environment to which an organism must adapt is not perfectly predictable from the cues (Brunswik, 1943). A particular distal stimulus does not always imply specific proximal effects, because under some conditions an effect may not be present. Similarly, particular proximal effects do not always imply a specific distal stimulus, because under some conditions the same effect may be caused by other distal stimuli. For example, the strength of a defendant's community ties may be a better predictor of him or her absconding than the seriousness of the offense with which he or she is charged. However, not all defendants without a permanent residence will abscond, and some defendants with a permanent residence may abscond.

Proximal cues are, therefore, only probabilistic indicators of a distal variable. Brunswik (1940, 1952) proposed to measure the ecological (or predictive) validity of a cue by the correlation between the cue and the distal variable (Brunswik, 1940, 1952). The proximal cues are themselves interrelated, thus introducing redundancy (or intra-ecological correlations) into the environment (Brunswik, 1952). For example, when considering the cues that a judge may use to predict absconding on bail, the defendant's gender may be highly related to the type of offense with which he or she is charged (e.g., male defendants may be more likely than female defendants to be charged with committing sexual offenses).

It was assumed that organisms learn the ecological validity of cues and their intercorrelations through experience. In his concepts of *vicarious mediation* and *vicarious functioning*, Brunswik (1952) emphasized the equipotentiality of cues in the environment and the equifinality of an organism's responses, respectively. The environment enables the achievement of a distal variable via alternative proximal cues, and an organism must cope with this uncertainty by learning to alternate between different cues through the "equivalence and mutual intersubstitutability" of these cues (Brunswik, 1952, pp. 675–676). This is particularly the case when previously used cues become unreliable or are unavailable (Brunswik, 1934). In other words, distal variables can be attained vicariously through proximal variables, and proximal variables can themselves be used vicariously through other proximal variables. For example, when predicting whether a defendant will abscond while released on bail, a judge may consider the strength of the defendant's community ties. However, in some cases this information may be unavailable, and so a judge may consider the defendant's age, if he or she believes age to be related to community ties.[1]

Although achievement (also called *functional validity*; Brunswik, 1952) can be maximized by using cues according to their ecological validities (Brunswik, 1943, 1955c, 1957), the fact that cues are only probabilistic indicators of the distal variable means that achievement is at best probabilistic (Brunswik, 1943, 1952). For example, there may be no single cue or combination of cues that allows the perfect prediction of absconding on bail. Brunswik (1940, 1943, 1952) proposed that achievement should be studied at the level of the individual and be measured over a series of responses through the correlation between the distal variable and an organism's responses. For example, to examine how well a judge makes bail decisions, one could examine the correlation between the decisions the judge made over a series of cases with whether the defendant later absconded.

In sum, for Brunswik (1943), the primary aim of psychological research is to discover probabilistic laws that describe an organism's adaptation, in terms of distal achievement, to the causal texture of its environment. The questions to be answered are the following: How is an organism perceiving and responding to its probabilistic environment to achieve a distal variable? Can the findings of such an experiment be used to predict future achievement in that environment? The first question refers to the study of vicarious functioning, and the second refers to the generalizability (external validity) of research findings.

Representative Design: Brunswik's Alternative to Experiments at the Fringes of Reality

Brunswik (1955c, p. 198) proposed that in order to accomplish these two aims experimenters "must resist the temptation . . . to interfere" with the environment and instead strive to retain its natural causal texture in the stimuli presented to participants. In his view, systematic design and its policy of isolating and controlling selected variables destroys the naturally existing causal texture of the environment to which an organism has adapted (Brunswik, 1944). The tasks presented to participants typically "are based on an idealized black–white dramatization of the world, somewhat in a Hollywood style" (Brunswik, 1943, p. 261). He argued that controlling mediation, as is done in systematic design, "leaves no room for vicarious functioning" (Brunswik, 1952, p. 685). Therefore, stimuli should be representative of a defined population of stimuli in terms of their number, values, distributions, intercorrelations, and ecological validities of their variable components (Brunswik, 1956); otherwise, the generalizability of findings will be limited, and researchers will need to be satisfied with "plausibility generalizations, . . . always precarious in nature—or [be] satisfied with results confined to a self-created ivory tower ecology . . . an exercise in neatness" (Brunswik, 1956, p. 110). In other words, Brunswik argued that psychology's preferred method of systematic design destroys the very phenomenon under investigation, or at least, it alters the processes in such a way that the obtained results are no longer representative of people's actual functioning in their ecology. Moreover, he proposed to disentangle variables not at the design stage of research, as is done in systematic design, but at the analysis stage, through techniques such as partial correlation. Nowadays, researchers may use techniques such as analysis of covariance that deal with covariation among independent variables (Cohen & Cohen, 1983), although these techniques are not without problems.

Brunswik also pointed to the "double standard" in the practice of sampling in psychological research (Brunswik, 1943, 1944). To quote Hammond (1998):

> Why, he wanted to know, is the logic we demand for generalization over the subject side ignored when we consider the input or environment side? Why is it that psychologists scrutinize subject sampling procedures carefully but cheerfully generalize their results—without any logical defense—to conditions outside those used in the laboratory. (p. 2)[2]

Since Brunswik's time, the idea that human functioning is shaped by environmental structures has been generally accepted. Furthermore, concerns with the limited generalizability of many research findings have been expressed periodically, and there have been calls for greater external validity, variously termed, of research in many fields of psychology (e.g., Cronbach, 1975; Einhorn & Hogarth, 1981; Elms, 1975; Jenkins, 1974; Klein, Orasanu, Calderwood, & Zsambok, 1993; Koch, 1959; Neisser, 1978).[3] Indeed, Ulric Neisser's (1978, 1982) ecological approach to the study of memory and Urie Bronfenbrenner's (1977) proposal to study the ecology of human development both acknowledge the importance of the environment in shaping human behavior and embody a deep concern for the external validity of research. There have also been specific calls for representative design in the study of clinical judgment (Hammond, 1955; Maher, 1978), interpersonal perception (Crow, 1957), animal behavior (Petrinovich, 1980), aging (Poon, Rubin, & Wilson, 1989), and judgment and decision making (Hammond, 1990). The ecological psychology of Roger Barker (1968) was heavily influenced by Brunswikian ideas. Half a century ago, however, Brunswik's call for a principled alternative to systematic design remained virtually unheard.

Proposals for Implementing a Representative Design

Brunswik (1955c) suggested three ways of achieving a representative design. The first, and preferred way, is by *random sampling* (also referred to as *situational*, *representative*, and *natural sampling*) of stimuli from a defined population of stimuli (reference class) to which the experimenter wishes to generalize the findings. This is one form of what today is called *probability sampling*, in which each stimulus has an equal probability of being selected. To obtain a representative sample using random sampling, the sample size has to be sufficiently large. Brunswik (1944) attempted to randomly sample stimuli by making observations at randomly sampled time intervals in an experiment on size con-

[1] Wolf (1999) more comprehensively defined *vicarious functioning* as the ability for "accumulating, checking, weighting, interchanging, questioning, partly utilizing, rejecting, or substituting" (p. 2) proximal cues in search of alternative pathways to the distal stimulus.

[2] In actuality, as Keppel (1991) pointed out, researchers rarely randomly sample participants in psychological experiments. They more often use random assignment of a very small slice of humankind (e.g., students).

[3] One term that has unfortunately become a popular substitute for external validity and representative design is *ecological validity,* which Brunswik (1943) defined as the validity of a cue in predicting a criterion.

stancy that was later replicated by Dukes (1951). Time sampling, however, is not synonymous with random sampling, as intervals may be sampled systematically (for a review of this sampling method, see Czikszentmihalyi & Larson, 1992). In Brunswik's 1944 study, an experimenter periodically followed a participant (student) going about her daily routine over a 4-week period. At random intervals during the student's day, the experimenter asked her to stop what she was doing and estimate the size of the object that happened to be in her perceptual field at the time. Although he had selected the dependent variables of interest—namely, object size, projective size, and distance—it was the participant who "chose" the object of study. After the student had given her estimate of size, the experimenter also provided an estimate and so acted as a control to the participant. Finally, the experimenter made the objective measurements of the variables. This procedure was repeated in 180 situations, thus justifying generalization to the stimulus population.

The second way of achieving a representative design is through what Brunswik (1955c, 1956) called *canvassing* of stimuli and what today is known as *nonprobability sampling*, namely, stratified, quota, proportionate, or accidental sampling. However, these procedures provide only a "primitive type of coverage of the ecology" (Brunswik, 1955c, p. 204). Third, representative design could theoretically be achieved by a complete coverage of the whole population of stimuli, although Brunswik (1955c) recognized that this might be infeasible.

Brunswik's conception of representative design had some major difficulties, and although others have used these as objections to his method (e.g., Birnbaum, 1982; Björkman, 1969; Hochberg, 1966; Leeper, 1966; Mook, 1989; Sjöberg, 1971), he acknowledged many of them himself. First, there are practical problems in terms of the lack of experimental control and the time-consuming and cumbersome nature of research conducted outside the laboratory (Brunswik, 1944, 1955c, 1956). Brunswik (1956) acknowledged that representative design was a "formidable task in practice" (p. viii) and "ideally would require huge, concerted group projects, involving as it does a practically limitless number of variables" (Brunswik, 1955a, p. 239). In later experiments, he used photographs of objects in the environment to make the research more manageable (Brunswik, 1945; Brunswik & Kamiya, 1953). Hochberg (1966) criticized this approach because static pictures do not allow use of cues afforded by motion and because the pictures were not randomly sampled. Nevertheless, as Doherty (2001) argued, this research "stands as a serious and virtually unprecedented attempt to quantify ecologically relevant aspects of the environment" (p. 212).

Second, there are theoretical problems of defining the appropriate reference class (Brunswik, 1956; see Hoffrage & Hertwig, in press). Brunswik (1956) conceded that his 1944 experiment on size constancy lacked a precisely defined reference class of stimuli. Indeed, population parameters may be influenced by a participant's attention, attitude, and motivation. To avoid the problem of generalizing over attitudes, Brunswik (1944) had obtained observations of behavior influenced by a number of different experimenter-induced attitudes, including naïve and analytical attitudes.

Although the theoretical problems of sampling have not been resolved, attempts have been made to overcome the practical difficulties associated with representative design. Brunswik (1944) himself proposed that it is "generally possible" and "practically often very desirable" (p. 42) to use a hybrid design in which the researcher introduces certain elements of systematic design into an experiment in which a representative design is used. For instance, in a social perception experiment, Brunswik and Reiter (1937) used a truncated factorial design in which unrealistic schematized faces created by a factorial design were omitted. Earlier, in an experiment on perceptual constancy, Holaday (1933) systematically stripped an exemplary stimulus of its complexity through "successive omission" of cues. Brunswik (1956) lamented that hybrid designs would be more popular than representative design, but even his own experiments moved toward such designs.

Whereas the theoretical necessity of representative design in psychological research may seem more reasonable at present, Brunswik's (1943) call for a fundamental shift in the method of psychology did not occur. His contemporaries were united in their wish to maintain the status quo (Postman, 1955). Despite the fact that his ideas were published in leading journals, they were largely ignored, misunderstood, and treated with skepticism and hostility (Feigl, 1955; Hilgard, 1955; Hull, 1943; Krech, 1955; Postman, 1955), even by those who shared his belief in the importance of studying the ecology (Lewin, 1943). Hilgard (1955) demonstrated his skepticism when he argued that "representative designs are no more foolproof than the other types of systematic design" (p. 227). Brunswik's (1955a) rejoinder was insufficient. Suffering from a painful and intractable disease, Brunswik committed suicide on July 7, 1955, at the age of 52. By his untimely death, he relinquished an opportunity to fully rebut his critics and to refine the concept of representative design.

After Brunswik: Hammond's Interpretation of Representative Design

Those close to Brunswik also pointed to the difficulties associated with designing experiments to meet his criteria for representativeness (Gibson, 1957/2001; Hammond, 1966; Tolman, 1955). Björkman (1969) suggested that the concept of representativeness be replaced by the idea of ecological relevance, so "*rather than* [samples] *being representative they should be samples biased towards those aspects which are considered important for man's adaptation in the real world*" (p. 153). Hochberg (1966) and Leeper (1966) proposed that some form of unrepresentative, extreme group sampling procedure would be sufficient to assess the validity of generalizations based on selective samples of variables. However, these proposals seem to dilute the notion of representative design. The person who most determinedly took up the challenge to refine and further develop the concept was Kenneth R. Hammond. Hammond, a student of Brunswik at Berkeley, extended Brunswik's ideas to the study of higher cognition, namely, judgment and decision making (e.g., Hammond, 1955; Hammond, Stewart, Brehmer, & Steinmann, 1975). Over the past 50 years, Hammond has used the notion of external validity and the requirements of representative design to critically evaluate the merits of psychological research in general and research on the normative analysis of rational behavior in particular (Hammond, 1948, 1951, 1954, 1955, 1978, 1986, 1990, 1996a, 2000; Hammond, Hamm, & Grassia, 1986; Hammond & Wascoe, 1980).

Hammond (1966) asked of representative design "Can it be done?" (p. 66). In an attempt to overcome its practical difficulties, he differentiated between the concepts of *substantive situational sampling* and *formal situational sampling*. The former focuses on the content of the task (e.g., size constancy) with its inherent formal properties and is analogous to Brunswik's original definition of representative design. Formal situational sampling, on the other hand, focuses on the formal properties of the task (i.e., number of cues, their values, distributions, intercorrelations, and ecological validities), irrespective of its content. The formal properties define the universe of stimulus (or situation) populations. For instance, cue number, values, and distributions range from 0 to infinity, and the ecological validities of cues and their intercorrelations range from –1 to +1. Any population of situations lies within these boundaries. Brunswik (1955c) believed that "there will be a limited range and a characteristic distribution of conditions and condition combinations" (p. 199). A researcher who uses formal situational sampling can sample various combinations of formal properties (e.g., various numbers of cues, ecological validities, and intercue correlations). In addition, a comprehensive study of a particular issue is not excluded a priori (B. Brehmer, 1979).

Whereas Brunswik (1944) had taken the researcher and the participant outside the laboratory, formal situational sampling returns them both to the laboratory, where it permits the construction and presentation of stimuli that are formally representative of the natural stimulus population. This approach is supported by Brunswik's (1934) early work.

As an illustration of formal situational sampling in practice, consider a study conducted by Hammond, Hamm, Grassia, and Pearson (1987). Hammond et al. (1987) captured the policies of 21 expert highway engineers on judging highway safety. The distal criterion was the rate of accidents divided by the number of miles traveled, averaged over 7 years, for each of 40 highways. The highways were described in terms of 10 cues (e.g., lane width) that highway safety experts identified as essential information for judging road safety. The values, intercorrelations, distributions, and ecological validities of eight cues were deduced from highway department records, and the properties of two cues were measured by the experimenters from visual inspection of videotapes of each highway. The researchers were primarily interested in examining how presentation of the task affects the mode of cognition used (e.g., analytical), and so the cue information for each highway was presented via filmstrips, bar graphs, and formulas in the three presentation conditions.

To identify the formal properties of the task to be represented in the experiment, it is clear that researchers should first familiarize themselves with the task by conducting some form of task analysis (Cooksey, 1996; Hammond, 1966; Hammond et al., 1975; Petrinovich, 1979). This may be based, for example, on interviews with individuals experienced with the task, observations of individuals performing the task, or document analysis of past cases. Stewart (1988) proposed that if formal properties such as intercue correlations are unavailable or unknown, subjective ones might be extracted from participants. Thus, researchers can sample from a population of situations as perceived by the participant. Moreover, researchers can sample from both an existing population and from possible populations of situations, as in multiple-cue probability learning experiments (for a review, see Klayman, 1988).

As B. Brehmer (1979) argued, however, formal situational sampling is "no easy road to success" (p. 198) either, and because the number of all possible combinations may be extremely large, the researcher needs to know which are the important combinations to study. Indeed, if *important* is used to mean representative, and the researcher is interested in sampling from an existing population of situations, then the problem of defining a reference class or sampling frame looms large. Nevertheless, Hammond (1972) concluded that formal situational sampling is "clearly feasible," and he advocated that until technological advances allowed substantive situational sampling, researchers should use formal situational sampling (Hammond, 1966). Nowadays, computers, film, and tape are readily available and can be effectively used to capture and reproduce environments (e.g., P. N. Juslin, 1997; Shaw & Gifford, 1994; Stewart, Middleton, Downton, & Ely, 1984). In sum, whereas Brunswik proposed sampling real stimuli from the environment, Hammond (1966) advocated creating stimuli that were representative in terms of the formal informational properties of the environment.

Social Judgment Theory, Policy Capturing, and the Importance of Expertise

Throughout his career, Hammond has promoted and developed a Brunswikian approach to the study of human judgment and decision making. In 1975, he and his colleagues synthesized research conducted within the lens model framework under the rubric of *social judgment theory* (Hammond et al., 1975). This is not a theory that provides any testable hypotheses about the nature of human judgment but a meta-theory that provides a framework to guide research of "life relevant" judgment problems (Hammond et al., 1975). Here, researchers aim to describe the nature of human judgment in specific task domains with a view to developing ways for improving performance (A. Brehmer & Brehmer, 1988; Hammond et al., 1975).

