Information Integration in Child Welfare Cases: An Introduction to Statistical Decision Making

John Ruscio
Brandeis University

Workers in the field of child maltreatment are required to make many complex and far-reaching decisions every week. In this article, two general methods for formulating decision-making policies are presented, along with a discussion of the considerable research literature demonstrating the superior predictive validity of statistical decision models over clinical prediction. A series of illustrative contrasts between the two approaches highlights the desirable mathematical properties of statistical equations as well as the cognitive biases and limitations inherent in unaided human judgment. Reasons for practitioners' adherence to the clinical approach are explored, with specific reference to child welfare decision making. Finally, recommendations are provided for enhancing the efficiency, validity, and ethical defensibility of decision making that seriously impacts the lives of children and their families.

Imagine yourself as an investigator in the child welfare system. On a weekly basis, you are asked to evaluate a number of cases of alleged child maltreatment to decide whether a child should remain in his or her family or be placed outside of the home. Each case that you review contains a wide array of information regarding family history, current home environment, and other types of data obtained through interviews. Your caseload and your deadline pressures are high. How do you make these decisions that will seriously affect the lives of children and their families? That is, how do you determine the relative risk in each case and set appropriate cutoffs for alternative decisions? The answer to these questions amounts to the formation of a decision policy, a system that you will use to combine the available information with the goal of maximizing correct decisions.

The efficacy of your decision policy can be examined through the use of a simple fourfold classification table crossing the optimal outcome for each child (kept at home vs. placed into care) with the decision that is reached. There are two types of correct decisions, or "hits," that are possible: True positives are decisions that place children into care when appropriate, and true negatives are decisions that keep children at home when appropriate. There are also two types of incorrect decisions, or "misses," that are possible: False positives are decisions that unnecessarily place children into care, and false negatives are decisions that fail to place children into care when placement is necessary. Based on this classification table, the effectiveness of a decision policy may be evaluated in several ways. For instance, one could determine how many of the decisions to place a child into foster care were correct (true positives divided by the sum of true and false positives); how many children who optimally should have been kept in the home actually were (true negatives divided by the sum of true negatives and false positives); or how many placement decisions, overall, were correct (the sum of true positives and true negatives divided by the total number of cases).

The latter index of overall correct decisions, called the *hit rate*, is arguably the most appropriate gauge of the efficacy of placement decision policies because both false positives and false negatives can result in serious harm to the child and his or her family (Berger, Rolon, Sachs, & Wilson, 1989). For any decision

Author's Note: This research was supported by a grant to the author from the National Center on Child Abuse and Neglect (90-CA-1576). Their support is gratefully acknowledged. The author thanks Theodore P. Cross, Ayelet Meron Ruscio, and two anonymous reviewers for helpful suggestions at several points in the development of this article. Correspondence concerning this article should be addressed to John Ruscio, Department of Psychology (MS 062), Brandeis University, Waltham, MA 02254-9110; e-mail ruscio@brandeis.edu.

CHILD MALTREATMENT, Vol. 3, No. 2, May 1998 143-156
© 1998 Sage Publications, Inc.

policy, a wide range of threshold values, or cut scores, could be used for choosing between alternative decisions. However, there always exists an optimal cut score that maximizes the hit rate (Meehl & Rosen, 1955). Although decision policies can be custom tailored to selectively reduce either false positives or false negatives by shifting this cut score, such a shift reduces one type of error only at the expense of an increase in the other type of error. For example, if the cut score for placing a child into foster care is set at a very low level to prevent false negatives (failures to remove children from their homes when necessary) the rate of false positives (unnecessary removals) will be increased. Moreover, the total number of erroneous decisions increases with any deviation from the cut score at which the hit rate is at a maximum. Unless one can convincingly argue that one type of incorrect decision has a considerably more negative impact than the other, the safest course is to use the optimal cut score (Meehl & Rosen, 1955).

The efficacy of a cut score hinges critically on the reliability and validity of the decision policy. Ideally, child welfare workers would use a policy in placement decision making that had perfect reliability and validity. Perfect reliability indicates that decisions are replicable, meaning that they are consistent both within and between workers. A child welfare worker should therefore consistently apply one decision policy and agree in recommendations for each case with other workers. Perfect validity indicates that workers successfully identify those and only those cases optimally served by placement outside the home. However, actual decision policies fall far short of this goal. Given two realistic approximations, Lindsey (1992) developed a numerical illustration of the effect of imperfect reliability on 100 typical placement decisions. Based on a nationwide survey (Lindsey, 1991), he estimated that 30% of all cases referred to child welfare services are placed into foster care. Second, based on his extensive review of the literature, he estimated that the coefficient of concordance (reliability) of placement decisions is .25. A coefficient of .00 would represent purely chance-level agreement, or blind guessing, whereas a coefficient of 1.00 would represent perfect agreement. A coefficient of .25, therefore, means that any randomly selected pair of child welfare workers are likely to agree with one another at a rate 25% above chance levels. The reliability of decision making is an issue taken up in much more detail later, but it is worth noting at this point that the .25 estimate is actually the highest reliability coefficient that Lindsey (1992) encountered in his review, yielding a fairly charitable scenario.