Social judgment theory is characterized by four varieties of the lens model. The *single-system design* refers to a situation in which the criterion variable is either unavailable or is of no interest and thus represents only the subject side of Brunswik's (1952) lens model. In the *double-system design*, the parameters of the ecological side of the lens model are known (see Figure 1). This design has been used to study achievement or judgment accuracy (Hammond, 1955), multiple-cue probability learning (Hammond & Summers, 1965; Klayman, 1988), and cognitive feedback (Balzer, Doherty, & O'Connor, 1989; F. J. Todd & Hammond, 1965). The *triple-system design,* which is an expansion of the lens model, involves a task situation and two people making use of the same probabilistic cues. This has enabled the study of interpersonal learning (Earle, 1973; Hammond, 1972; Hammond, Wilkins, & Todd, 1966) and interpersonal conflict (B. Brehmer, 1976; Hammond, 1965, 1973; Hammond, Todd, Wilkins, & Mitchell, 1966; Mumpower & Stewart, 1996). Finally, the *N-system design* involves more than two people and may or may not include an outcome criterion, and so it enables a study of group judgment (e.g., Rohrbaugh, 1988). For pictorial representations of these designs, see Hammond (2001).

Most research has been conducted using the single-system design, in which there is no outcome criterion.[4] Here, researchers simply capture and describe an individual's judgment policy (for a review, see A. Brehmer & Brehmer, 1988). To extract a person's policy, social judgment theorists use the techniques of *judgment analysis* (Christal, 1963) or *policy capturing* (Bottenberg & Christal, 1961). These techniques have been fully explicated in a number of publications (Cooksey, 1996; Stewart, 1988). Suffice it to say that individuals make decisions on a set of either real or hypothetical cases that comprise a combination of cues. Each individual's judgment policy is then inferred from his or her behavior, traditionally through the use of multiple linear regression analysis. An individual's judgment policy is described in terms of, among other things, the number and nature of cues used to make judgments. Achievement may be measured by correlating the individual's judgments with the criterion values and comparing his or her policy with a model of the task. Agreement in decisions and policies among individuals may also be examined. Similarly, participants' self-reported policies may be compared with their captured policies, in what is traditionally considered a measure of self-insight (e.g., Cook & Stewart, 1975; Summers, Taliaferro, & Fletcher, 1970). Finally, intra-individual inconsistency in judgments may be studied by comparing the judgments made in a test–retest situation.

In his study of elementary processes such as perception, Brunswik could safely assume that the participant was a "natural expert" at the task. By contrast, social judgment theorists study higher cognitive processes in what are often socially constructed tasks. Therefore, an important issue for the reliability of the captured policy is the experience of the participant. Participants who are experienced with the task will be more sensitive to the face and construct validity of the task (Shanteau & Stewart, 1992). They may recognize, for example, that some of the relevant cues have not been presented, and this may prevent them from expressing their natural judgment behavior. Inexperienced participants, or novices, by contrast, will not have any developed policy to be captured (A. Brehmer & Brehmer, 1988).

The judgment policies of professionals have been captured in a variety of domains, including medicine, education, social work, and accounting (for reviews, see Wigton, 1996; Heald, 1991; Dalgleish, 1988; and Waller, 1988, respectively). Although there are exceptions (see A. Brehmer & Brehmer, 1988), periodic reviews have concluded that studies yield consistent findings, irrespective of the number and type of decision makers sampled and the nature and content of the judgment tasks studied (A. Brehmer & Brehmer, 1988; B. Brehmer, 1994; Cooksey, 1996; Hammond et al., 1975; Libby & Lewis, 1982; Slovic & Lichtenstein, 1971). In general, achievement is high and judgments are considered to be the result of a linear, additive process in which a few, differentially weighted cues are used. People show some degree of inconsistency in their judgments, however, and there are interindividual differences in judgment policies for the same task. Finally, self-reports of policies tend to differ from captured policies.

Part 2: Is Representative Design Used to Capture Judgment Policies?

Social judgment theorists adopt a Brunswikian approach to studying judgment and decision making. They have expressed a commitment to the method of representative design, which, they argue, differentiates them from other researchers in cognitive psychology in general and judgment and decision making in particular (e.g., see B. Brehmer, 1979; Cooksey, 1996; Hammond et al., 1975; Hammond & Wascoe, 1980; Hastie & Hammond, 1991). For instance, Hastie and Hammond (1991) claimed that "the Lens model researchers' commitment to 'representative design' is explicit (and enthusiastic)" (p. 498). Similarly, Cooksey (1996) stated that "The critical dimension of . . . research which distinguishes it from nearly all other research endeavors in the social and behavioral sciences is its insistence upon applying the principle of *representative design* to guide the structure of specific investigations" (p. 98). We now analyze the extent to which these "neo-Brunswikian" researchers have lived up to their stated ideal of implementing a representative design in their studies.

In this section we compare neo-Brunswikian research practices with that of a parallel yet largely unconnected research tradition. It is interesting that the techniques of policy capturing and judgment analysis have also been used by researchers who do not conduct their research within the lens model framework (e.g., Dudycha & Naylor, 1966; Madden, 1963; Naylor, Dudycha, & Schenck, 1967; Naylor & Wherry, 1964, 1965). For instance, Bottenberg and Christal (1961) coined the phrase *policy capturing*, which refers to the analysis of judgment data using multiple regression techniques. Christal (1963) developed judgment analysis as a statistical technique for analyzing group judgment that reveals similarities and differences among group members, thereby helping them reach a consensus in their judgment policy. Although these researchers also capture judgment policies in applied domains, they tend to focus solely on describing agreement among individuals and do not study achievement. In contrast to neo-Brunswikians, they are less likely to be aware of the concept of representative design and so may be less likely to use it. Therefore, research conducted outside the Brunswikian tradition provides a "baseline condition" in the use of representative design against which neo-Brunswikian practices can be compared.

Method

To identify the relevant population of studies, we conducted a literature search on the following databases: PsycINFO (journals and books) from 1935 to July 1999, Social Sciences Citation Index from 1973 to July 1999, Sociofile from 1974 to July 1999, and EconLit from 1969 to July 1999. The applied nature of the research necessitated coverage of a broad range of social science databases. We selected entries whose title or abstract contained at least one of the central theoretical and methodological concepts associated with Brunswik, social judgment theory, policy capturing, and

[4] An outcome criterion may be unavailable for quite valid reasons. First, an outcome criterion may not be useful because there is no correct answer, as, for example, in the diagnosis of a mental illness (Doherty, 1995, as cited in Cooksey, 1996). Second, an outcome criterion may be difficult to obtain because of concerns with confidentiality, ethics, or legality. Third, collection of all outcomes may be theoretically impossible (e.g., Dhami & Ayton, 2001). Fourth, an outcome criterion may be unavailable during the study period. Fifth, studies using hypothetical cases or cases that represent future situations by their very nature preclude the use of an outcome criterion. One way to overcome problems in obtaining an outcome criterion is to use expert judgments as environmental criterion measures (e.g., Hammond & Adelman, 1976; Mumpower & Adelman, 1980).

judgment analysis researchers. The keywords were *representative design*, *probabilistic functionalism*, *social judg(e)ment theory*, *lens model*, *judg(e)ment analysis*, and *policy-capturing*.

A cursory glance at the abstracts of entries selected at the first stage of searching indicated that many entries in which the keywords appeared were not relevant to our analysis. We therefore analyzed the contents of each abstract in more detail, in search of the population of *empirical* studies with which we were concerned. The keywords used at this stage were *subjects*, *participants*, *study*, *method*, *procedure*, *empirical*, *data*, *findings*, and *results*. This procedure identified entries containing empirical studies. We discarded the following entries: nonempirical studies; empirical studies reporting data without presenting the method or data (the majority were abstracts of unpublished dissertations); empirical studies in which cues were not combined to create stimuli but where participants simply ranked, rated, or made paired comparisons of the cues; empirical studies on policy learning or policy feedback; studies from the field of visual perception that concerned the lens of the eye (elicited by the *lens model* keyword); and studies on Sherif and Hovland's (1961) social judgment theory. There were also numerous repeats among databases, which we discarded. This second stage of searching yielded a total of 130 entries reporting on 143 empirical studies (some entries reported more than 1 study). These articles are listed on the Brunswik Society Web site: http://brunswik.org/resources/RepDesignRefs.pdf.

Next, we developed and used a structured coding scheme to analyze the methodology used in the studies. The coding scheme contained a number of checklist variables pertaining to the methodological details of studies. The variables were chosen on the basis of our review of the literature on representative design, social judgment theory, policy capturing, and judgment analysis. These variables provide an empirical basis for our evaluation of the extent and nature of representative design used in research. A copy of the full scheme is available on request from Mandeep K. Dhami. The variables can be grouped into the categories, as shown in Table 1.

Mandeep K. Dhami familiarized herself with the coding scheme by piloting it on a set of 100 studies selected randomly from the 143 studies. It was evident that some studies reported details of the methodology in the introduction section as well as the method section, so both were analyzed. The scheme was then used to code the designs used by the complete population of 143 studies. Ambiguous cases were discussed and resolved among all three authors.

Results

We found that 74 studies could be classified as neo-Brunswikian because they cited one or more of Brunswik's (1943, 1944, 1952, 1955c, 1956) articles or cited articles that discuss the theory and method of social judgment theory or lens modeling (A. Brehmer & Brehmer, 1988; B. Brehmer, 1979, 1980; Cooksey, 1996; Hammond, 1966, 1972; Hammond et al., 1975; Petrinovich, 1979; Stewart, 1988). The remaining 69 studies were classified as being outside of the Brunswikian tradition and were considered a baseline condition. Of this latter group, 21 studies cited one or more articles that introduced the techniques of policy capturing and judgment analysis (Bottenberg & Christal, 1961; Christal, 1963; Dudycha, 1970; Dudycha & Naylor, 1966; Madden, 1963; Naylor et al., 1967; Naylor & Wherry, 1964, 1965), and 48 studies cited none of the above; however, rather than discarding them we included them in the baseline condition because they did not explicitly demonstrate an awareness of Brunswik's concept of representative design. The two research traditions were largely unconnected, with one exception that made reference to literature from both traditions (Hoffman, Slovic, & Rorer, 1968); we placed this study in the neo-Brunswikian tradition category as it showed an awareness of Brunswik's ideas.

Thus, we compared the research practices of neo-Brunswikians with researchers working outside of the Brunswikian tradition who also capture judgment policies but do not share the neo-Brunswikians' theoretical position. The aims of the studies in each tradition, the judgment domain being investigated, the participants, their experience with the task being studied, the relevance of the task to them, and the dependent variable being measured are presented in Table 2. One can see that the two groups of studies differ on only a few dimensions. (Because we were comparing two populations rather than sample means, we did not use statistical tests of the differences.)

Do neo-Brunswikians study achievement? Influenced by Karl Bühler's biologically motivated concern with the success of or-

Table 1
Summary of Coding Scheme

Category	Item
Intellectual tradition	Did researchers cite Brunswik's articles on probabilistic functionalism and representative design or Hammond's work on the development of social judgment theory and representative design?
	Did researchers cite the literature involved in the development of the judgment analysis or policy-capturing techniques?
Scope of analysis	Did researchers study achievement and/or agreement?
Stimuli presented	Did researchers use real or constructed cases?
Substantive situational sampling	How were real cases sampled?
Formal situational sampling	How was the task analysis conducted?
	What task information was elicited?
	How were cues combined to form cases?
	How were cues coded?
	Did researchers claim the task to be realistic?
Participants	Who were the participants?
	Were they experienced with the task?
	Did researchers claim the task to be relevant to participants?
Dependent measures	What was the dependent variable?
	How was it measured?
Other information	What was the judgment domain studied?
	What methods and procedures were used to collect data?

Table 2
Overview of Neo-Brunswikian Studies and Studies Outside the Brunswikian Tradition

Characteristic	Neo-Brunswikian studies ($N = 74$)		Studies outside the Brunswikian tradition ($N = 69$)	
	%	*n*	%	*n*
Scope/aim of study				
Study agreement only	72	53	97	67
Study agreement and achievement	28	21	3	2
Domain of study				
Clinical	16	12	1	1
Education	7	5	22	15
Legal	1	1	4	3
Management/business	8	6	41	28
Marketing/advertising	4	3		
Medical	19	14	3	2
Personnel/occupational	7	5	19	13
Other professional	4	3	1	1
Social perception	14	10	4	3
Other	20	15	4	3
Participants				
Professionals	50	37	71	49
Students	35	26	26	18
Both	4	3	1	1
Other	8	6	1	1
No information provided	3	2		
Participants' level of experience				
Experienced	81	60	75	52
Familiar	12	9	17	12
Inexperienced	1	1	6	4
Combination	5	4	1	1
Relevance of task to participants				
Relevant	62	46	61	42
Not relevant				
No information provided	38	28	39	27
No. of participants	Range = 1 to 283, *mdn* = 38		Range = 4 to 2,052, *mdn* = 54	
Dependent variable				
Decision/judgment	12	9	9	6
Estimation/prediction	10	7		
Likelihood	15	11	13	9
Money			9	6
Probability	5	4	1	1
Ranking			4	3
Rating	35	26	35	24
Other	14	10	10	7
Combination of the above	10	7	15	10
No information provided			4	3
Scale used to measure dependent variable				
Binary	5	4	6	4
Continuous/rating	62	46	67	46
Multicategorical	5	4	6	4
Other	16	12	7	5
Combination of the above	11	8	15	10

Note. Percentages may not sum to 100 because of rounding.

ganisms in their environments, Brunswik focused on the study of achievement.

In fact, he believed that the study of distal achievement should be the primary aim of psychological research (Brunswik, 1943). In this sense, for theoretical rather than methodological reasons, Brunswik always placed great emphasis on the utility of including a distal variable in one's research, as achievement can be studied only through reference to a distal variable. Thus, it is interesting to examine whether the two groups of studies conform to Brunswik's theoretical position. They mostly do not (see Table 2). The majority of studies outside the Brunswikian tradition (97%) and the neo-Brunswikian studies (72%) described participants' judgment policies and compared policies among participants without reference to their degree of achievement. In some studies, this was clearly due to the choice of a research topic, which, for principled reasons, led to the unavailability of an outcome criterion. For instance, York (1992) used court records to capture the policies of federal district and appellate courts in deciding sexual harassment cases. It was not feasible to discover whether their decisions were accurate.

Although the relative neglect of the study of achievement by neo-Brunswikians is surprising in light of Brunswik's (1943, 1952) emphasis on achievement as *the* topic of psychological research, the lack of a distal variable does not automatically imply that a study does not use representative design. The inclusion or exclusion of a distal variable is orthogonal to the question of whether the design is representative. To the extent that researchers who do not include a distal variable nevertheless represent the situation toward which they are generalizing, they are considered to be using a representative design. Thus, one may interpret the rarity of studies including distal variables as a reflection of the fact that neo-Brunswikians are not tied to achievement-oriented research questions.

Do neo-Brunswikians study "mature" policies? Inexperienced participants will not have any developed policy to be captured, and so neo-Brunswikian policy-capturing research should be committed to studying people who are experienced at the task.

Consistent with this standard, half of the studies in the neo-Brunswikian tradition relied on professionals as participants, and four-fifths relied on experienced participants (see Table 2).[5] These proportions were also comparatively high in studies conducted outside the Brunswikian tradition. In this sense, the practices in both traditions differ dramatically from practices in psychology. For instance, Sieber and Saks (1989) reported that, in 1987, 74% of studies reported in the *Journal of Personality and Social Psychology* used students from department subject pools. Table 2 also shows that a large proportion of neo-Brunswikian studies were conducted in the clinical and medical domains. This may be explained by Hammond's (1955) early application of Brunswikian principles to clinical judgment.

Have neo-Brunswikians dispensed with the "double standard" in sampling, and do they achieve representative stimulus samples? To examine the representativeness of the stimuli (cases) presented to participants, we first distinguished between studies that sampled real cases from the environment and studies that created hypothetical cases.

The use of real cases maps onto Brunswik's (1944, 1955c, 1956) original conception of how a representative design may be achieved and what Hammond (1966) referred to as *substantive situational sampling*. The construction of hypothetical cases using formal situational sampling refers to Hammond's (1966) proposal for achieving a representative design. In 43% ($N = 32$) of neo-Brunswikian studies, participants were presented with real cases, compared to only 19% ($N = 13$) of studies outside the Brunswikian tradition. The remaining 57% of neo-Brunswikian studies ($N = 42$) used hypothetical cases constructed by the experimenter, compared to 81% ($N = 56$) of studies outside the neo-Brunswikian tradition. We first review the sampling of real cases by neo-Brunswikians (excluding studies outside the Brunswikian tradition because of their small sample size).

As stated earlier, Brunswik (1957) placed equal emphasis on the environment and the organism. He accused his peers of using a double standard in applying the logic of induction to the person but not to the environment (Brunswik, 1943). As Hammond and Stewart (2001b; see also Hammond, 1948, 1954; Wells & Windschitl, 1999) pointed out,

> Articles regularly appear in American Psychological Association (APA) journals describing experiments in which many subjects (conforming to the logic of induction on the subject side) but only one or two or three "person–objects" are used, thus ignoring the need for sampling on the object, or environmental, side. (p. 5)

According to Hammond (1986), this propensity to substitute the number of participants for the number of conditions in the test of the null hypothesis is an error endemic to the use of systematic design and has led to many false rejections of the null hypothesis.

It seems fair to conclude that neo-Brunswikians using what Hammond (1966) called *substantive situational sampling* have made a serious effort to sample participants and objects (stimuli). In fact, the median number of stimuli presented to participants was approximately twice as large as the median number of participants involved in the study (i.e., 51 and 26, respectively). As shown in Table 3, probability sampling was frequently used to sample cases, typically via random time intervals. Nonprobability sampling also was common. Rarely was the whole population of stimuli sampled. On a more practical note, it is interesting to observe that approximately one third of studies that sampled real cases used audio, video, and photographic technology, as envisaged by Hammond (1966). In sum, neo-Brunswikian studies that used substantive situational sampling involved both participant and stimulus sampling. This conclusion is tempered only by the fact that approximately one third of neo-Brunswikian studies that used real cases did not provide any information about the stimuli-sampling procedure used.

We now turn to those studies in both groups that used hypothetical cases. These studies are particularly interesting. Although the representativeness of real cases is a function of the sampling procedure used, with hypothetical cases representativeness is possible if researchers conduct what most likely amounts to a time-consuming and effortful task analysis. On the basis of the outcome of this analysis, they can construct cases that are representative of the environment to which they want to generalize. Therefore, the use of hypothetical cases represents a more demanding test of neo-Brunswikians' commitment to representative design.

At first glance, the commitment seems serious, because the sampling of cases and participants received equal weight (i.e., the median numbers of sampled cases and participants are 49 and 52, respectively). A second glance, however, calls their commitment into question. The key to this observation lies in the paucity of data collected by researchers' task analyses. In Hammond et al.'s (1987) study, cited earlier as an ideal example of formal situational sampling, the researchers' task analysis provided information on the cues, their values, distributions, intercorrelations, and ecological validities. On the basis of this information it was possible to generate representative hypothetical cases (i.e., highways). The data in Table 4 show that in a large majority of studies in both groups a task analysis was conducted and that neo-Brunswikians typically learned about the task by interviewing or conducting a pilot study with individuals experienced with the task or by reviewing literature on the topic. However, details of the formal properties of the environment were often either vague or entirely

[5] About one third of the neo-Brunswikian studies involved student samples. However, such studies typically required students to perform a judgment task with which they were experienced. For example, Aarts, Verplanken, and Knippenberg (1997) studied the modes of transport students favored when traveling to university.

Table 3
Substantive Situational Sampling (of Real Cases) in Neo-Brunswikian Studies and Studies Outside the Brunswikian Tradition

	Neo-Brunswikian studies (N = 32)		Studies outside the Brunswikian tradition (N = 13)	
Characteristic	%	*n*	%	*n*
Sampling procedure used				
Probability/random sampling	38	12	31	4
Nonprobability sampling	22	7	8	1
Sampled whole population	3	1		
No information provided	38	12	62	8
No. of cases sampled	Range = 6 to 330, *mdn* = 51		Range = 13 to 702, *mdn* = 77	
Method used to collect data				
Audio/videotape/photo	31	10		
Computer	3	1		
Document analysis	6	2	39	5
Observation	6	2		
Paper and pencil	44	14	54	7
No information provided	9	3	8	1

Note. Percentages may not sum to 100 because of rounding.

absent, especially regarding the distribution of the cues, their intercorrelations, and ecological validities.

The manner in which cues were combined to form the cases provides us with some additional evidence as to researchers' commitment to representative design. As shown in Table 4, the large majority of neo-Brunswikian studies combined cues using a factorial design and presented cues that had been interpreted by the researcher.[6] Factorial combinations of cue values create rectangular distributions and zero intercue correlations. This suggests that most studies did not maintain representative cue distributions and intercorrelations. Figure 2 illustrates the percentage of studies in both groups that endeavored to represent formal properties of the task in the hypothetical cases. As mentioned earlier, factorial designs increase the risk of studying highly improbable stimuli or stimuli that do not exist in the population (e.g., the "bearded lady"; Brunswik, 1955c). One relatively effortless step to reduce this risk is to screen stimuli and discard or replace atypical or unrealistic cases. We found that none of the neo-Brunswikian studies, and only two of the studies outside the neo-Brunswikian tradition that used a factorial design, filtered out unrealistic cases in this manner, which further calls into question the researchers' commitment to representative design.

Discussion

Social judgment theorists have expressed an explicit and enthusiastic commitment to the method of representative design (e.g., Cooksey, 1996; Hastie & Hammond, 1991). Our analysis has revealed that the actual degree of commitment appears to be a function of a distinction that Hammond (1966) made. The majority of studies that used substantive situational sampling (akin to Brunswik's, 1955c, 1956, original formulation) sampled cases so as to preserve the properties of the ecology. By contrast, studies that used formal situational sampling often failed to represent the ecological properties toward which generalizations were intended. For instance, researchers rarely combined cues to preserve their intercorrelations—an essential condition for the operation of vicarious functioning. In fact, in terms of representing the ecology, there was little difference between the practices of neo-Brunswikians and those working outside the Brunswikian tradition.

Although we can only speculate as to why researchers using formal situational sampling have parted from the Brunswikian ideal of veridically representing the ecology, two possible explanations are worth mentioning. First, considerable time, effort, and resources are required to conduct a task analysis that extracts the formal properties of the ecology (e.g., cue intercorrelations and ecological validities) and that allows researchers to translate them into representative hypothetical cases.

Second, neo-Brunswikian theorists such as Thomas S. Stewart and Ray W. Cooksey have endorsed methodological simplifications that led to substantial deviations from Brunswik's original conception of representative design. These simplifications have often been driven by a concern for data analysis and the requirements of certain statistical tools. For example, Brunswik (1952) considered *vicarious functioning*—the idea that an adaptive system can rely on multiple cues that can be substituted for each other—to be a defining feature of psychological inquiry. Social judgment theorists have modeled vicarious functioning using multiple regression (see Cooksey, 1996; Hammond, Hursch, & Todd, 1964; Stewart, 1976). This tool, however, has certain requirements, for instance, a high case-to-cue ratio to establish stable beta weights

[6] It is argued that cues should be presented in their original units of measurement, being coded in a concrete rather than abstract manner (A. Brehmer & Brehmer, 1988; B. Brehmer, 1979; Cooksey, 1996). For example, age should be expressed in terms of years, rather than as "young," "middle aged," or "old." There is evidence that representation may affect weights attached to cues (Wigton, 1988) and that mode of presentation may affect the type of processing used (e.g., Hammond et al., 1987). If cues are coded in an appropriately abstract manner, it is assumed that the perceptual element of the judgment process has been bypassed (A. Brehmer & Brehmer, 1988).

Table 4
Formal Situational Sampling in Neo-Brunswikian Studies and Studies Outside the Brunswikian Tradition

Characteristic	Neo-Brunswikian studies (N = 42) %	Neo-Brunswikian studies (N = 42) n	Studies outside the Brunswikian tradition (N = 56) %	Studies outside the Brunswikian tradition (N = 56) n
Basis of task analysis				
Interview/pilot study	29	12	20	11
Literature review	26	11	38	21
Real cases	2	1	4	2
Combination	22	9	25	14
No task analysis conducted	19	8	14	8
No information provided	2	1		
Method of combining cues to construct cases				
Factorial design	79	33	77	43
Not factorial, not representative	2	1	2	1
Not factorial, representative	7	3	13	7
No information provided	12	5	9	5
Coding of cues				
Uninterpreted by researcher	19	8	14	8
Interpreted by researcher	67	28	79	44
No information provided	14	6	7	4
Claimed realism of cases				
Realistic	55	23	46	26
Not realistic	2	1		
No information provided	43	18	54	30
No. of cases sampled	Range = 8 to 190, *mdn* = 49		Range = 4 to 250, *mdn* = 50	
Method used to collect data				
Audio/videotape/photograph				
Computer	17	7	5	3
Document analyses				
Observation	2	1		
Paper and pencil	76	32	86	48
No information provided	5	2	9	5

Note. Percentages may not sum to 100 because of rounding.

(Cohen & Cohen, 1983). Thus, Stewart (1988) recommended that the number of cues presented "should be kept as small as possible" (p. 43). Concern with data analysis is also a reason provided for why researchers typically remove intercue correlations. To establish the effect of each cue on the judgments independent of the effect of other cues, some researchers reduce or eliminate intercue correlations, as advocated in the literature (e.g., Cooksey, 1996; Stewart, 1988) and endorsed by reviews of the literature (e.g., A. Brehmer & Brehmer, 1988).

Finally, although Brunswik did not seriously consider the issue of the representativeness of the dependent variable in terms of how participants' responses are elicited and measured, it is clear that this is important. There is evidence that response mode affects all stages of the decision process (e.g., Billings & Scherer, 1988; Westenberg & Koele, 1992). Cooksey (1996) argued that the responses should be measured in the same units as the outcome criterion; otherwise, the task may become more cognitively demanding, as participants convert the units of measurement with which they are familiar into the ones used in the study. Our analysis revealed that the large majority of studies in both groups measured the dependent variable (i.e., judgment or decision) on a continuous rating scale. In some instances, this seemed inappropriate. For example, Harries, St. Evans, Dennis, and Dean (1996) asked physicians to judge the likelihood of them prescribing a lipid-lowering drug on a continuous scale, when in reality physicians make a categorical decision to, or not to, prescribe. Thus, researchers have tended to use representations of the dependent variable that meet the requirements of the statistical techniques they used (e.g., multiple linear regression). In Part 4, we consider whether the implementation of representative stimulus sampling can be fostered by liberating research on vicarious functioning from the straitjacket of multiple regression.

Part 3: Does Representative Design Matter?

Brunswik (1955c, 1956) highlighted two major limitations of systematic design that he believed could be overcome by representative design: (a) Systematic design does not allow researchers to elicit the natural process of vicarious functioning, and as a consequence (b) it does not enable them to generalize the findings of an experiment beyond the experimental situation. These limitations can be translated into two questions concerning research that aims to capture judgment policies. First, do the captured judgment policies reflect participants' "true" policies? If representative design matters, one would expect to find differences in the characteristics of the policies captured under more and less representative conditions. Second, can the captured policies be used to predict participants' behavior on the judgment task for cases

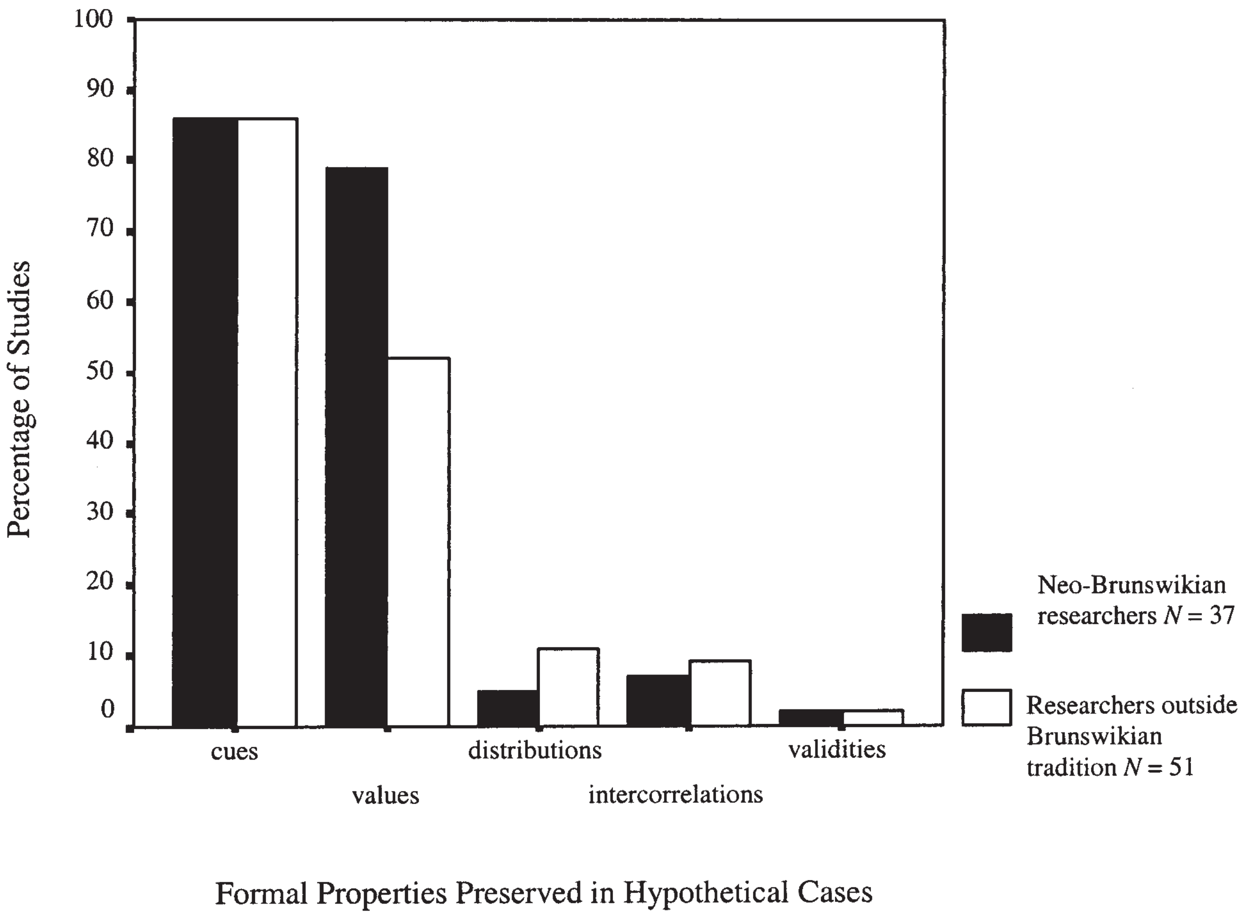

Figure 2. Formal properties of a task preserved in the hypothetical cases constructed by neo-Brunswikian researchers and those working outside the Brunswikian tradition.

outside the laboratory? If representative design matters, one would also expect to find little generalization of policies extracted in an unrepresentative condition to policies applied outside the laboratory.

Although Brunswik (1956) laid out the theoretical rationale for representative design, he did not have available any findings of a comparative study of systematic and representative designs. Perhaps in practice there is little or no significant difference in the findings obtained by studies using either design. To our knowledge there has not been any comprehensive review and systematic comparison of the effects of systematic and representative designs in policy-capturing research. One simple explanation is that neo-Brunswikian researchers have often fallen short of Brunswik's vision for the application of his method of representative design; consequently, there are only very few representatively designed studies that could be compared to those using a systematic design. We attempted to deal with this problem in multiple ways; unfortunately, several approaches we initially took were unsuccessful.

First, using the 143 studies reviewed in Part 2, we tried to identify significant differences in the findings of those studies that were representatively designed and those that were not. For instance, on average, were participants in the former group of studies more consistent, did they use fewer cues, and did they demonstrate greater insight into their policies? However, there were too few studies for meaningful analysis after controlling for variations among studies in terms of, for example, the experience of participants, the dependent measure, and the level of analysis (i.e., capturing aggregate policies or individual policies). Second, we attempted to assess the validity of the findings of studies using unrepresentatively designed cases by examining whether models of the captured policies were successfully cross-validated: Do they correctly predict individuals' responses to a set of real cases? This attempt, however, also was unsuccessful, because although cross-validation is often recommended (e.g., Cooksey, 1996), it is rarely practiced.

In light of these unsuccessful approaches, we chose two alternative ways of addressing the question of whether representative design matters. First, we reviewed the results obtained by a small set of studies that compared judgment policies captured using representative and unrepresentative cases. Second, we relied on a set of studies that compared the calibration of confidence judgments and hindsight bias under representative and unrepresentative stimulus sampling conditions. We now describe the findings of these reviews.

Do Judgment Policies Alter as a Function of Design?

Possibly the most rigorous test of the effect of representative design is a within-subject comparison of policies captured for individuals under both representative and unrepresentative conditions. We are aware of only two published studies that have reported such a rigorous test. Phelps and Shanteau (1978) analyzed livestock experts' judgments of the breeding quality of pigs (see Table 5). They found that the judgment policies differed when experts were presented with unrepresentatively designed pigs and representative pigs: Experts seemed to have used significantly more cues in the unrepresentative condition than in the representative condition. Moore and Holbrook (1990) captured the car purchasing policies of MBA students (see Table 5) and found a significant difference between the weights attached to two cues in individuals' policies captured using representative and unrepresentative stimuli.