(a) Validity = 1.00
Hit rate = (16 + 56)/100 = 72%

		OPTIMAL OUTCOME		
		Place into care	Keep at home	
ACTUAL DECISION	Place into care	16	15	31
	Keep at home	13	56	69
		29	71	

(b) Validity = .00
Hit rate = (9 + 49)/100 = 58%

		OPTIMAL OUTCOME		
		Place into care	Keep at home	
ACTUAL DECISION	Place into care	9	22	31
	Keep at home	20	49	69
		29	71	

(c) Validity = .40
Hit rate = (12 + 52)/100 = 64%

		OPTIMAL OUTCOME		
		Place into care	Keep at home	
ACTUAL DECISION	Place into care	12	19	31
	Keep at home	17	52	69
		29	71	

FIGURE 1: Classification Tables for Three Different Child Placement Decision Policies With Poor Reliability (.25): (a) Perfect Validity, (b) Zero Validity, (c) Moderate Validity
NOTE: Note that approximately 30% of all cases are placed into care and that placement into care is the optimal outcome for approximately 30% of all cases. These marginal totals, based on Lindsey's (1992) example, are held constant across tables.

Without introducing any bias or prejudice into the system (i.e., assuming perfect validity), Lindsey (1992) derived the classification results shown in Figure 1a. Although the hit rate is 72%, both the false positive and false negative rates are alarming. Simple calculations reveal that 48% (15 / [15 + 16]) of the placements into care were unnecessary and that 45% (13 / [13 + 16]) of the children needing such placement were denied it. These results represent the best decisions that can be made when reliability is so drastically constrained because the validity of the decision policy was assumed to be perfect. However, a more realistic estimate of validity is needed to more accurately represent the actual hit rates of clinical decision policies currently in use.

An estimate of the actual hit rate can be obtained by interpolating between the theoretical minimum and maximum values. Holding constant the marginal totals in Lindsey's (1992) example, the classification results of a decision policy possessing zero validity (a completely random process) are shown in Figure 1b. The hit rate of this random decision policy is 58%. The actual hit rate that would be obtained using a decision policy of moderate validity therefore falls in the interval ranging from 58% (zero validity) to 72% (perfect validity). Because it will make little difference for the comparisons made later, the actual hit rate will be assumed to fall two fifths of the way from the low end to the high end of the range, or approximately at the 64% mark (see Figure 1c). This means that 64 out of 100 cases are decided correctly and that the remaining 36 children and their families suffer the negative consequences of unnecessary placement into foster care or the denial of necessary placement.

Few would maintain that such a decision policy realizes the goals of effective child protection. The remainder of this article will compare two general approaches to decision making using the situation outlined above to demonstrate that substantial improvements in the hit rate of child welfare decisions can be obtained through the use of a more efficient technique: statistical decision making.

TWO GENERAL APPROACHES TO DECISION MAKING

There are two alternative methods for making decisions, and these are traditionally referred to as the clinical and statistical approaches (Meehl, 1954; Sarbin, 1944; Sawyer, 1966; Wiggins, 1981). The clinical approach involves nothing more sophisticated than a human judge evaluating available information and arriving at a decision. This technique does not necessarily involve a professional clinician—the term is broadly applied whenever a human judge forecasts outcomes. The statistical technique involves the use of an equation derived from quantitative information to maximize the accuracy of predictions. Information from one sample of cases is used to develop a statistical decision-making equation that is then tested, or cross-validated, on an independent sample of cases.

In the end, any decision policy is subsumed under one of these two approaches. Although a common proposition is that any reasonable decision maker uses a combination of these methods (e.g., Kleinmuntz, 1990), there is an important sense in which this cannot literally be true. When outcomes recommended by these approaches agree, the distinction between approaches becomes irrelevant. However, when these recommendations do not agree, it becomes important to understand the relative strengths and weaknesses of each approach to arrive at a decision. When we seek to resolve discrepancies between the outcomes of these approaches, we come to the body of research investigating these very differences (Meehl, 1957).

In such circumstances, it can be argued that the only fair and ethically defensible decision policy is the one with a demonstrably superior predictive validity (Dawes, 1994). So long as one can assess an appropriate, measurable outcome arising from a good decision policy, the two approaches' hit rates (or predictive validities) can readily be compared in an objective, empirical fashion.