We also found seven other studies that compared policies captured under representative and unrepresentative conditions using a less rigorous procedure. The methods used, and the results obtained in these studies, are summarized in Table 5. Overall, the findings are mixed: Whereas some researchers found that there are differences in the policies captured using representative and unrepresentative stimuli (Ebbesen & Konecni, 1975, 1980; Ebbesen, Parker, & Konecni, 1977; Hammond & Stewart, 1974), others concluded that there is no difference (Braspenning & Sergeant, 1994; Olson, Dell'omo, & Jarley, 1992; Oppewal & Timmermans, 1999). When differences were observed, they occurred across a range of variables, including the linearity of the policy (Hammond & Stewart, 1974), the number of significant cues (Ebbesen & Konecni, 1975; Phelps & Shanteau, 1978), and the weighting and combination of cues (Ebbesen et al., 1977).

Unfortunately, the fact that the studies vary in terms of their methods (e.g., between-subjects vs. within-subject design), procedures, and analyses makes it difficult to compare their findings. In addition, some of the results are not as clear cut as the authors' conclusions. For instance, Oppewal and Timmermans (1999) based their conclusion of "no differences" on the fact that the same cue was accorded the greatest weight in both conditions, yet there were only 4 statistically significant cues in the unrepresentative condition compared with 10 in the representative condition. At the same time, however, differences should not be overweighted. For instance, although Moore and Holbrook (1990) found significant differences in the cue weights, these differences did not affect the predictive power of the two captured policies as both were equally good at predicting responses on a set of holdout cases.

The results from this small set of studies do not enable us to draw any definite conclusions regarding the effects of representative design on research findings. They do, however, point to the necessity of conducting more studies that directly compare policies captured for individuals under both representative and unrepresentative conditions (e.g., Dhami, 2004). Fortunately, we do not need to end with the clichéd call for more empirical work, because there is research that affords us further opportunity to examine the effects of representative design. Although research on the overconfidence effect and hindsight bias emanated outside of the neo-Brunswikian tradition, some researchers have recently drawn on Brunswik's notion of representative design for theoretical inspiration and demonstrated the effects of representative stimulus sampling.

Does "Overconfidence" Disappear in Representative Designs?

In Brunswik's (1943) view, psychology should aim to investigate organisms' adjustment to the inherently uncertain environments in which they function. Adjustment is considered in terms of an organism's use of proximal cues to achievement of a distal variable. The core premise of representative design is that the informational properties of the experimental task presented to participants represent the properties of the ecology to which experimenters wish to generalize. Although many contemporary psychologists may not necessarily share Brunswik's view of psychology as a science of organism–environment relations, they are interested in describing people's cognitive and behavioral achievements in an uncertain world. For illustration, consider the influential heuristics-and-biases research program (e.g., Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1996). This research has demonstrated a large collection of departures of human reasoning from classic norms of rationality, including phenomena such as base rate neglect, insensitivity to sample size, misconceptions of chance, illusory correlations, overconfidence, and hindsight bias. These phenomena have been described as *cognitive illusions* (Kahneman & Tversky, 1996), and these demonstrations of irrationality are explained in terms of heuristics on which people, equipped with limited cognitive resources, need to rely when making inferences about an uncertain world (Kahneman et al., 1982). In light of these cognitive illusions, many researchers have arrived at a bleak assessment of human reasoning—it is "ludicrous," "indefensible," and "self-defeating" (see Krueger & Funder, in press). However, concerns about whether studies demonstrating irrationality preserved an isomorphism between environmental and experimental properties have given rise to a Brunswikian perspective, initially in research on the overconfidence effect.

The overconfidence effect is prominent among the cognitive illusions catalogued by the heuristics-and-biases program, and it has received considerable research attention (for reviews, see Alba & Hutchinson, 2000; Hoffrage, in press; Lichtenstein, Fischhoff, & Phillips, 1982). Overconfidence research is essentially concerned with achievement, measured in terms of the extent to which people are calibrated to the accuracy of their knowledge. Studies demonstrating the overconfidence effect typically present participants with a general knowledge question of the following kind: Which city has more inhabitants, Atlanta or Baltimore? The participant is asked to choose one of the two options and then indicate his or her confidence that the chosen option is correct. This task requires participants to determine which of two objects scores higher on a criterion, and so it is a special case of the more general problem of estimating which subclass of a class of objects has the highest value on a criterion. Examples of such tasks are treatment allocation (e.g., which of two patients to treat first in the emergency room, with life expectancy after treatment as a criterion), and financial investment (e.g., which of two securities to buy, with profit as a criterion). The classic finding of overconfidence research is that out of the questions in which people say they are 100% confident, only about 80% are accurately answered. Out of the cases in which people are 90% confident, the proportion correct is about 75%, and so on. Quantitatively, the overconfidence effect is defined as the difference between the mean confidence rating and the mean percentage correct. This discrepancy has been interpreted as an error in reasoning and, like many other cognitive illusions, it is considered difficult to avoid.

However, this conclusion was challenged in the early 1990s by researchers advocating a Brunswikian theory of confidence (Gigerenzer, Hoffrage, & Kleinbölting, 1991). Gigerenzer et al. (1991) emphasized the importance of how questions had been sampled by experimenters. According to their "probabilistic mental models" theory, people solve questions such as which city is more populous by generating a mental model that contains probabilistic cues and their validities, akin to the process depicted in Brunswik's (1952) lens model. Therefore, in the choice between Atlanta and Baltimore, an individual may retrieve the fact that Atlanta has one of among the world's 50 busiest airports and that Baltimore does not, and that cities with such a busy airport tend to be more populous than those without. Consequently, the individual may conclude that Atlanta has a greater population than Baltimore and, according to probabilistic mental models theory, would report the validity of the cue as his or her subjective confidence that this choice is correct.

The manner in which questions are sampled is relevant for the demonstration of the overconfidence effect. For instance, assume that an individual has only one cue (e.g., airport cue) with which to determine the population size of U.S. cities. Among the 50 largest cities in the United States, the airport cue has an ecological validity of .6 (see Soll, 1996), where the ecological validity of a particular cue is defined as the percentage of correct choices using this cue alone. If the participant's subjective validity approximates this ecological validity, as suggested by Brunswik's (1943, 1952) assumption that people are well adjusted to their environment and empirically supported by a rich literature demonstrating that people automatically pick up frequency information in the environment (e.g., Hasher & Zacks, 1979), then the judge will be appropriately calibrated. That is, for the confidence category 60%, the relative frequency of correct choices should be 60%. This, however, holds true only if the experimenter selects questions in a representative way (randomly) from the reference class of the 50 largest cities in the United States. Alternatively, the experimenter could systematically sample more city pairs in which the airport cue would lead to a wrong choice (it so happens that Baltimore is more populous than Atlanta) or to a correct choice. The former sampling method would produce overconfidence, and the latter would yield underconfidence.

Therefore, Gigerenzer et al. (1991) suggested that the overconfidence effect may stem from the fact that researchers did not sample general-knowledge questions randomly but tended to overrepresent items in which cue-based inferences would lead to wrong choices. If so, then overconfidence does not reflect fallible reasoning processes but is an artifact of how the experimenter sampled the stimuli and ultimately misrepresented the cue–criterion relations in the ecology. Figure 3 shows the results of Gigerenzer et al.'s (1991) Study 1. Indeed, consistent with this interpretation, people were well calibrated when questions included randomly sampled items from a defined reference class (here, German cities). Percentage correct and mean confidence were 71.7 and 70.8, respectively. Figure 3 shows that Gigerenzer et al. (1991) also replicated the well-known overconfidence effect when people were presented with a nonrandomly selected set of items. Here, the percentage correct was 52.9, mean confidence was 66.7, and overconfidence was 13.8. The same result has been independently predicted and replicated by Peter Juslin and his colleagues (e.g., P. Juslin, 1993, 1994; P. Juslin & Olsson, 1997; P. Juslin, Olsson, & Björkman, 1997).

Following the publication of the probabilistic mental models theory, many researchers have examined whether unrepresentative sampling accounts for the overconfidence effect. Therefore, the impact of representative design can now be studied across a wide range of studies. Recently, P. Juslin, Winman, and Olsson (2000) conducted a review of 130 overconfidence data sets to quantify the effects of representative and unrepresentative item sampling. Figure 4 depicts the over- and underconfidence scores (regressed onto mean confidence) observed in those studies. Consistent with Gigerenzer et al.'s (1991) argument, the overconfidence effect was, on average, pronounced with selected item samples and close to zero with representative item samples. The results hold even when controlling for item difficulty—a variable to which the disappearance of overconfidence in the initial Gigerenzer et al. (1991) studies has sometimes been attributed.

Does Hindsight Bias Disappear in Representative Designs?

The impact of representative stimulus sampling on the existence of cognitive illusions is not restricted to probabilistic reasoning but also extends to memory illusions such as the *hindsight bias*: the tendency to falsely believe, after the fact, that one would have correctly predicted the outcome of an event. In an early study, Fischhoff and Beyth (1975) asked a group of students to judge a variety of possible outcomes of President Nixon's visits to Peking and Moscow before they occurred in 1972. Participants rated their confidence in the truth of assertions such as "The United States will establish a permanent diplomatic mission in Peking, but not grant diplomatic recognition" before the visits. After the visits, participants were asked to recall their original confidence. They exhibited hindsight bias: Participants' recalled confidence for events they thought had happened was higher than their original confidence, whereas their recalled confidence for events they thought had not happened was lower. These findings were supported by later research (for reviews, see Hawkins & Hastie, 1990, and Hoffrage & Pohl, 2003).

To examine whether representative sampling can reduce hindsight bias, Winman (1997) presented participants with general-knowledge questions of the kind used in overconfidence research. For instance, participants read questions such as: "Which of these two countries has a higher mean life expectancy, Egypt or Bulgaria?" Then, they were told the correct answer (Bulgaria) and were asked to identify the option they would have chosen had they not been told the correct answer. Akin to research on the overconfidence effect, Winman (1997) constructed selected sets of items and representative sets in which the countries that are included in the paired comparison task were drawn randomly from a specified reference class. The differences were striking. In Experiment 1, 42% of items in

Table 5
Results of Studies Testing the Validity of Judgment Policies Captured Under Representative and Unrepresentative Conditions

Study	Method	Results
Hammond & Stewart (1974)	Thirty-two students were divided into four equal groups. All were trained to infer the amount of foreign capital invested in hypothetical underdeveloped nations on the basis of two cues. (We report the results of the two relevant groups.) In the unrepresentative condition, participants were each presented with three blocks of 25 cases comprising a factorial combination of two cues. In the representative condition, participants were presented with 75 cases of the sort with which they had learned to deal.	Individual participants' policies were captured using multiple linear regression analysis. Five of the 8 participants in the unrepresentative condition had nonlinear policies, whereas all participants in the representative condition had linear, additive policies.
Ebbesen & Konecni (1975)	In the unrepresentative experiment, 18 San Diego court judges were each presented with 8 different cases. These were selected from a set of 36 cases comprising a factorial combination of four cues. Participants decided the amount of bail to be set on the cases. In the representative experiment, observations were made of the bail set on 105 cases by 5 of the judges, and details of the cases were recorded.	An analysis of variance was performed on the responses of all participants in the unrepresentative experiment, and multiple linear regression analysis was used to capture the policies of judges in the representative experiment. Three cues were significant in the former model compared to one in the latter.
Ebbesen et al. (1977)	In the unrepresentative condition, 20 students with driving licenses were presented 20 hypothetical cases comprising a factorial combination of two cues (they did practice and replications of whole design). Participants first decided whether it was safe for a car to cross an intersection. Then, participants judged the risk of crossing on a 200-mm scale anchored at each end. In the representative condition, 2,350 observations were made of 2,058 drivers who made left turns at four crossings in a locale.	An analysis of variance was performed on the data in each condition separately. Although some of the results converged, there was evidence for differences in how the cues were weighted and combined.
Phelps & Shanteau (1978)	In the unrepresentative condition, each of seven livestock experts were presented with 64 hypothetical paper cases that comprised a fractional factorial combination of 11 cues. Experts rated the breeding quality of these pigs. Two months later, in the representative condition, the same participants were asked to provide ratings on the overall quality of 8 pigs in photographs and provide ratings on the 11 cues.	Analyses of variance were performed on each individual's responses in the unrepresentative condition and a stepwise multiple regression was computed for each individual in the representative condition. In the unrepresentative condition experts seemed to have used more cues (i.e., 9–11) than in the representative condition (i.e., fewer than 3) as suggested by the number of statistically significant cues in the captured policies.
Ebbesen & Konecni (as cited in Ebbesen & Konecni, 1980)	In the unrepresentative condition, court judges were presented with 72 hypothetical cases comprising a factorial combination of five cues. Participants sentenced convicted defendants either to different types of penal establishments or to probation. In the representative condition, document analyses and observations were conducted, and the cues were recorded for 3,000 cases (although the results presented reflect the 800 cases analyzed at that point).	An aggregate policy was captured for participants in each condition using analysis of variance. The cue weights differed in both conditions, with more interactions in the representative condition and more cues used in the unrepresentative condition.
Moore & Holbrook (1990)	These authors reported three studies involving students who were each asked to rate the likelihood of purchasing cars on a 0-to-10 scale. In the first study, 62 participants were asked to make ratings on two sets of cases after a 1-month interval. The first set (unrepresentative condition) included 18 hypothetical cases comprising an orthogonal combination of five cues. The other set (representative condition) included these same cases but were altered to have plausible intercue correlations (which, however, were lower than found in the environment). Participants also completed a set of 17 holdout cases used for cross-validation, which included 9 orthogonal and 8 representative cases comprising three cues.	For each individual, a multiple regression analysis was computed on their responses to the orthogonal and representative sets of cases separately. Although both models were equally good at predicting responses on the holdout set, there was a significant difference between the policies across participants in terms of the weights attached to two of the cues. The other two studies replicated these results.
Olson et al. (1992)	In the unrepresentative condition, 19 arbitrators were each presented 32 hypothetical cases of police disputes comprising a fractional factorial combination of nine cues. Participants made decisions on whether to award wage increases on the basis of the union's offer. In the representative condition, the same participants had made decisions to award wage increases on 208 real teacher disputes, and these records were available for analysis.	An aggregate policy was captured for participants in each condition using a standard probit model. Two of the four comparable cues were significant in both conditions. The cues were used in the same direction and had the same order of relative weights.

Table 5 (*continued*)

Study	Method	Results
Braspenning & Sergeant (1994)	In the unrepresentative condition, 28 general practitioners were each presented 50 hypothetical cases comprising a random variation of the levels of eight cues. Participants rated on 9-point scales the degree to which patients had a mental health problem and the degree to which they showed a somatic condition. In the representative condition, 90 videotaped doctor–patient consultations were coded, and data were collected on six cues comparable with those used in the unrepresentative condition. Doctors provided ratings on one scale comparable to those used in the unrepresentative condition.	The two ratings made in the unrepresentative condition were moderately correlated and so were collapsed into one that was on a scale comparable to that used in the representative condition. Then, an aggregate policy was captured for participants in the unrepresentative condition using multiple linear regression analysis and the six cues that were comparable to those in the representative condition. An aggregate policy was captured for participants in the representative condition also using regression analysis. There was no significant difference between the weights attached to the cues in both conditions.
Oppewal & Timmermans (1999)	In a between-subjects design, 396 members of the general public evaluated the public space appearance in shopping centers on a 9-point scale. In the unrepresentative condition, 214 participants were each presented with 1 of 16 sets of 16 hypothetical cases. Altogether, 128 different hypothetical cases comprising a fractional factorial combination of 10 cues plus selected interactions were used. (Varying information on 3 other cues was added.) In the representative condition, 214 participants rated 29 existing shopping centers they visited frequently on 13 cues. In the validation condition, 182 participants completed the same task as in the representative condition.	An aggregate policy was captured for participants in each condition using multiple linear regression analysis. The same cue was most important in both conditions, but there were only 4 significant cues in the unrepresentative condition, compared with 10 in the representative condition. The models from the unrepresentative and representative conditions were equally good at predicting the responses of the new sample of participants.

the selected set produced hindsight bias, whereas in the representative set of items only 29% did so. A within-subject analysis revealed that for the representative sample only 3 out of 20 participants had a higher accuracy in hindsight than in foresight (indicating "I knew it all along"), whereas for the selected sample this was the case for 14 out of 20 participants.

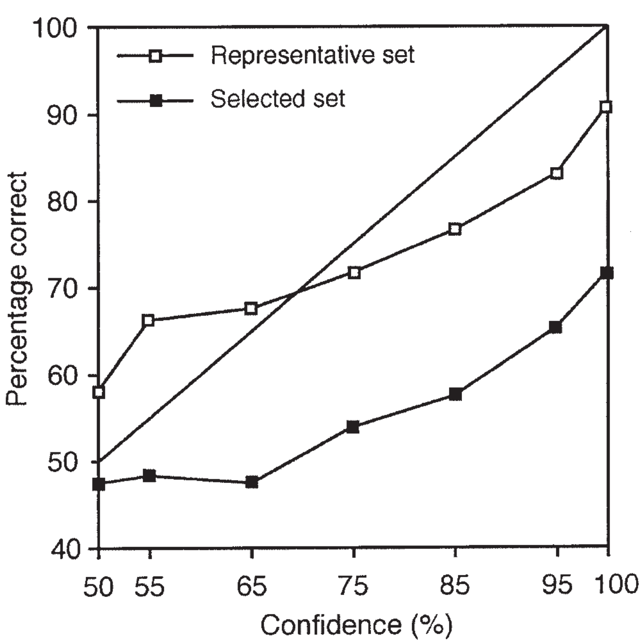

Figure 3. Calibration curves for the systematically selected (black squares) and representative sets (open squares). From "Probabilistic Mental Models: A Brunswikian Theory of Confidence," by G. Gigerenzer, U. Hoffrage, and H. Kleinbölting, 1991, *Psychological Review, 98,* p. 514. Copyright 1991 by the American Psychological Association. Adapted with permission.

In his most comprehensive attempt to study the impact of representative sampling, Winman (1997, Experiment 3) used large independent samples of up to 300 items per participant. In addition, he compared the results of a within- and between-subjects design. In the between-subjects design he obtained slight overes-

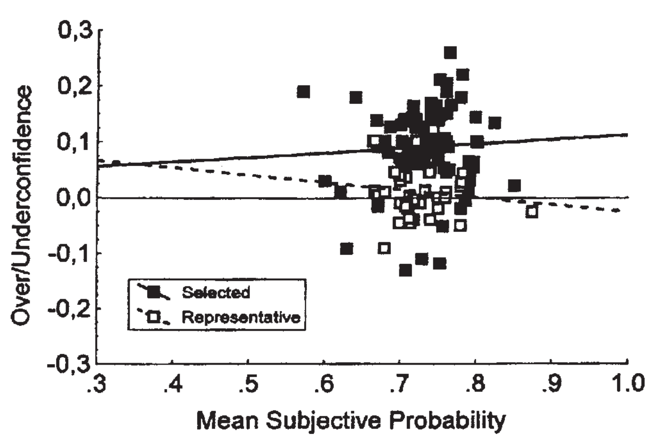

Figure 4. Regression lines relating over- and underconfidence scores to the mean subjective probability for systematically selected (black squares) and representative samples (open squares). From "Naive Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard–Easy Effect," by P. Juslin, A. Winman, and H. Olsson, 2000, *Psychological Review, 107,* p. 391. Copyright 2000 by the American Psychological Association. Reprinted with permission.

timations of the frequency with which participants believed they would have used extreme response categories—an effect that Winman attributed to the lack of familiarity with the confidence scale and not a reflection of the hindsight bias. Consistent with this view, in the within-subject design the hindsight bias completely disappeared when items were sampled representatively (here participants had the opportunity to use the confidence scale in foresight). To further appreciate that representative sampling eliminates hindsight bias, it is worth noting that the hindsight bias has resisted most debiasing techniques (Fischhoff, 1982) and that it appears most pronounced in the domain of general-knowledge items (Christensen-Szalanski & Willham, 1991).

Discussion

The key imperative of representative design is to represent the situation toward which the researcher intends to generalize. If one does not, the danger is that the processes studied are altered in such a way that the obtained results are no longer representative of people's actual functioning in their ecology. Whereas research comparing the effects of representative and systematic design on judgment policy capturing has yielded mixed results, our review of studies on the effects of representative item sampling on the overconfidence effect and hindsight bias demonstrates that design does matter (for an example of how representative item sampling influences the availability effect, see Sedlmeier, Hertwig, & Gigerenzer, 1998).

In a presidential address to the Society for Judgment and Decision Making, Barbara Mellers observed that decades of research on judgment and decision making have created a "lopsided view of human competence" (as cited in Mellers, Schwartz, & Cook, 1998, p. 3). Indeed, research in the heuristics-and-biases program involves carefully setting up conditions that produce cognitive biases (Lopes, 1991). The extent to which these findings generalize to conditions outside the laboratory is unclear. Lopes (1991) argued that this research depicts heuristic processing as negative and people as cognitively lazy and has led to a pessimistic view of human abilities. Imagine what our conception of human judgment would be if researchers had aimed to examine individual achievement and environmental adaptation rather than judgmental errors and maladaptiveness and if they had also used representative rather than systematic stimulus sampling. On the basis of our review of studies in Part 3, we can speculate that conclusions about human competence and rationality might have been quite different (see also Gigerenzer, 1991, 1996; Hammond, 1990) and, in fact, much more optimistic.

Finally, to the extent that representatively designed studies can yield information on how people function in their environments, research findings may be used to improve performance by intervening on the cognitive side, the environment side, or both. As mentioned at the outset of this article, representatively designed studies have had a significant impact on public policy and thus people's lives. For example, Hammond and his colleagues have successfully used Brunswik's method and theory to examine issues such as managing air quality, designing safe highways, choosing police handgun ammunition, and improving clinical judgment (for reviews, see Hammond, 1996a; Hammond & Stewart, 2001a; and Hammond & Wascoe, 1980).

Part 4: General Discussion

Brunswik stressed that psychological processes are adapted to the environments in which they have evolved and in which they function. He argued that psychology's accepted methodological paradigm of systematic design was incapable of fully examining the processes of vicarious functioning and achievement. As an alternative, he proposed the method of representative design. What is representative design, and how can it be done? In Part 1, we outlined Brunswik's (1955c, 1956) proposal that representative design could ideally be achieved by random sampling of real stimuli from the participant's environment. We further showed how Hammond later revised the concept of representative design by distinguishing between *substantive situational sampling* (akin to Brunswik's, 1955c, 1956, original formulation) and *formal situational sampling* (in which stimuli are created to preserve the formal informational properties of the environment).

Has representative design been used? In Part 2, we examined the research practices of individuals who have been committed to the notion of representative design. Two major findings emerged from our review of neo-Brunswikian policy-capturing research. First, most of the studies that presented participants with real cases satisfied Brunswik's (1955c, 1956) recommendation of probability or nonprobability sampling of stimuli. Second, there was a striking discrepancy between Hammond's formula for achieving formal situational sampling and the research practices of most neo-Brunswikian studies that presented participants with hypothetical cases. Neo-Brunswikians using formal situational sampling often failed to represent important aspects of the ecology toward which their generalizations were intended.

Does design even matter? In Part 3, we discussed whether representative sampling matters for the results obtained. Unfortunately, only a small body of research has compared judgment policies captured under representative and unrepresentative conditions, and their results are mixed. Whereas some studies reported that representative conditions affected judgment policies—for instance, in terms of cue weights—others concluded that captured policies were independent of the representativeness of the stimuli. To date, the strongest evidence for the effect of representative stimulus sampling stems from research on the overconfidence effect and on hindsight bias. With regard to the former, a recent review of studies that manipulated the sampling procedure of experimental stimuli demonstrated that representative item sampling reduces—in fact, almost eliminates—the overconfidence effect.

We conclude this article first with a discussion of the impact that Brunswik's (1955c, 1956) notion of representative design has had in fields beyond judgment and decision making. Second, we argue that representative design is seldom used not only because of the practical difficulties involved but also because it is so rarely taught. Third, we reveal that there are various practical and theoretical reasons why the ideal of representative design may be more attainable at the beginning of the 21st century than when it was originally proposed. Fourth, we suggest that representative and systematic designs may each be useful for different purposes. Finally, we discuss other ecological approaches to cognition.

Representative Stimulus Sampling in Brunswik's Field of Study and Beyond

The issue of how stimulus sampling affects empirical findings is pertinent to all areas of experimental psychology. We have already reviewed stimulus sampling in judgment and decision-making research. In this section, we discuss the role of stimulus sampling in the domains of visual perception, social perception, and language processing. Our intention is to provide a snapshot of the developments in these areas with reference to individuals such as James J. Gibson, David Funder, and Herbert Clark.

Perception. Whereas neo-Brunswikians have been mostly concerned with human judgment and decision making, much of Brunswik's empirical work was conducted in the field of perception (Brunswik, 1934, 1956). Like Brunswik, Gibson (1957/2001) recognized that the method of systematic design was unsuitable for the study of perception because of the multiplicity of variables and the complexity of relations among variables. Gibson (1979) developed an ecological approach to the study of perception in which he emphasized the need to study people performing meaningful tasks in their natural environments rather than responding to artificial (two-dimensional) stimuli in the laboratory. Like Brunswik, Gibson believed that research should be focused on perceptual achievement. Gibson's approach differs from Brunswik's in that, for example, whereas Brunswik saw limitations to the information for achievement provided by environmental cues, Gibson believed the environment afforded complete, valid information that enabled direct, veridical perception without intervening mediation (for a comparison between the two approaches, see Cooksey, 2001, and Kirlik, 2001). Gibson's approach has received considerable attention and has demonstrated that a serious analysis of perception requires a simultaneous analysis of the environment and the information that it affords (see Reed, 1988). However, a search in PsycINFO revealed that there is very little exchange between neo-Brunswikian researchers and perception researchers such as those pursuing the tradition of Gibson's theory of "direct perception" (Kirlik, 2001).

This does not mean that the notion of representative stimulus sampling is unimportant and unknown in perception research. Indeed, it is clear that perceptual achievement is codetermined by how stimuli are selected. For example, in color constancy, even under poor yet natural light conditions, such as before sunrise or after dawn, a red car will be identified as a red car; that is, under natural conditions, color constancy is preserved. However, under conditions unrepresentative of those in which our perceptual system has evolved, such as light emanating from sodium or mercury vapor lamps, color constancy breaks down (Shepard, 1992). (Note that in this example *representative sampling* refers more to the sampling of lighting conditions and less to the sampling of objects as, for instance, in studies of size constancy.) Therefore, we suggest that although Brunswik's work is rarely cited in perception research, the gist of his methodological innovation—namely, the representative sampling of naturally occurring objects and conditions—is understood and embraced across a wide range of perception studies. These include pattern recognition, color constancy, "ecological optics," visual perception (Field, 1987), and perception of sounds (Ballas, 1993).

To the best of our knowledge, only a few perception researchers have systematically compared the effects of representative and selected samples of stimuli. Among these is James Cutting (e.g., Cutting, Bruno, Brady, & Moore, 1992; Cutting, Wang, Flückiger, & Baumberger, 1999). His results suggest, however, that there is little or no difference between these two sets of stimuli. In an effort to reconcile the findings of Cutting and his colleagues with those that suggest the opposite (see Part 3), we refer to differences in the perceptual and reasoning processes studied. People's performance in perception experiments may indeed be less dependent on the selection of stimuli than is their performance in reasoning experiments. When applying his lens model framework to perceptual and cognitive judgment tasks, Brunswik (1955b) highlighted differences between the two systems. In his view, perception is predominantly "probability geared," whereas thinking is predominantly "certainty geared" (Brunswik, 1955b; this distinction is also central to Hammond's, 1996a, *cognitive continuum theory*; see also Goldstein & Wright, 2001). In other words, perception is characterized "by a relative paucity of on-the-dot precise responses counterbalanced by a relatively organic and compact distribution of errors free of gross absurdity" (Brunswik, 1955b, p. 109). Thinking, by contrast, "proves highly vulnerable in practice; inadvertent task-substitutions and other derailment-type errors often reaching bizarre proportions belie the relatively large number of absolutely precise responses" (Brunswik, 1955b, p. 109). Thus, if Brunswik was correct about the different distributions of errors—many, but small, for perception and few, but large, for reasoning—this may explain the (possibly) differential impact of representative design in research on perception and reasoning.

Social perception. In his studies on social perception, Brunswik advocated the importance of measuring achievement and presenting participants with a representative sample of *person-objects* (a phrase referring to the person to be judged; Brunswik, 1945, 1956; Brunswik & Reiter, 1937). Indeed, one would be using a double standard to not representatively sample the persons to be judged in social perception experiments even though one samples the persons doing the judging. Inspired by Brunswik's ideas, Funder (1995) criticized social perception researchers' emphasis on perceptual errors and use of artificial, arbitrary, and systematically designed stimuli that are presented to participants in the laboratory. He urged researchers to study accuracy in social perception and to use representative sampling of person-objects. He argued that personality traits can be considered as distal stimuli that emit multiple, interrelated, probabilistic cues, for example, via behavior or physical appearance. Funder's (1995) *realistic accuracy model* describes the accuracy of personality judgment as a function of the availability, detection, and utilization of behavioral cues to personality. The personality judgment may be used to predict how a person-object behaves in the future. To study the accuracy of social perception, videotape technology is first used to capture natural human interaction patterns (for a review, see Funder, 1999). These are then presented to participants who may or may not already be acquainted with the person-objects. Finally, participants' accuracy in judging traits is measured with reference to various criteria, including, for example, the results of personality tests completed by the person-object or ratings (including self-ratings) of the person-object.

Although some other researchers have used representative design in studying the perception of personality (e.g., Gifford, 1994) and the perception of rapport (e.g., Bernieri, Gillis, Davis, & Grahe, 1996), social psychologists have generally paid little attention to representative design. In a review of social psychological experimentation, Wells and Windschitl (1999) highlighted a "serious problem that plagues a surprising number of experiments" (p. 1115), namely, the neglect of "stimulus sampling." In their view, stimulus sampling is imperative whenever individual instances within categories (e.g., gender and race) vary from one another in ways that affect the dependent variable. For example, relying on only one or two male confederates to test the hypothesis that men are more courteous to women than to men "can confound the unique characteristics of the selected stimulus with the category," and "what might be portrayed as a category effect could in fact be due to the unique characteristics of the stimulus selected to represent that category" (Wells & Windschitl, 1999, p. 1116). In other words, it is a flagrant disregard of sampling theory when researchers report findings about the ability of people in general to make accurate judgments about other persons despite sampling only one or a few other persons. It is interesting that in an early publication Hammond (1954) described how a reputable psychologist conducted a prominent study using a large sample of subjects and a sample of only two person-objects. Almost half a century later, Wells and Windschitl demonstrated that the same double standard persists.

Language. Clark (1973) published a controversial article that criticized language and (semantic) memory researchers' practice of generalizing their findings beyond the specific sample of language they studied to language in general. For illustration, imagine a sentence verification experiment in which participants are presented with word pairs in sentences such as "A daisy is a flower" or "A daisy is a plant." In actual experiments, such word pairs were constructed from word triplets that included a word and its two most frequently elicited superordinates (for which one was the superordinate of the other as well), such as *daisy–flower–plant*. The triplets can be sampled randomly, and so for half of the triplets the second word (e.g., *flower*) would be more frequent than the third word (e.g., *plant*), and for the other half the relationship would be reversed, thus approximating actual occurrence. Clark pointed out that if researchers sampled only one kind of triplet, they would conclude that there were highly reliable differences in the verification response time for the pairs *daisy–plant* versus *daisy–flower* when in fact the differences are small and less reliable. Thus, researchers would have committed what Clark called the *language-as-fixed-effect fallacy*; that is, they would have treated linguistic objects as a fixed instead of a random effect and implicitly assumed that the objects chosen by investigators constitute the complete population to which they wish to generalize. Clark (1973) demonstrated this implicit assumption across a range of studies from the field of semantic memory, and he pointed out that researchers' neglect of sampling objects (e.g., letter strings, words, and sentences) has serious consequences: "The studies are shown to provide no reliable evidence for most of the main conclusions drawn from them" (p. 335).

Clark (1973) proposed different remedies, and he emphasized the use of appropriate statistics to enable generalizations from the stimulus sample to the population in question (see also Raaijmakers, Schrijnemakers, & Gremmen, 1999). Although Clark's (1973) analysis was not derived from the vantage point of Brunswik's ideas, in a manner akin to Brunswik's, he stressed that "if the investigator is to treat language as a random effect, then he must draw a sample at random from the language population he wishes to generalize to" (p. 350). To do so, he argued, the researcher has to define the language population, draw an unbiased sample from this population, and apply a sampling procedure that others can replicate.

Summary. Although we have demonstrated that leading researchers in several areas have recognized the importance of representatively sampling stimuli, it is also clear that others have not embraced this approach. For instance, Raaijmakers et al. (1999) stated that Clark's (1973) statistical solutions to the language-as-a-fixed-effect fallacy are inadequate. Cutting et al.'s (1992, 1999) research suggests that representative design may have little impact on perception research. Finally, as Wells and Windschitl (1999) pointed out, research on social perception continues to practice the double standard in sampling.