EMPIRICAL EVALUATION OF THE TWO APPROACHES

The clinical approach to decision making has been associated with a desire for causal rather than probabilistic explanations (Dawes, 1991, 1994, 1995) as well as with a search for narrative truth—a plausible "good story"—as opposed to a purely factual or historical account (Sarbin, 1986; Spence, 1982). It has also been identified as a more risky approach, given its practitioners' refusal to accept the error inherent in statistical techniques by aspiring to perfect predictability (Einhorn, 1986). Although there is a long-standing debate regarding the relative efficacy of these two approaches, research spanning more than four decades has consistently indicated that the predictive validity of human judgment is inferior to that of statistical prediction equations (for reviews of the evidence, see Dawes, Faust, & Meehl, 1989; Meehl, 1954; Sawyer, 1966; Wiggins, 1981). A partial sampling of the wide range of domains in which statistical prediction surpasses clinical prediction includes academic success, business bankruptcy, longevity, military training success, myocardial infarction, neuropsychological diagnosis, parole violation, police officer termination, psychiatric diagnosis, and violence (Dawes, Faust, & Meehl, 1993). Findings indicate that a probabilistic relationship is more readily obtained and verified than a causal understanding, that historical truth is more accurate than narrative truth, and that the acceptance of a fixed amount of error leads to a minimal number of incorrect decisions.

One particular review of the evidence is of special relevance here. Sawyer (1966) compared clinical and mechanical (statistical) approaches in all empirical studies that were available at that time. He made comparisons separately at the stages of data collection and data combination. Clinical data collection includes, for example, questionnaires and interviews that lack normative standards against which to compare individual responses. In contrast, mechanical data collection includes testing and assessment devices that have been standardized in their format, administration, and/or scoring. Methods of clinical and mechanical data combination have already been discussed. Sawyer's review of empirical investigations examined the validity of four combinations of approaches: clinical collection and combination, clinical collection with mechanical combination, mechanical collection with clinical combination, and mechanical collection and combination.

The results clearly indicated that mechanical procedures were independently superior at both stages of the decision-making process (Sawyer, 1966). That is, regardless of the method of data collection, information is combined more validly by a statistical equation than by human judgment. Therefore, even though clinical data collection is the norm in child welfare decision making, it would still be profitable to employ mechanical data combination.

Paul Meehl, a bold proponent of the statistical approach, has urged decision makers to attend to these findings since his review of the logical and empirical status of the controversy in 1954. By 1986, he was able to cite a vast array of research supporting his initial position.

> There is no controversy in social science that shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing 90 investigations [as of 1992, the number is 175 (Meehl, 1992)], predicting everything from the outcome of football games to the diagnosis of liver disease and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion. (Meehl, 1986, p. 374)

The comparatively high predictive validity of statistical decision making has been attributed to two sources: desirable properties of statistical techniques and undesirable cognitive biases of human judges.

EXPLANATIONS FOR THE EFFICACY OF STATISTICAL DECISION MAKING

The comparatively high predictive validity of statistical decision making has been attributed to two sources: desirable properties of statistical techniques and undesirable cognitive biases of human judges. Statistical equations possess many advantageous mathematical properties, and contrasted with human judges, they are extremely effective in detecting relationships amid considerable variation. Cognitive biases serve to broaden the gap between statistical and clinical predictions. Meehl (1992), for example, lists 37 sources of error contributing to the faulty judgments of both clinicians and scientists. Goldberg (1991) outlines five illustrative contrasts between the two approaches that help to explain the superiority of statistical decision making, each of which will be discussed in turn.

Validity

Statistical techniques detect and adjust for differential predictor, or "cue," validities (Dawes, 1971, 1979; Dawes & Corrigan, 1974). Valid cues, or predictors that are empirically related to an important outcome, are weighted more heavily than are invalid cues, or predictors that are not empirically related to the outcome measure. In this way, information is used to the extent that it is empirically valuable. Human judges, in contrast, bring various biases to the prediction task (Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980; Turk & Salovey, 1988). In particular, information processing often relies on three common heuristics, or mental shortcuts. The availability heuristic biases judgment in the direction of instances that are selectively recalled from memory (Tversky & Kahneman, 1973). The representativeness heuristic causes judges to make decisions based on the perceived similarity or goodness of fit of variables rather than on their actual relatedness (Kahneman & Tversky, 1972). Finally, the anchoring heuristic biases judgment toward a rough estimation or partial solution that is insufficiently adjusted in its magnitude (Tversky & Kahneman, 1974). Demonstrations of each of these heuristics appear in the appendix.

Several other types of biases result from the application of these heuristics. For example, superstitious beliefs can develop when people mistakenly perceive relationships in random sequences of events (Gilovich, 1991; Gilovich, Vallone, & Tversky, 1985; Tversky & Kahneman, 1974); preconceived notions about relationships often shape judgment more powerfully than do empirical data (Crocker, 1981; Jennings, Amabile, & Ross, 1982); illusory correlations formed from hunches or suspicions can be surprisingly resistant to disconfirmation (Chapman & Chapman, 1967, 1971; Kurtz & Garfield, 1978); and base rate information, the proportion of cases actually in each of various relevant categories, is underused (Kahneman & Tversky, 1973; Nisbett, Borgida, Crandall, & Reed, 1976). This is but a partial listing of the many biases that contribute to the faulty cue weighting of human judges.