Representative Design: Not Yet Part of Our Collective Toolbox

Brunswik's contemporaries rejected his methodological critique and innovations. In Gillis and Schneider's (1966) view, this response was due to their epistemological conviction that the "road to truth" is in the variation of one or few variables while all other variables are held constant. Thus, only systematic design can produce valid knowledge. This conviction is as old as modern experimental psychology and was espoused by its founders. The concept of systematic design has evolved since those early years. Recognizing the limitations of the one-variable design, Sir Ronald Fisher (1925), for instance, offered factorial design and analysis of variance as alternatives. Whereas representative design was a methodological consequence of probabilistic functionalism, reflecting what Brunswik believed to be the aims of psychological research, factorial design was developed for the aims of agricultural research. Here, researchers were testing the effect of specific treatments (conditions) on plants and animals (participants), and the generalization of findings focused on participants, not conditions. Researchers could control the conditions, thus making their findings applicable by creating a new environment for participants. By contrast, psychologists studying perception or judgment and decision making, for instance, tend not to want to change their participants but to identify the processes that have emerged as a result of people's adaptations to their natural environments. Although factorial design provided relief from the straitjacket of the one-variable design, it eliminates the natural covariation among variables and artificially combines factor levels that may be impossible outside of the laboratory. From a Brunswikian perspective, Fisher's time-honored methodological approach, therefore, still risks altering the very phenomena under investigation.

We suspect that generations of psychologists have little or no familiarity with the concept of representative stimulus sampling. This is because they are taught using textbooks that pay little attention to Brunswik's methodology. We conducted an analysis of 43 textbooks on experimental design or research methods and searched for keywords such as *Brunswik*, *sampling*, *representative* *design*, *ecological validity*, and *external validity*.[7] The concept of external validity was explained in about half of them, but only 3 textbooks made reference to Brunswik; of these, 1 explained representative design. Whereas almost all of the 43 textbooks stressed the importance of representatively sampling participants so as not to impede generalizability to the subject population, only 7 also mentioned the importance of representatively sampling stimuli. Therefore, we expect that many current researchers do not know about the notion of representative design. Those few who are aware of representative design may not wish to transgress the norm of systematic design. Indeed, Hammond (1996b) frankly admitted that one reason for his neglect of representative design was that he feared that otherwise, like Brunswik, "I would become isolated and ostracized" (p. 245).

Nevertheless, there is a growing discomfort with the unrepresentativeness of laboratory studies in general, and with systematic design in particular, which is reflected in an increased attention to concepts such as external validity and ecological validity. A search on PsycINFO for the root *external* valid** shows a healthy growth in the number of entries from 1956, when Brunswik's *Perception and the Representative Design of Psychological Experiments* was published (i.e., 9 hits from 1956 to 1966), to the present day (i.e., 168 hits from 1999 to October 2001). Similarly, the root *ecolog* valid** yielded 2 hits from 1956 to 1966 and 141 hits from 1999 to October 2001. However, this development has a Janus face. As Hammond (1978, 1996b) has pointed out, terms such as *external validity* are inferior substitutes for representative design because they do not specify the practice of sampling stimuli from the participant's environment. Similarly, frequent use of the terms *ecological validity* and *representative design* interchangeably confuses the concepts of achievement and generalizability. Indeed, Brunswik (1956) recognized the possibilities of the erosion of his terminology. We hope the present article redresses such confusions.

Representative Design in the 21st Century

Since Brunswik (1956) acknowledged that representative design was a formidable task in practice, it faces the charge of infeasibility. Although the theoretical problems of sampling have not been resolved, Hammond (1966) and others proposed modifications (e.g., formal situational sampling) intended to make representative design a more workable tool. Yet it is still perceived as overly costly and time consuming. Vivid examples of how time consuming and cumbersome it was to implement (see Brunswik's, 1944, size-constancy experiment and Gifford's, 1994, interpersonal perception experiment) put representative design at a disadvantage in a methodological marketplace where fast, inexpensive, and convenient methods prevail (Hertwig & Ortmann, 2001). This however, may soon change.

Virtual environments. One reason is that modern technologies can help to re-create natural instances of a person's environment in the laboratory. Today, spatial cognition and perceptual control of action, for instance, can be explored in virtual environments that provide a great realism of the displayed images (e.g., Bülthoff & van Veen, 2001). Similarly, "micro-worlds" can be used to study decision making, problem solving, and cognition (e.g., B. Brehmer & Dörner, 1993; Fiedler, Walther, Freytag, & Plessner, 2002; Omodei & Wearing, 1995); learning (e.g., Kordaki & Potari, 1998); human–machine dynamics (e.g., Wastell, 1996); navigation strategies (e.g., Burns, 2000); and management (e.g., Hirsch & Immediato, 1999). These micro-worlds are dynamic computer simulations of real environments with which participants repeatedly interact in the laboratory. The construction of virtual environments or simulated micro-worlds, however, still requires an ecological analysis of the natural world. Nevertheless, they can enormously simplify the task of data collection by providing efficient and flexible tools with which to implement real-world contingencies among variables in a small-scale laboratory environment.

Computer simulations. Since Brunswik's time, the methodological toolbox of cognitive psychologists has become more diverse. Although experimentation remains important, it has been supplemented by other methods, such as computer simulation. We suggest that the issue of sampling is as important in the design of simulations as in the design of experiments. For example, there are process models available for both the overconfidence effect (e.g., Gigerenzer et al., 1991) and hindsight bias (e.g., Hoffrage, Hertwig, & Gigerenzer, 2000) that can be, and have been, implemented in computer programs (Hertwig, Fanselow, & Hoffrage, 2003; Pohl, Eisenhauer, & Hardt, 2003). The way in which samples are drawn from the ecology, and the formal properties of the ecology (i.e., number of cues, their values, distributions, intercorrelations, and ecological validities), can be manipulated, thus making it possible to replicate and explicate the impact of representative sampling on the overconfidence effect and hindsight bias. We believe that researchers who specify parameters and samples of their simulated ecology benefit theoretically from representative sampling, and vice versa, that representative design may prove its feasibility and utility by being used in the context of simulations.

Computer records of environments. According to Gibson (1957/2000), to Brunswik,

> *the experimenter has little or no basis for knowing in advance whether his experiment is representative or unrepresentative*. It might seem to him lifelike, yet prove to be not so. The experimenter's only policy is to keep on experimenting so as to sample the world adequately. He must operate, if not in darkness, at least in theoretical twilight. (p. 246)

We suggest that with the availability of large databases that capture aspects of people's environment, the twilight has become less dusky. For instance, J. R. Anderson and Schooler (1991) were interested in studying the statistics of the information-retrieval demands that people face in their environments. To gather such statistics required detailed records of people's experiences in the world. As J. R. Anderson and Schooler (2000) pointed out, "ideally, researchers would follow people around, tallying their informational needs" (p. 560)—as in Brunswik's (1944) size-constancy study. Rather than burdening themselves with this task, they made

[7] We searched through the library catalogues of the Max Planck Institute for Human Development in Berlin and the University of Mannheim for books that (a) had either *research methods* or *experimental design* in the title, descriptor field, or keyword field; (b) covered psychology or consumer research; (c) were published in the first edition after 1955, when Brunswik's central article on representative design was published; (d) had at least a second edition; and (e) were written in English. The list of 43 textbooks that fulfilled these criteria that we analyzed can be obtained from Mandeep K. Dhami.

use of human-communication databases (i.e., transcripts of 25 hr of children's speech interactions, 2 years of *New York Times* front-page headlines, and 3 years of J. R. Anderson's e-mail messages) that were available as computer records and so could be subjected to computer-based analyses. Therefore, with the availability of large databases that capture aspects of people's ecologies, experimenters can analyze informational structures in the environment and assess the extent to which their experiments succeed in re-creating these structural properties.

The fast and frugal lens. Finally, representative design may be more feasible today than it was previously because of advances in the tools used to capture judgment processes. As we noted in Part 1, neo-Brunswikians have modeled vicarious functioning—the process of substituting correlated cues—using multiple linear regression. To meet the requirements of this statistical tool, researchers tend to use fewer cues than may be relevant, and they orthogonally manipulate cues. Multiple regression relies on two fundamental processes, namely, weighting of cues and the summing of cue values (Kurz & Martignon, 1999). According to Gigerenzer and Kurz (2001), weighting and summing capture only part of vicarious functioning at the expense of two cognitive processes that are ignored by unweighted or weighted linear models such as multiple regression, namely, the order in which cues are searched and the point at which search is stopped.

As an alternative to a lens function that is modeled in terms of multiple regression, Gigerenzer and Kurz (2001) proposed a fast and frugal lens (also see Dhami & Harries, 2001). The terms *fast* and *frugal* signify cognitive processes that allow one to make judgments in a limited amount of time, with limited information, and without having to optimize. A fast and frugal lens that relies on "one-reason decision making," for instance, would terminate the search for information as soon as the first cue has been found that renders a decision possible. Different heuristics will have different search, stop, and decision-making rules. For example, the take-the-best heuristic models two-alternative choice tasks (Gigerenzer & Goldstein, 1996). It orders and searches for cues according to their validities and bases a choice on the first cue that discriminates between two alternatives. Computer simulations have demonstrated that take-the-best sometimes outperforms the regression model in predicting environmental criteria (see Gigerenzer, Todd, & the ABC Research Group, 1999). Experimental studies have shown that take-the-best proves to be a better predictor of human judgment than linear models when an information search has to be performed under time pressure (Rieskamp & Hoffrage, 1999) and when information is costly (Bröder, 2000). Other fast and frugal heuristics, such as the matching heuristic, perform as equally well as the regression model (Dhami & Harries, 2001) and outperform other compensatory strategies (Dhami, 2003; Dhami & Ayton, 2001) in predicting professionals' judgments.

The notion of a fast and frugal lens promises to be key in making representative design more feasible for several reasons. First and foremost, it avoids the problems that a statistical tool such as multiple regression poses for users of representative design. Specifically, by assuming that cues are not integrated (or not integrated in a complex way), the fast and frugal lens allows researchers to study intercorrelated cues. Under this condition, multiple regression, by contrast, suffers from the problem of multicollinearity that may make it impossible to determine the cues actually used by the decision maker. Furthermore, in a fast and frugal lens model there is no need to adhere to a particular case-to-cue ratio, thus enabling examination of a larger number of cues. Finally, the problem faced by someone conducting a multiple regression of a reduced stimulus set when cases include missing cue values is overcome in a fast and frugal lens model by the fact that many of the heuristics have built-in rules for dealing with missing data (e.g., search the next cue). In fact, by precisely describing the processes of information search, stop, and decision making, researchers can now directly explore these psychological issues.

Representative design and response dimensions. Representative design has mostly been considered in terms of the sampling of experimental stimuli (conditions) and how well they represent the population of stimuli toward which generalization is intended. Little attention has been paid to how participants' responses measured in experiments are representative of their "natural" response dimensions outside of the laboratory. For illustration, let us return to research in the heuristics-and-biases program that has examined the extent to which people are able to reason in accordance with the rules of probability and statistics (e.g., Kahneman et al., 1982; Kahneman & Tversky, 1996). The dominant view was that the human mind is not built to work by the rules of probability. This was inferred from experimental demonstrations of *cognitive illusions*—systematic errors in people's probabilistic reasoning. Many of the tasks designed to demonstrate the reality of cognitive illusions, however, required participants to express their responses in terms of a mathematical probability. Mathematical probabilities are relatively recent representations of uncertainty; they were first devised in the 17th century (Gigerenzer et al., 1989), and for most of the time during which the human mind evolved, information was encountered in the form of natural frequencies, that is, counts of the occurrence of events (which were not normalized with respect to base rates). Thus, to the extent that the human mind has evolved to reason about uncertainty in terms of natural frequencies, asking participants to respond in terms of natural frequencies is more representative of the way people express uncertainty outside of the laboratory (Cosmides & Tooby, 1996). Consequently, their "achievement" in probabilistic reasoning tasks may depend on, among other factors, the format in which they are required to express their judgments about uncertainty (Gigerenzer & Hoffrage, 1995; for a debate on this issue, see Mellers, Hertwig, & Kahneman, 2001).

Therefore, the notion of representative design not only is applicable to the issue of sampling participants and experimental stimuli but also extends to the sampling of response formats (e.g., Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) and response modes (e.g., Billings & Scherer, 1988; Hertwig & Chase, 1998; Westenberg & Koele, 1992). It is clear that more research needs to examine the effect of representativeness of response dimensions on people's performance in cognitive tasks.

In sum, we suggest that, for various reasons, representative design has shifted from an unattainable ideal to a prudent goal. As a consequence, researchers can think afresh about when and why it can and ought to be used—a discussion that is no longer ailed by the sense of detached methodological idealism.

Representative and Systematic Design: Complementary Experimental Tools?

Brunswik pitted representative design against the existing experimental practices in psychology, thus perhaps contributing to the belief that representative and systematic design are antagonistic strategies. We disagree with this view. Rather, we propose that there is a division of labor between both experimental strategies, and we provide an illustration of the distinct yet complementary purposes that the two designs may serve. In a recent study, Rieskamp and Hoffrage (1999) aimed to discover which of eight compensatory (e.g., weighted pros) and noncompensatory (e.g., lexicographic) choice strategies best described people's choices under conditions of low and high time pressure. From groups of four publicly held companies, participants had to select the one with the highest yearly profit. The companies were sampled from a population of 70 real companies and were described in terms of six cues, such as share capital and number of employees.

First, Rieskamp and Hoffrage (1999) provided participants with 30 groups of four companies, each of which was randomly drawn from the reference class. In this representative condition, they found that under high time pressure, individuals searched for less information, searched for the most important cues, spent less time looking at information, and demonstrated a cue-wise information search pattern. Although these results indicated that participants were using one of the noncompensatory strategies, the results did not distinguish the lexicographic strategy (in which the authors were particularly interested) from the other strategies. Indeed, the strategies made the same predictions in most of the cases (the percentage of identical predictions averaged across all possible pairs of strategies was 92%). Consequently, the proportions of participants' choices the strategies correctly predicted fell in a small range (between 79% and 83%).

Next, to investigate which strategy each participant most likely used, Rieskamp and Hoffrage (1999) selected a choice set that distinguished between the strategies so that different strategies made different predictions. Participants were then provided with 30 systematically selected groups of companies. In this systematic condition, the percentage of identical predictions among strategies was, by design, reduced (i.e., to 50% averaged across all possible pairs of strategies). Consequently, the range of the proportions of participants' choices that the strategies correctly predicted was considerably greater (between 33% and 66%), rendering it possible to more clearly distinguish between the strategies participants used. Overall, participants tended to switch from weighted pros to lexicographic strategies under conditions of low and high time pressure, respectively.

Rieskamp and Hoffrage's (1999) study underscores the complementary nature of the two designs. If the researchers had solely sampled in an unrepresentative way, they would have severely underestimated the descriptive validity of the choice strategies: On average, the proportion of predicted choices dropped from 81% in the representative condition to 51% in the systematic condition. If, however, the researchers had sampled only in a representative way, they would not have been able to discriminate between the strategies that people used. The general issue is that a representative sample may contain only a few of those crucial items that allow differentiation of the candidate strategies, leading to low statistical power.[8]

Thus, Rieskamp and Hoffrage's (1999) study demonstrates that the decision of which design to use cannot be subjected to a dogmatic or mechanical rule but requires an informed judgment in light of the investigation's goals.[9] If the goal is to evaluate competing models, then research may involve stimuli that discriminate between these hypotheses, namely, systematically designed stimuli. For instance, the work of Norman H. Anderson (1981) involves examining how algebraic models describe information-integration processes (e.g., additive and multiplicative) rather than measuring achievement, and so he has used systematic design in his research on person perception. On the other hand, if the goal of a study is to understand how an organism functions and achieves in its environment, then representative design will be the method of choice. It is important to note that using a representative design will not necessarily lead to conclusions that people are perfectly adjusted to the informational properties of their environment or that people demonstrate high achievement. Researchers should not assume that perfect achievement can be observed in an uncertain world or that people always sample the world representatively (see Fiedler & Juslin, in press). Fiedler (2000) demonstrated that judgments are often remarkably accurate in light of the samples drawn, although people may draw inaccurate conclusions about the population of stimuli to the extent that they sample unrepresentatively.

A researcher's choice of experimental design can determine the outcome of an investigation. The solution to this dilemma should not be that all researchers use the same standardized design, thus circumventing the potential reality of conflicting results. Rather, we believe that one way to deal with the interaction between experimental designs and results is to be clear about the goal of the investigation. Referring to Roger Shepard's ideas and experiments on evolutionary internalized regularities and the lessons drawn from animal studies on circadian rhythms, Schwarz (2001) wrote that

> uncovering evolutionary constraints requires the use of abnormal experimental conditions . . . if a constraint does embody a regularity occurring in the environment, it will remain hidden in ordinary circumstances. . . . To discover constraints on circadian rhythms it

[8] To increase the power of the chosen design, some social judgment theorists have advocated *efficient designs,* which rely on extreme but plausible cases as a compromise (e.g., McClelland, 1997; McClelland & Judd, 1993).

[9] As another example, we can point to Reimer and Hoffrage's (2003) study of the *hidden-profile effect,* which emerges when the distribution of information among group members is such that an omniscient group member (with access to all information the individual members have) makes a decision that none of the individual members would make. Producing and testing such hidden-profile scenarios proves useful because it enables researchers to make different predictions for different information-exchange strategies (Stasser & Titus, 1985). However, as Reimer and Hoffrage (2003) discovered, these scenarios are rare. Using Monte Carlo studies, they found that a hidden-profile effect occurred in only 456 of 300,000 (0.15%) randomly generated scenarios. Thus, to the extent that the effect is helpful in distinguishing between different information-exchange strategies, sampling representatively from the 300,000 cases would be futile, and so researchers would benefit from systematic sampling. Finally, this example also illustrates that the ultimate decision of what design to use benefits from an analysis of the task environment.