Although comparisons of clinical and statistical approaches to risk assessment have not been made in the field of child welfare, such comparisons have been conducted in investigations of parole decisions (e.g., Quinsey & Maguire, 1986; Wormith & Goldstone, 1984). One study found that when rating the dangerousness of men in a maximum security psychiatric institution, experienced forensic clinicians weighted the frequency of aggression in the institution fairly heavily in their decision policy. Although intuitively compelling, the weighting of institutional aggression was in fact unwarranted. Institutional aggression was found to be unrelated to recidivism outcomes collected over a period of 11 years or more, resulting in overall dangerousness ratings uncorrelated with this outcome (Quinsey & Maguire, 1986). Statistical prediction systems weighted information according to its validity, thereby surpassing the accuracy of clinical prediction.

Units of Measurement

Statistical techniques align the metrics (or scales) of cues with that of the criterion, allowing optimal use of cues regardless of their units of measurement. Human judges, on the other hand, have difficulty combining cues that are measured in different ways and tend to overuse cues that are presented in the same metric as the criterion (Slovic & MacPhillamy, 1974). In one study, for example, participants were presented with one ranked cue (an ability test) and one categorical cue (a strong vs. weak pattern of extracurricular activities) and were asked to predict college admissions decisions in terms of either ranks or categories (Tversky, Sattath, & Slovic, 1988). Individuals asked to predict ranks weighted the ranked cue significantly higher than did those asked to predict categorically, with the opposite pattern of results obtained for categorical cue weights.

The units of measurement of information typically available in child welfare decision making vary widely. Many types of information are strictly categorical in nature (e.g., child sex and race, the relationship of an alleged perpetrator to the child), whereas other types of information are more continuous in nature (e.g., child age, the duration and number of abuse incidents). Without a common metric, it is remarkably difficult for human judges to effectively combine such disparate types of information.

Reliability

Judgmental reliability indicates replicability, or the extent to which decisions are consistently made. Reliability can be established within one judge by determining how consistently a single decision policy is applied, which is analogous to test-retest reliability, or between several judges by determining the consistency in judgments across multiple decision policies, which is analogous to parallel forms reliability. The validity of decisions is constrained by a lack of either type of reliability. For example, suppose an individual decides to place a child outside the home on one occasion and is then surreptitiously given the same exact information at a later date to assess judgmental consistency. If the decision is now to keep the child at home, indicating very poor reliability, at most one of these decisions is correct, and validity is at chance levels—the decision maker may as well be flipping a coin. Likewise, if two different decision makers make different decisions regarding this case, at most one of them can be correct, and the validity of one or both of the decision makers is necessarily low. The maximum attainable validity will be constrained by an unreliable judgment process.

Poor reliability (both within and between judges) is perhaps the largest problem faced in the domain of child welfare decision making. Many authors have attributed this problem to a lack of well-defined institutional goals and to the predominantly subjective nature of decisions made in the absence of a theoretical or empirical foundation for prediction (Alter, 1985; Berger et al., 1989; Corby & Mills, 1986; DePanfilis & Scannapieco, 1994; Gleeson, 1987; Haskett, Wayland, Hutcheson, & Tavana, 1995; Jones, 1993; Miller, Fisher, & Sinclair, 1993; Schuerman & Vogel, 1986; Stein & Rzepnicki, 1983; Wick & Schoech, 1988). The result of these limitations is poor within and between judge agreement, even among experts in the field (Lindsey, 1992; Mandel, Lehman, & Yuille, 1994; Sicoly, 1989; but see Alter, 1985, for a counterexample). After reviewing the extant research on child welfare decision making, Lindsey (1992) seriously questioned the foundation for the dramatic interventions prevalent today, arguing that such actions lack sufficient scientific and clinical bases. However, whereas Lindsey concludes that the reliability of professional judgment is so poor as to be completely useless, it is proposed here that significant improvements in reliability can be obtained with relative ease.

Statistical techniques generate predictions in a perfectly reliable manner. Given the same data on repeated occasions, statistical equations generate identical decisions. It is precisely this consistency that accounts for the success of "judgmental bootstrapping" techniques, in which equations are used to make predictions based on statistical models of human judges (Camerer, 1981; Dawes, 1971, 1979; Dawes & Corrigan, 1974; Goldberg, 1965, 1970). The human judge possesses some skill for selecting the important cues to incorporate into predictions but is incapable of consistently applying any one prediction formula (Goldberg, 1986). The statistical model takes advantage of intuitive weighting of the different cues but predicts with perfect reliability. As will be demonstrated later, one major consequence of improved reliability is that the hit rate of decision makers will necessarily increase.

Redundancy

When deriving prediction equations, statistical techniques automatically assess the interrelationships among cues and account for any redundancies by reducing the weight of cues that provide little additional information, or incremental validity. Not only is it unlikely that human judges are capable of taking all of these interrelationships into account, but evidence suggests that judges may prefer redundant information because it bolsters confidence (Kahneman & Tversky, 1973; Slovic, 1966). Pieces of information collected in child welfare investigations may appear distinct yet be merely redundant with one another. For example, children's race and socioeconomic status may be highly correlated (Mandel, Lehman, & Yuille, 1995). Such correlations between case variables greatly increase the complexity and difficulty of the decision-making task.

Regression Effects

One special case of valid cue weighting is purely statistical in nature. Given that any cue is imperfectly correlated with the criterion, the predicted value should always be less extreme (in standardized units) than the cue value to maximize accuracy. For example, a child of parents that each have IQs of 150 cannot reasonably be expected to also have an IQ of 150 because there is an imperfect correlation between the intelligence of parents and their children. The actual expectation depends on the strength of the true correlation (in this case determined by a combination of a heritability coefficient and the reliability of IQ measurement) and is probably closer to 120 or 130. Statistical techniques incorporate the effects of this regression toward the mean into their decisions, whereas human judges ordinarily fail to make regressive predictions (Kahneman & Tversky, 1973; Nisbett & Ross, 1980).

ADHERENCE TO THE CLINICAL APPROACH

Despite an overwhelming preponderance of empirical evidence favoring statistical prediction and a lack of theoretical or logical support for clinical prediction, practitioners in many fields demonstrate an untenable adherence to the clinical approach to decision making. One survey, for example, found that a majority of neuropsychologists favor the clinical prediction of intellectual deficit over the use of statistical equations whose higher validity has repeatedly been demonstrated (Guilmette, Faust, Hart, & Arkes, 1990). Meehl (1986) attributes unwarranted adherence to the clinical approach to seven sources, each of which will be explored in the context of child welfare decisions.

Knowledge Deficit

Individuals are often not aware of the many biases that affect human judgment nor of the statistical alternatives available for decision-making tasks (see Faust, 1986, for an excellent discussion of the application of the judgment literature to clinical practice). Anyone even casually familiar with this literature surely understands the need for decision aids to supplement clinical judgment (Arkes, 1991; Kleinmuntz, 1990). Because little has appeared in the child welfare decision-making literature on statistical decision making, it is hoped that this article will alert child welfare workers to the existence and viability of the statistical approach to decision making (see Sicoly, 1989, however, for a discussion of some related technical issues).

Fear of Unemployment

Many individuals fear the threat of technological unemployment. They worry that an equation performing professional functions may displace some professionals, perhaps even themselves. This is an unfounded fear, particularly in social work domains in which researchers have noted many social constraints detrimental to decision making: a lack of necessary information, administrative pressures such as deadlines or excessive paperwork, high caseloads, fear of legal reprisal, and a lack of necessary training and experience (Murdach, 1994). The presence of severe constraints on workers' decisions highlights the critical importance of improving decision making efficiency. The employment of a statistical decision-making tool would do just that. Less time would be spent on activities at which case workers are relatively poor, such as combining information to reach decisions, freeing time for activities at which they are invaluable, such as collecting case information and providing services to children and families. A proper matching of skills to tasks would ensure that workers be provided the opportunity to accomplish realistic goals and provide valuable services.

Belief in the Efficacy of One's Own Judgment

The steadfast belief of many individuals in the validity of their own judgments is a frequent obstacle to the employment of statistical techniques. The persistence of a "controversy" over statistical versus clinical prediction is no doubt due largely to an unwillingness to admit that equations can perform some professional functions better than human judges. Decision makers commonly point to their training and experience as support for the quality of their decisions. Considerable research has been conducted to determine whether individuals with proper, formal training and extensive experience with a decision task can compete favorably with statistical equations. In fact, Meehl (1954, 1959, 1967) proposed that certain task characteristics, most notably configural (curvilinear or interactive) relationships between variables, would favor the highly trained and experienced, or "configural," judge. He maintained that complex decision-making tasks required human judges capable of such complex information integration. Contrary to these hopes and expectations, the search for such configural judges has been in vain (Goldberg, 1991; Slovic & Lichtenstein, 1971; Stewart, 1988). Substantial bodies of research have demonstrated that neither level of training nor amount of experience influences the quality of clinical judgment (Berman & Norton, 1985; Dawes, 1989, 1994; Faust, 1986; Faust & Ziskin, 1988; Garb, 1989; Goldberg, 1959; Oskamp, 1965; Ziskin & Faust, 1988). Appeals to formal training and clinical experience as support for human judgment are what Dawes (1994) refers to as "arguments from a vacuum," or unsubstantiated claims. In this case, the claims are in fact directly contradicted by empirical research.