> was necessary to remove the animals from their ordinary environment and place them in artificially created settings. (p. 627)

Another way to deal with the interaction between experimental designs and results is to adopt a multimethod approach in which the same phenomenon is investigated using more than one design. Here, systematic and representative designs are not the only choices; other designs include, for instance, Birnbaum's (1982) *systextual* design, which deals with the interaction of cognition and environment by systematically manipulating the latter (e.g., in terms of the intercorrelations among cues), thus ultimately rendering possible a comprehensive theory of context and how it affects human judgment.

Different routes to the generality of results. The systematic manipulation of our world in order to learn its secrets has frequently been hailed as the royal road to knowledge (others being, for instance, deduction from first principles). To learn the secrets, however, it is often necessary to bring some manageable fraction of the world into the laboratory under circumstances that render possible inferences about the underlying cause–effect relations. Brunswik (1955c) acknowledged that any systematic experiment represents at least one actual or potential ecological instance and in this sense is a fraction of reality. Yet the question remains whether the experimental results observed in this fraction of reality generalize to related but nonidentical situations outside of the laboratory (an issue that may be less worrisome in, for instance, areas of physics—see Hacking's, 1983, notion of the creation of phenomena that do not exist outside of certain kinds of apparatus). Brunswik's solution is to investigate representative slices of those conditions to which generalization is intended.

Proponents of systematic design also offer an alternative defensible route to generality of results. To the extent that manipulating features of the environment succeeds in identifying a causal model, this model fosters not only understanding but also prediction and control within the target domain for which the model was conceived. This route to the generality of results clearly relies not on the logic of induction (as representative design does) but on the logic of deduction. We suggest that both routes—representative sampling from the environment and testing the model's capacity of prediction and control within the target domain—are defensible, powerful, and nonexclusive routes to the challenge of generality of results. Moreover, the existence and acceptance of the other route may raise awareness of the costs and benefits of the route chosen.

Ecological Approaches to Cognition

It is not only opposing methodological convictions that have prevented representative design from being added to the methodological toolbox of mainstream psychology. Another obstacle, one to which Brunswik pointed us, is psychology's conceptual focus. He described it as follows:

> If there is anything that still ails psychology in general, and the psychology of cognition specifically, it is the neglect of investigation of environmental or ecological texture in favor of that of the texture of organismic structures and processes. Both historically and systematically psychology has forgotten that it is a science of organism–environment relationships, and has become a science of the organism. (Brunswik, 1957, p. 6)

Brunswik offered the theory of probabilistic functionalism as a framework for structuring investigations of organism–environment relations. He is not alone in this emphasis on the study of ecological structures. Other prominent research programs in cognitive psychology have also focused on organism–environment relations, albeit outside of the lens model framework and with a different emphasis on the specific nature of the interaction. It is possible that the most immediate link between environment and person was proposed by Gibson (1979). Unlike Brunswik, who was inspired by Helmholtz's unconscious inferences and Fritz Heider's functionalism, Gibson did not accept the distinction between distal stimuli and proximal cues. Rather, his notion of *direct perception* implies that light from an object is a property of the world that already contains all the information sufficient to make an accurate inference, and so mediating processes and uncertain inferences are unnecessary (for a discussion of the commonalities and differences between Heider, Brunswik, Gibson, and Marr, see Looren de Jong, 1995, and for more information on the roots of Gibson's approach and its connections to the work of William James and Roger Barker, see Heft, 2001).

The metaphor characterizing Roger N. Shepard's view of organism–environment relations is that of mirrors (P. M. Todd & Gigerenzer, 2001). The organism and the environment mirror each other inasmuch as the mind has internalized regularities in the environment that help one arrive at accurate representations of, for instance, an object's position, motion, and color (see Bloom & Finlay, 2001; Shepard, 1994/2001). Consequently, Shepard (1990) believes that "we may look into that window [on the mind] as through a glass darkly, but what we are beginning to discern there looks very much like a reflection of the world" (p. 213). Similarly, the rational analysis proposed by John R. Anderson (e.g., J. R. Anderson, 1990; also see Oaksford & Chater, 1998) holds that any study of a psychological mechanism should be preceded and informed by an analysis of the environment. For instance, J. R. Anderson and Schooler's (1991) analysis of *New York Times* headlines suggests that the general form of forgetting, in which the rate of forgetting slows down over time, reflects a function similar to that in the environment. In other words, the human memory system has learned environmental regularities and in essence has wagered that when information has not been used recently, it will probably not be needed in the future (for a similar approach, see Marr's, 1982, functional analysis of a particular task at the computational level).

For Brunswik, more than for Shepard and Anderson, the mind infers the world more than it reflects it. In contrast to these authors, Herbert Simon (1956, 1990) has argued that people are boundedly rational to the extent that they must function under conditions of limited time, knowledge, and computational capacity. This, however, does not imply that people arrive at irrational inferences; rather, the inferences are boundedly rational to the extent that people's cognitive strategies succeed in exploiting the informational structures of the environment. For Simon, human (boundedly) rational behavior can be understood only in terms of both the environment and the mind. He captured the interplay between mind and environment using a scissors metaphor: "Human rational behavior is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990, p. 7). As this quote illustrates, human (boundedly) rational behavior can be understood only in terms of both sides—environment and mind, much like a scissors needs two blades to cut. More recently, the idea of *ecological rationality*, or the match between cognitive processes and the statistical structure of the environments in which they function, has inspired research on boundedly rational heuristics (see Gigerenzer et al., 1999).

Beyond these emerging theoretical frameworks there have been powerful calls to more naturalistic research in cognitive psychology. For instance, Neisser (1978, 1982; Neisser & Winograd, 1988) has argued that research should be focused on natural memory phenomena, such as memory for life events, and be conducted in more naturalistic settings, using, for instance, techniques such as diary studies. The naturalistic decision-making perspective that has emerged in the field of judgment and decision making involves studying the decisions of domain experts in field settings (see Zsambok & Klein, 1997). In addition, there are fascinating case studies of various aspects of cognition and behavior, such as Hutchins's (1995) studies on navigation in the wild, Tweney's (2001) analysis of scientific diaries, Dunbar's (1995) in vivo analysis of scientific research practices, and Norman's (1988) analysis of the design of everyday things.

Of course, this list of researchers who pursue an ecological agenda is by no means comprehensive—many others deserve to be mentioned (for an overview, see Woll, 2002). Yet it seems fair to conclude that in mainstream cognitive psychology (e.g., research on attention, perception, memory, language, categorization, and thinking) the focus remains on the organism with a relative neglect of the ecology in which it evolved and functions. This focus makes it less imperative to study the mind's mechanisms in a context that *preserves*, in Brunswik's terminology, the naturally existing causal texture of the ecology. J. R. Anderson (1998) recently speculated on the future of cognitive psychology:

> If I were to prophesy the future, it would be that there will be more of this interplay between the architectural assumptions about mechanisms and the careful analysis of how these mechanisms are adapted to the structure of the world. (p. vii)

If Anderson is correct, and cognitive psychologists do direct attention toward the study of organism–environment relations, one may speculate that representative design will eventually make it into psychologists' methodological toolbox.

Final Remarks

Brunswik polarizes. For some, he was brilliant and creative; for others, his theoretical concepts are unintelligible and overly philosophic. For some, he wrote beautiful prose; others believe he could not write well. Some claim that representative design is impractical and, even worse, nonsensical, whereas others argue that is it iconoclastic and offers a methodology indigenous to psychology. There is no doubt that Brunswik's ideas and others' admiration or contempt for them invite a polemic—but a polemic was not our goal. Rather, our goal was to contribute to an understanding of his theory and methodology such that we can all appreciate the challenge they continue to pose.

References

Aarts, H., Verplanken, B., & Knippenberg, A. (1997). Habit and information use in travel mode choices. *Acta Psychologica, 96,* 1–14.

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research, 27,* 123–156.

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

Anderson, J. R. (1998). Foreword. In M. Oaksford & N. Chater, *Rational models of cognition* (pp. v–vii). New York: Oxford University Press.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408.

Anderson, J. R., & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 557–570). London: Oxford University Press.

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 250–267.

Balzer, W. K., Doherty, M. E., & O'Connor, R., Jr. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106,* 410–433.

Barker, R. G. (1968). *Ecological psychology.* Palo Alto, CA: Stanford University Press.

Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology, 70,* 110–129.

Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes, 41,* 1–19.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Erlbaum.

Björkman, M. (1969). On the ecological relevance of psychological research. *Scandinavian Journal of Psychology, 10,* 145–157.

Bloom, P., & Finlay, B. L. (Eds.). (2001). The work of Roger Shepard [Special issue]. *Behavioral and Brain Sciences, 24*(4).

Bottenberg, R. A., & Christal, R. E. (1961). *An iterative technique for clustering criteria which retains optimum predictive efficiency.* Lackland Air Force Base, TX: Personnel Laboratory, Wright Air Development Division.

Braspenning, J., & Sergeant, J. (1994). General practitioners' decision making for mental health problems: Outcomes and ecological validity. *Journal Clinical Epidemiology, 47,* 1365–1372.

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75–114). New York: North-Holland.

Brehmer, B. (1976). Social judgment theory and the analysis of interpersonal conflict. *Psychological Bulletin, 83,* 985–1003.

Brehmer, B. (1979). Preliminaries to a psychology of inference. *Scandinavian Journal of Psychology, 20,* 193–210.

Brehmer, B. (1980). Probabilistic functionalism in the laboratory: Learning and interpersonal (cognitive) conflict. In K. R. Hammond & N. E. Wascoe (Eds.), *Realizations of Brunswik's representative design* (pp. 13–24). San Francisco: Jossey-Bass.

Brehmer, B. (1994). The psychology of linear judgment models. *Acta Psychologica, 87,* 137–154.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9,* 171–184.

Bröder, A. (2000). Assessing the empirical validity of the "take the best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1332–1346.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513–531.

Brunswik, E. (1934). *Wahrnehmung und Gegenstandswelt: Grundlegung einer Psychologie vom Gegenstand her* [Perception and the world of objects: The foundations of a psychology in terms of objects]. Leipzig, Germany: Deuticke.

Brunswik, E. (1940). Thing constancy as measured by correlation coefficients. *Psychological Review, 47,* 69–78.

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

Brunswik, E. (1944). Distal focussing of perception: Size constancy in a representative sample of situations. *Psychological Monographs, 56,* 1–49.

Brunswik, E. (1945). Social perception of traits from photographs. *Psychological Bulletin, 42,* 535.

Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1, No. 10, pp. 656–760). Chicago: University of Chicago Press.

Brunswik, E. (1955a). In defense of probabilistic functionalism: A reply. *Psychological Review, 62,* 236–242.

Brunswik, E. (1955b). "Ratiomorphic" models of perception and thinking. *Acta Psychologica, 11,* 108–109.

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

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Brunswik, E. (1957). Scope and aspects of the cognitive problem. In H. Gruber, K. R. Hammond, & R. Jessor (Eds.), *Contemporary approaches to cognition* (pp. 5–31). Cambridge, MA: Harvard University Press.

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of "proximity" and of other Gestalt factors. *American Journal of Psychology, 66,* 20–32.

Brunswik, E., & Reiter, L. (1937). Eindrucks-Charaktere schematisierter Gesichter [Impression-characteristics of schematized faces]. *Zeitschrift fur Psychologie, 142,* 67–134.

Bülthoff, H. H., & van Veen, H. A. H. C. (2001). Vision and action in virtual environments: Modern psychophysics in spatial cognition research. In M. Jenkin & L. Harris (Eds.), *Vision and attention* (pp. 233–252). New York: Springer.

Burns, C. M. (2000). Navigation strategies with ecological displays. *International Journal of Human–Computer Studies, 52,* 111–129.

Christal, R. E. (1963). *JAN: A technique for analyzing individual and group judgment*. Lackland Air Force Base, TX: Personnel Research Laboratory, Aerospace Medical Division.

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior & Human Decision Processes, 48,* 147–168.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cook, R. L., & Stewart, T. R. (1975). A comparison of seven methods for obtaining subjective descriptions of judgmental policy. *Organizational Behavior and Human Performance, 123,* 31–45.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications.* San Diego, CA: Academic Press.

Cooksey, R. W. (2001). On Gibson's review of Brunswik and Kirlik's review of Gibson. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 242–244). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1–73.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127.

Crow, W. J. (1957). The need for representative design in studies of interpersonal perception. *Journal of Consulting Psychology, 21,* 323–325.

Cutting, J. E., Bruno, N., Brady, N. P., & Moore, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General, 121,* 364–381.

Cutting, J. E., Wang, R. F., Flückiger, M., & Baumberger, B. (1999). Human heading judgments and object-based motion information. *Vision Research, 39,* 1079–1105.

Czikszentmihalyi, M., & Larson, R. (1992). Validity and reliability of the experience sampling method. In M. W. De Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. 43–57). New York: Cambridge University Press.

Dalgleish, L. I. (1988). Decision making in child abuse cases: Applications of social judgment theory and signal detection theory. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 317–360). New York: North-Holland.

Dhami, M. K. (2001). *Bailing and jailing the fast and frugal way: An application of social judgement theory and simple heuristics to English magistrates' remand decisions*. Unpublished doctoral dissertation, City University, London.

Dhami, M. K. (2003). Psychological models of professional decision-making. *Psychological Science, 14,* 175–180.

Dhami, M. K. (2004). *Effects of systematic and representative stimulus design on judgment policy-capturing*. Manuscript submitted for publication.

Dhami, M. K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making, 14,* 141–168.

Dhami, M. K., & Harries, C. (2001). Fast and frugal versus regression models of human judgment. *Thinking & Reasoning, 7,* 5–27.

Doherty, M. E. (2001). Demonstrations for learning psychologists: "While God may not gamble, animals and humans do." In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 192–195). New York: Oxford University Press.

Dudycha, A. L. (1970). A Monte Carlo evaluation of JAN: A technique for capturing and clustering raters' policies. *Organizational Behavior and Human Performance, 5,* 501–516.

Dudycha, A. L., & Naylor, J. C. (1966). The effect of variations in the cue R matrix upon the obtained policy equation of judges. *Educational and Psychological Measurement, 26,* 583–603.

Dukes, W. F. (1951). Ecological representativeness in studying perceptual size-constancy in childhood. *American Journal of Psychology, 64,* 87–93.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: MIT Press.

Earle, T. C. (1973). Interpersonal learning. In L. Rappoport & D. A. Summers (Eds.), *Human judgment and social interaction* (pp. 240–266). New York: Holt, Rinehart & Winston.

Ebbesen, E. B., & Konecni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology, 32,* 805–821.

Ebbesen, E. B., & Konecni, V. J. (1980). On the external validity of decision-making research: What do we know about decisions in the real world? In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 21–45). Hillsdale, NJ: Erlbaum.

Ebbesen, E. B., Parker, S., & Konecni, V. J. (1977). Laboratory and field analyses of decisions involving risk. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 576–589.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory:

Processes of judgment and choice. *Annual Review of Psychology, 32,* 53–88.

Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist, 30,* 967–976.

Feigl, H. (1955). Functionalism, psychological theory and the uniting sciences: Some discussion remarks. *Psychological Review, 62,* 232–235.

Fiedler, K. (2000). Beware of samples! A cognitive–ecological sampling approach to judgment biases. *Psychological Review, 107,* 659–676.

Fiedler, K., & Juslin, P. (Eds.). (in press). *Information sampling and adaptive cognition*. Cambridge, England: Cambridge University Press.

Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment biases in a simulated classroom—A cognitive–environmental approach. *Organizational Behavior and Human Decision Processes, 88,* 527–561.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society A: Optics, Image Science, and Vision, 4,* 2379–2394.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, England: Cambridge University Press.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior and Human Performance, 13,* 1–16.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver & Boyd.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102,* 652–670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to social perception*. San Diego, CA: Academic Press.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gibson, J. J. (2001). Survival in a world of probable objects. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 244–246). Oxford, England: Oxford University Press. (Original work published 1957)

Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology, 66,* 398–412.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond heuristics and biases. In W. Stroebe & H. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 83–115). Chichester, England: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103,* 592–596.

Gigerenzer, G. (2001). Ideas in exile: The struggles of an upright man. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 445–452). Oxford, England: Oxford University Press.

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

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704.

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

Gigerenzer, G., & Kurz, E. (2001). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 342–347). Oxford, England: Oxford University Press.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge, England: Cambridge University Press.

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

Gillis, J., & Schneider, C. (1966). The historical preconditions of representative design. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 204–236). New York: Holt, Rinehart and Winston.

Goldstein, W. M., & Wright, J. H. (2001). "Perception" versus "thinking": Brunswikian thought on central responses and processes. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 249–256). Oxford, England: Oxford University Press.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge, England: Cambridge University Press.