Theoretical Identifications and Overconfidence

Another sizable body of research has demonstrated that in addition to predicting more poorly than statistical equations, human judges ordinarily hold greater confidence in their judgments than their validity merits (Dawes et al., 1989; Faust & Ziskin, 1988; Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff, & Phillips, 1982; Ziskin & Faust, 1988). This overconfidence has been observed across a diverse range of tasks and procedures, such as the estimation of uncertain quantities (Alpert & Raiffa, 1969) and the formation of clinical case-study judgments (Oskamp, 1965). Strong theoretical identifications and beliefs are a major contributor to overconfidence.

A decision maker's confidence is worthy of careful scrutiny because it is often mistakenly used as an estimate of validity. For example, research has found that court testimony may have a greater impact on a judge or jury when it is stated confidently, regardless of its actual validity (Faust & Ziskin, 1988; Ziskin & Faust, 1988). Fischhoff (1988) suggests that one appeal of the clinical approach to decision making is that it appears to be capable of producing perfect prediction. To significantly improve predictive validity by adopting a statistical decision-making policy, one must first accept that a fixed percentage of decisions will necessarily be wrong. Many practitioners seem unwilling to do this, ignoring in the process the unknown, but larger, degree of error inherent in their own judgments. In this sense, professionals' overconfidence may arise from nothing more substantial than pure wishful thinking (Dawes, 1991, 1994). Unfortunately, "the simplest principle is that past behavior is the best predictor of future behavior and it isn't very good" (Dawes, 1991, p. 259). Wishful thinking may inflate the perceived quality of professional decisions, but it only serves to hurt the people whose lives are affected by these decisions.

The Dehumanizing Feel of Statistical Equations

Many individuals erroneously believe that the statistical integration of information denies the uniqueness of individuals. In truth, a statistical equation can be constructed to use the very same information available to a clinical decision maker. Therefore, the difference between the two approaches has little to do with the uniqueness of individuals but has a great deal to do with combining information in the most reliable and valid manner possible. Moreover, a statistical equation is a dynamic decision-making policy open to public scrutiny and continual revision, refinement, and improvement. In comparison, human judgment is relatively static, privately held, potentially biased, and—as the research suggests—unlikely to improve much with training or experience. Therefore, the uneasy feeling that the use of validated decision models is somehow dehumanizing—a belief closely linked to the mistaken conception of ethics described later—can be most charitably labeled as misinformed.

Fear of Computers

Because statistical decision making can readily be implemented on computerized systems, practitioners who are relatively unfamiliar or inexperienced with computers tend to cling to the clinical approach to prediction. Because computers become more prevalent in the workplace with every passing day, an understanding of their use has become a virtual prerequisite for many occupations. Indeed, it is difficult to imagine performing many office tasks without the use of a computer. Some modern computer programs have become so sophisticated and user friendly that one now risks the pitfall of relying on computerized techniques owing only to their simplicity. However, relying on newly developed, user-friendly, computerized statistical decision models of undemonstrated validity is every bit as foolish as relying on clinical judgment of undemonstrated (or surpassed) validity (Matarazzo, 1986). It is important to keep in mind that it is the empirical validity of a decision-making policy, and not simply its ease of use, that ensures its practical utility and ethical acceptability.

Conception of Ethics

A disturbing conception of ethics is frequently advanced by opponents of statistical prediction. Due perhaps in part to several of the factors already discussed, statistical decision making is often viewed by these opponents as unethical. Examined carefully, however, this position contains a serious contradiction. Few would disagree, for example, that the most ethically defensible course of action is that which has the highest demonstrated validity—the best track record of correct decisions with the fewest mistakes. To use a decision-making policy known to systematically err more often than a statistical equation (or some as yet undiscovered approach that surpasses even statistical models) is grossly unfair to those individuals being served by the system. Meehl (1986) has addressed this issue in the following way:

> If I try to forecast something important about a college student, or a criminal, or a depressed patient by inefficient rather than efficient means, meanwhile charging this person or the taxpayer 10 times as much money as I would need to achieve greater predictive accuracy, that is not a sound ethical practice. That it feels better, warmer, and cuddlier to me as predictor is a shabby excuse indeed. (p. 374)

THE RANGE OF STATISTICAL ALTERNATIVES

The evidence in favor of statistical over clinical decision making is robust in another sense: Several different types of linear equations can be used to outpredict human judges. Not only will equations that use sample-specific information optimally outpredict human judges but so will equations that preserve only the direction of relationships—positive or negative—and weight the predictors, or cues, either equally (unit weighting), randomly (random weighting), or based on a model of an individual's own judgments (Dawes, 1979; Wainer, 1976). The superiority of unit-weighted models over human judges prompted Dawes and Corrigan (1974) to conclude that "the whole trick is to know what variables to look at and then to know how to add" (p. 105). The ability of even random-weighted models to outpredict human judges speaks strongly against the predictive validity of the clinical approach.