Hammond, K. R. (1948). Subject and object sampling—A note. *Psychological Bulletin, 45,* 530–533.

Hammond, K. R. (1951). Relativity and representativeness. *Philosophy of Science, 18,* 208–211.

Hammond, K. R. (1954). Representative vs. systematic design in clinical psychology. *Psychological Bulletin, 51,* 150–159.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychology Review, 62,* 255–262.

Hammond, K. R. (1965). New directions in research in conflict resolution. *Journal of Social Issues, 21,* 255–262.

Hammond, K. R. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 15–80). New York: Holt, Rinehart and Winston.

Hammond, K. R. (1972). Inductive knowing. In J. R. Royce & W. W. Rozeboom (Eds.), *The psychology of knowing* (pp. 285–320). New York: Gordon & Breach.

Hammond, K. R. (1973). The cognitive conflict paradigm. In L. Rappoport & D. A. Summers (Eds.), *Human judgment and social interaction* (pp. 188–205). New York: Holt, Rinehart & Winston.

Hammond, K. R. (1978). *Psychology's scientific revolution: Is it in danger?* Technical Report No. 211, University of Colorado, Center for Research on Judgment and Policy.

Hammond, K. R. (1986). Generalization in operational contexts: What does it mean? Can it be done? *IEEE Transactions on Systems, Man, and Cybernetics, 16,* 428–433.

Hammond, K. R. (1990). Functionalism and illusionism: Can integration be usefully achieved? In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 227–261). Chicago: Chicago University Press.

Hammond, K. R. (1996a). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavailable injustice.* New York: Oxford University Press.

Hammond, K. R. (1996b). Upon reflection. *Thinking & Reasoning, 2,* 239–248.

Hammond, K. R. (1998). *Representative design*. Retrieved October 9, 1998, from http://www.brunswik.org/notes/essay3.html

Hammond, K. R. (2000). Coherence and correspondence theories in judgment and decision making. In T. Connolly & H. R. Arkes (Eds.), *Judgment and decision making: An interdisciplinary reader* (2nd ed., pp. 53–65). New York: Cambridge University Press.

Hammond, K. R. (2001). Expansion of Egon Brunswik's psychology, 1955–1995. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 464–478). New York: Oxford University Press.

Hammond, K. R., & Adelman, L. (1976, October 22). Science, values, and human judgment. *Science, 194,* 389–396.

Hammond, K. R., Hamm, R. M., & Grassia, J. (1986). Generalizing over conditions by combining the multitrait–multimethod matrix and the representative design of experiments. *Psychological Bulletin, 100,* 257–269.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytic cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics, 17,* 753–770.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71,* 438–456.

Hammond, K. R., & Stewart, T. R. (1974). *The interaction between design and discovery in the study of human judgment.* Report No. 152, University of Colorado, Institute of Behavioral Science.

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

Hammond, K. R., & Stewart, T. R. (2001b). Introduction. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 3–11). Oxford, England: Oxford University Press.

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. 271–317). New York: Academic Press.

Hammond, K. R., & Summers, D. A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review, 72,* 215–224.

Hammond, K. R., Todd, F. J., Wilkins, M., & Mitchell, T. O. (1966). Cognitive conflict between persons: Application of the "lens model" paradigm. *Journal of Experimental Social Psychology, 2,* 343–360.

Hammond, K. R., & Wascoe, N. E. (Eds.). (1980). *Realizations of Brunswik's representative design.* San Francisco: Jossey-Bass.

Hammond, K. R., Wilkins, M. M., & Todd, F. J. (1966). A research paradigm for the study of interpersonal learning. *Psychological Bulletin, 65,* 221–232.

Harries, C., St. Evans, B. J., Dennis, I., & Dean, J. (1996). A clinical judgement analysis of prescribing decisions in general practice. *Le Travail Humain, 59,* 87–111.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Hastie, R., & Hammond, K. R. (1991). Rational analysis and the lens model. *Behavioral and Brain Sciences, 14,* 498.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311–327.

Heald, J. E. (1991). Social judgment theory: Applications to educational decision making. *Educational Administration Quarterly, 27,* 343–357.

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism.* Mahwah, NJ: Erlbaum.

Hertwig, R., & Chase, V. M. (1998). Many reasons or just one: How response mode affects reasoning in the conjunction problem. *Thinking & Reasoning, 4,* 319–352.

Hertwig, R., Fanselow, C., & Hoffrage, U. (2003). Hindsight bias: How knowledge and heuristics affect our reconstruction of the past. *Memory, 11,* 357–377.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24,* 383–451.

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

Hirsch, G., & Immediato, C. S. (1999). Microworlds and generic structures as resources for integrating care and improving health. *System Dynamics Review, 15,* 315–330.

Hochberg, J. (1966). Representative sampling and the purposes of perceptual research: Pictures of the world, and the world of pictures. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 361–381). New York: Holt, Rinehart and Winston.

Hoffman, P. J., Slovic, P., & Rorer, L. G. (1968). An analysis of variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin, 69,* 338–349.

Hoffrage, U. (in press). Overconfidence. In R. Pohl (Ed.), *Cognitive illusions: Fallacies and biases in thinking, judgement, and memory*. Hove, England: Psychology Press.

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

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 566–581.

Hoffrage, U., & Pohl, R. F. (Eds.). (2003). Hindsight bias [Special issue]. *Memory,* 11(4/5).

Holaday, B. E. (1933). Die Größenkonstanz der Sehdinge bei Variation der inneren und äußeren Wahrnehmungsbedingungen [Size constancy for visual objects under variation of internal and external perceptual conditions]. *Achiv für die Gesamte Psychologie, 88,* 419–486.

Hull, C. L. (1943). The problem of intervening variables in molar behavior theory. *Psychological Review, 50,* 273–291.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Jenkins, J. J. (1974). Remember that old theory of memory? Well, forget it! *American Psychologist, 29,* 785–795.

Juslin, P. (1993). An explanation of the "hard–easy effect" in studies of realism of confidence in one's general knowledge. *European Journal of Cognitive Psychology, 5,* 55–71.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57,* 226–246.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104,* 344–366.

Juslin, P., Olsson, H., & Björkman, M. (1997). Brunswikian and Thurstonian origins of bias in probability assessment: On the interpretation of stochastic components of judgment. *Journal of Behavioral Decision Making, 10,* 189–209.

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

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14,* 383–418.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103,* 582–591.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.

Kirlik, A. (2001). On Gibson's review of Brunswik. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 238–242). New York: Oxford University Press.

Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 115–162). New York: North-Holland.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Koch, S. (1959). Epilogue. In S. Koch (Ed.), *Psychology: Vol. 3. A study of a science* (pp. 729–788). New York: McGraw-Hill.

Kordaki, M., & Potari, D. (1998). A learning environment for the conservation of area and its measurement: A computer microworld. *Computers & Education, 31,* 405–422.

Krech, D. (1955). Discussion: Theory and reductionism. *Psychological Review, 62,* 229–231.

Krueger, J. I., & Funder, D. C. (in press). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*.

Kurz, E. M., & Hertwig, R. (2001). To know an experimenter. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 180–186). New York: Oxford University Press.

Kurz, E., & Martignon, L. (1999). Weighing, then summing: The triumph and tumbling of a modeling practice in psychology. In L. Magnani, N. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 26–31). Pavia, Italy: Cariplo.

Kurz, E. M., & Tweney, R. D. (1997). The heretical psychology of Egon Brunswik. In W. G. Bringmann, H. E. Lueck, R. Miller, & C. E. Early (Eds.), *A pictorial history of psychology* (pp. 221–232). Carol Stream, IL: Quintessence.

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.

Leeper, R. W. (1966). A critical consideration of Egon Brunswik's probabilistic functionalism. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 405–454). New York: Holt, Rinehart & Winston.

Lewin, K. (1943). Defining the "field at a given time." *Psychological Review, 50,* 292–310.

Libby, R., & Lewis, B. L. (1982). Human information processing research in accounting: The state of the art 1982. *Accounting, Organizations, and Society, 7,* 231–285.

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). New York: Cambridge University Press.

Looren de Jong, H. (1995). Ecological psychology and naturalism: Heider, Gibson and Marr. *Theory and Psychology, 5,* 251–269.

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

Madden, J. M. (1963). *An application to job evaluation of a policy-capturing model for analyzing individual and group judgment.* Lackland Air Force Base, TX: Personnel Research Laboratory, Aerospace Medical Division.

Maher, B. A. (1978). Stimulus sampling in clinical research: Representative design reviewed. *Journal of Consulting and Clinical Psychology, 46,* 643–647.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

McClelland, G. H. (1997). Optimal design in psychological research. *Psychological Methods, 2,* 3–19.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390.

Mellers, B. A., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269–275.

Mellers, B. A., Schwartz, A., & Cooke, A. D. J. (1998). Judgment and decision making. *Annual Review of Psychology, 49,* 447–477.

Mook, D. G. (1989). The myth of external validity. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and later life* (pp. 25–43). Cambridge, England: Cambridge University Press.

Moore, W. L., & Holbrook, M. B. (1990). Conjoint analysis on objects with environmentally correlated attributes: The questionable importance of representative design. *Journal of Consumer Research, 16,* 490–497.

Mumpower, J., & Adelman, L. (1980). The application of Brunswikian methodology to policy formation. In K. R. Hammond & N. E. Wascoe (Eds.), *Realizations of Brunswik's representative design* (pp. 49–60). San Francisco: Jossey-Bass.

Mumpower, J. L., & Stewart, T. R. (1996). Expert judgment and expert disagreement. *Thinking & Reasoning, 2,* 191–211.

Naylor, J. C., Dudycha, A. L., & Schenck, E. A. (1967). An empirical comparison of ρ_a and ρ_m as indices of rater policy agreement. *Educational and Psychological Measurement, 27,* 7–20.

Naylor, J. C., & Wherry, R. J. (1964). *Feasibility of distinguishing supervisors' policies in evaluation of subordinates by using ratings of simulated job incumbents.* Lackland Air Force Base, TX: Personnel Research Laboratory Aerospace Medical Division.

Naylor, J. C., & Wherry, R. J. (1965). The use of simulated stimuli and the JAN technique to capture and cluster the policies of raters. *Educational and Psychological Measurement, 25,* 969–986.

Neisser, U. (1978). Memory: What are the important questions? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 3–24). London: Academic Press.

Neisser, U. (Ed.). (1982). *Memory observed*. San Francisco: Freeman.

Neisser, U., & Winograd, E. (Eds.). (1988, September). Remembering reconsidered: Ecological and traditional approaches to the study of memory. In U. Neisser (Chair), *Emory Symposia in Cognition, 2*. Symposium conducted at Emory University, Atlanta, GA.

Norman, D. (1988). *The psychology of everyday things*. New York: Basic Books.

Oaksford, M., & Chater, N. (1998). *Rational models of cognition*. New York: Oxford University Press.

Olson, G. A., Dell'omo, G. G., & Jarley, P. (1992). A comparison of interest arbitrator decision-making in experimental and field settings. *Industrial and Labor Relations Review, 45,* 711–723.

Omodei, M. M., & Wearing, A. J. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision making behavior. *Behavior Research Methods, Instruments & Computers, 27,* 303–316.

Oppewal, H., & Timmermans, H. (1999). Modeling consumer perception of public space in shopping centers. *Environment and Behavior, 31,* 45–65.

Petrinovich, L. (1979). Probabilistic functionalism: A conception of research method. *American Psychologist, 34,* 373–390.

Petrinovich, L. (1980). Brunswikian behavioral biology. In K. R. Hammond & N. E. Wascoe (Eds.), *Realizations of Brunswik's representative design* (pp. 85–93). San Francisco: Jossey-Bass.

Phelps, R. H., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance, 21,* 209–219.

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory, 11,* 337–356.

Poon, L. W., Rubin, D. C., & Wilson, B. A. (Eds.). (1989). *Everyday cognition in adulthood and later life*. Cambridge, England: Cambridge University Press.

Postman, L. (1955). The probability approach and nomothetic theory. *Psychological Review, 62,* 218–225.

Postman, L., & Tolman, E. C. (1959). Brunswik's probabilistic functionalism. In S. Koch (Ed.), *Psychology: A study of science* (pp. 502–564). New York: McGraw-Hill.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misperceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Reed, E. S. (1988). *James J. Gibson and the psychology of perception.* New Haven, CT: Yale University Press.

Reimer, T., & Hoffrage, U. (2003). *Can simple group heuristics detect hidden profiles?* Manuscript submitted for publication.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC

Research Group (Eds.), *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.
Rohrbaugh, J. (1988). Cognitive conflict tasks and small group processes. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 199–226). New York: North-Holland.
Schwarz, R. (2001). Evolutionary internalized regularities. *Behavioral and Brain Sciences, 24,* 626–628.
Sedlmeier, P., Hertwig, R., & Gigerenzer, G. (1998). Are judgments of the positional frequencies of letters systematically biased due to availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 754–770.
Shanteau, J., & Stewart, T. R. (1992). Why study expert decision making: Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes, 53,* 95–106.
Shaw, K. T., & Gifford, R. (1994). Residents' and burglars' assessment of burglary risk from defensible space cues. *Journal of Environmental Psychology, 14,* 177–194.
Shepard, R. N. (1990). *Mind sights: Original visual illusions, ambiguities, and other anomalies*. New York: Freeman.
Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 495–532). New York: Oxford University Press.
Shepard, R. N. (2001). Perceptual–cognitive universals as reflections of the world. *Behavioral and Brain Sciences, 24,* 581–601. (Reprinted from *Psychonomic Bulletin and Review, 1,* 2–28). (Original work published 1994)
Sherif, M., & Hovland, C. I. (1961). *Social judgment.* New Haven, CT: Yale University Press.
Sieber, J. E., & Saks, M. J. (1989). A census of subject pool characteristics and policies. *American Psychologist, 44,* 1053–1061.
Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129–138.
Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41,* 1–19.
Sjöberg, L. (1971). The new functionalism. *Scandinavian Journal of Psychology, 12,* 29–52.
Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744.
Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65,* 117–137.
Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48,* 1467–1478.
Stewart, T. (1976). Components of correlations and extensions of the lens model equation. *Psychometrika, 41,* 101–120.
Stewart, T. (1988). Judgment analysis: Procedures. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 41–74). New York: North-Holland.
Stewart, T. R., Middleton, P., Downton, M., & Ely, D. (1984). Judgments of photographs vs. field observations in studies of perception and judgment of the visual environment. *Journal of Environmental Psychology, 4,* 283–302.
Summers, D. A., Taliaferro, J. D., & Fletcher, D. J. (1970). Subjective vs. objective description of judgment policy. *Psychonomic Science, 18,* 249–250.
Todd, F. J., & Hammond, K. R. (1965). Differential feedback in two multiple-cue probability learning tasks. *Behavioral Science, 10,* 429–435.
Todd, P. M., & Gigerenzer, G. (2001). Shepard's mirrors or Simon's scissors? *Behavioral and Brain Sciences, 24,* 704–705.
Tolman, E. C. (1955, November 11). Egon Brunswik, psychologist and philosopher of science. *Science, 122,* 910.
Tweney, R. (2001). *Brunswikian accounts of scientific thinking*. Retrieved October 30, 2001, from http://www.albany.edu/cpr/brunswik/newsletters/newsletter2001/2001news.pdf
Waller, W. S. (1988). Brunswikian research in accounting and auditing. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 247–272). New York: North-Holland.
Wastell, D. (1996). Human–machine dynamics in complex information systems: The "microworld" paradigm as a heuristic tool for developing theory and exploring design issues. *Information Systems Journal, 6,* 245–260.
Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115–1125.
Westenberg, M. R. M., & Koele, P. (1992). Response modes, decision processes and decision outcomes. *Acta Psychologica, 80,* 169–184.
Wigton, R. S. (1988). Use of linear models to analyze physicians' decisions. *Medical Decision Making, 8,* 241–252.
Wigton, R. S. (1996). Social judgment theory and medical judgment. *Thinking & Reasoning, 2,* 175–190.
Winman, A. (1997). The importance of item selection in "knew-it-all-along" studies of general knowledge. *Scandinavian Journal of Psychology, 38,* 63–72.
Wolf, B. (1999). *Vicarious functioning as a central process-characteristic of human behavior*. Retrieved May 27, 1999, from http://www.albany.edu/cpr/brunswik/essay4.html
Woll, S. (2002). *Everyday thinking: Memory, reasoning, and judgment in the real world*. Mahwah, NJ: Erlbaum.
Woodworth, R. (1938). *Experimental psychology*. New York: Holt.
York, K. M. (1992). A policy capturing analysis of federal district and appellate court sexual harassment cases. *Employee Responsibilities and Rights Journal, 5,* 173–184.
Zsambok, C. E., & Klein, G. (Eds.). (1997). *Naturalistic decision making.* Mahwah, NJ: Erlbaum.

Received August 1, 2003
Revision received February 5, 2004
Accepted June 4, 2004 ■