Of the "improper" (nonoptimal) linear models described above, one particular model is worthy of special note. An equation can be derived that closely models the predictions of a human judge; such an equation is called a "paramorphic representation" of that judge (Hoffman, 1960). This equation is representative of the manner in which this particular judge processes information. The technique of paramorphic representation was originally developed to learn from the judgments of experts. That is, it allowed novices in various prediction tasks to study the informational weighting processes of the most experienced and competent professionals in a field.

Although the technique served exceedingly well in modeling predictions (e.g., Goldberg, 1968; Oskamp, 1967), an unexpected result changed the course of its use: Predictions generated by a model were superior to those of the judge on whom it was based (Camerer, 1981; Dawes, 1971, 1979; Dawes & Corrigan, 1974; Goldberg, 1965, 1970). It was thus discovered that an equation can be gainfully substituted for the human judge from whom it was derived, a technique referred to both as Goldberg's Paradox, after its discoverer, and as judgmental bootstrapping.

This effect is due primarily to the perfect reliability of statistical decisions. Some mixture of research and practical experience typically suggests to an individual what information to include in a clinical prediction strategy, and the judge-modeled equation capitalizes on this information. However, the statistical equation can then profitably be substituted for the original judge because it bases its decisions on exactly this information but with perfect reliability. The practical value of this finding can hardly be overstated: Even in cases in which clinical decision makers are experienced experts, the statistical approach can nonetheless improve on their results.

IMPROVING DECISION-MAKING POLICIES

Let us return now to the formulation of a policy for child placement decisions. Your initial solution to the problem of decision making most likely involved an evaluation of the particular set of circumstances found in each case file, weighing factors that favored or opposed placement outside of the home as appropriate on a case-by-case basis. All in all, you would no doubt have espoused a policy founded on the clinical approach to decision making. This solution is still a common one, as evidenced by presentations at the 1997 meeting of research grantees of the National Center on Child Abuse and Neglect. In a panel discussion of child welfare decision making, for example, it was apparent that the statistical approach to decision making has not gained a foothold with practitioners. Not surprisingly, two researchers remarked on the unacceptably low agreement levels between expert case workers on risk assessment in child maltreatment decisions obtained in their studies (Baird, 1997; Schuerman, 1997).

As viable alternatives to problematic clinical techniques, consider instead methods founded on the statistical approach to decision making. Various authors have made recommendations to eliminate subjectivity through the development of more systematic instruments (Berger et al., 1989), professional training (Miller et al., 1993), or computerization of the decision-making process (Schuerman & Vogel, 1986; Schwab, Bruce, & McRoy, 1985, 1986; Schwab & Wilson, 1989; Sicoly, 1989; Wick & Schoech, 1988). A fruitful, three-stage, iterative approach to the construction of statistical models has been outlined by Schwab and Wilson (1989). First, define the problem as concretely and objectively as possible and analyze statistical relationships. Second, assess the validity and credibility of the model. Third, implement the model in usable form. Expanding on this approach, Sicoly (1989) noted the necessity of cross-validating classification models to prevent capitalization on chance by basing important decisions on the results obtained in one idiosyncratic sample of workers or cases. Once cross-validated,

> the statistical model that is developed simply represents the collective expertise of many workers and their experience with many cases. The process can be thought of as a computerized case conference that draws upon a vast, accumulated pool of knowledge. (Sicoly, 1989, pp. 53-54)

In keeping with the define, cross-validate, implement scheme described above, judgmental bootstrapping and statistical prediction represent steps in the continuous evolution of a sound decision policy. Imagine, once again, that you are a child welfare worker trying to determine how best to make placement decisions.

Judgmental Bootstrapping

We have seen that the largest impediment to valid child welfare decision making is the poor reliability of clinical judgment. To overcome this obstacle, you could derive an equation that models your own judgment and subsequently base your decisions on the results arising from this formula. By simply removing from your decision policy the inconsistency inherent in human judgment, improvements would be virtually guaranteed. After a small initial investment of time to generate your statistical model, you would save considerable amounts of time by applying it to the child welfare decisions that you are asked to make. This recovered time can be spent refining the statistical decision model, with the eventual goal of a statistical prediction system validated on outcome data; providing services to children and families; or performing other useful activities. More important, there would be an increased likelihood of positive outcomes for the children and families whose lives you affect through your placement decisions.

Methodology. Practically speaking, the development of an equation would require little effort. You would need to keep records of the information available for each case that you process as well as your decision on each case. Although not strictly necessary, standardization of the interview process and the use of a computerized spreadsheet (e.g., Lotus, Excel) or database (e.g., Dbase, Access) program would greatly facilitate the collection of information as well as its subsequent coding and integration. Once you have collected data on a sufficiently large number of cases, you need to quantify it. For continuous variables (e.g., child age, the duration and number of abuse incidents), this is straightforward. For categorical variables (e.g., child sex and race, the relationship of an alleged perpetrator to the child, your placement decision), you simply code each category as a unique number. Having done this, you are ready to generate your statistical model.

Deriving a statistical model of your decisions generally involves nothing more than ordinary least squares or logistic regression procedures. The technical details involved in deriving this statistical decision-making equation are not presented here; interested readers are referred to more comprehensive treatments of regression analysis (e.g., Cohen & Cohen, 1983; Pedhazur, 1982) and its application to human judgment and decision making (e.g., Cooksey, 1996; Dawes, 1979; Dawes & Corrigan, 1974; Stewart, 1988). You could take advantage of any major statistical software package (such as SPSS, SAS, or BMDP) to per-

(a) Validity = 1.00
Hit rate = (29 + 69)/100 = 98%

		OPTIMAL OUTCOME: Place into care	OPTIMAL OUTCOME: Keep at home	
ACTUAL DECISION	Place into care	29	2	31
	Keep at home	0	69	69
		29	71	

(b) Validity = .40
Hit rate = (17 + 57)/100 = 74%

		OPTIMAL OUTCOME: Place into care	OPTIMAL OUTCOME: Keep at home	
ACTUAL DECISION	Place into care	17	14	31
	Keep at home	12	57	69
		29	71	

(c) Validity = .60
Hit rate = (21 + 61)/100 = 82%

		OPTIMAL OUTCOME: Place into care	OPTIMAL OUTCOME: Keep at home	
ACTUAL DECISION	Place into care	21	10	31
	Keep at home	8	61	69
		29	71	

FIGURE 2: Classification Tables for Four Different Child Placement Decision Policies With Perfect Reliability: (a) Perfect Validity, (b) Moderate Validity, (c) Better Than Moderate Validity

NOTE: Note that approximately 30% of all cases are placed into care and that placement into care is the optimal outcome for approximately 30% of all cases. These marginal totals, based on Lindsey's (1992) example, are held constant across tables.

form a wide range of regression analyses. Alternately, you could hire a statistical consultant to clarify the procedures involved in setting up and running the analysis, to perform the analysis, and/or to aid in interpreting and implementing the output. Although the amount of time required to perform the appropriate analysis would depend on several factors, such as your statistical background, familiarity with computerized data analysis, and access to software and consultants as needed, it is unlikely to extend beyond a few hours or so. This statistical equation becomes your decision policy for future cases.

Decision-making improvement. If applied with as little discretionary judgment as possible, this policy will approach perfect reliability. Recall that the hit rate of the clinical approach to decision making, assuming moderate validity (two fifths of the difference between the zero-validity and the perfect-validity hit rates) and a reliability estimate of .25, was 64%. The hit rate of the bootstrapped decision policy can now be derived in parallel fashion.

Figure 2a depicts the perfect-validity classification results for a perfectly reliable, bootstrapped decision policy: The hit rate increases from the 72% level obtained with a constrained reliability to 98%. The results for zero-validity classification remain at a hit rate of 58% at all levels of reliability. Assuming, as before, that two fifths of this difference could be attained in practice, the hit rate of the bootstrapped decision policy increases from 64% to 74% (see Figure 2b). The 10 correct decisions in excess of those obtained through clinical prediction represent a 28% reduction in the original error rate of 36 cases out of 100, a substantial improvement by any standard.

Statistical Prediction

Once you have begun saving time through the use of judgmental bootstrapping, you can begin to develop and refine a statistical prediction model. The formal decision-making time that you conserve would become available to collect follow-up data on cases. When a large enough set of cases with follow-up data is available, analyses can be conducted to determine the predictive validity of case information using a range of potential criterion measures (and composites of those measures). Those models with the highest validity can then be tested against one another in cross-validation studies based on the continued collection of follow-up data. This process can—and ideally should—continue indefinitely, with refinement of the statistical model being an ongoing endeavor. Figure 3 graphically outlines this iterative approach.

Once you reach the point at which you become satisfied that the equation predicts outcomes well, you can begin using it in daily decision-making tasks. Whereas judgmental bootstrapping improves hit rates purely through an increase in reliability, statistical prediction takes advantage of all the desirable mathematical properties described earlier to increase the rate of correct decisions: weighting information according to its validity, aligning the metrics of all variables, increasing reliability, accounting for predictor redundancy, and adjusting for regression effects. Such a versatile decision-making policy would increase efficiency and, even more important, substantially increase the likelihood of positive outcomes for the children and families that you serve.

Methodology. There are two critical differences between this method and that of judgmental bootstrapping. The first difference is that, as noted above, follow-up data collection is required for the development and continued testing of statistical prediction

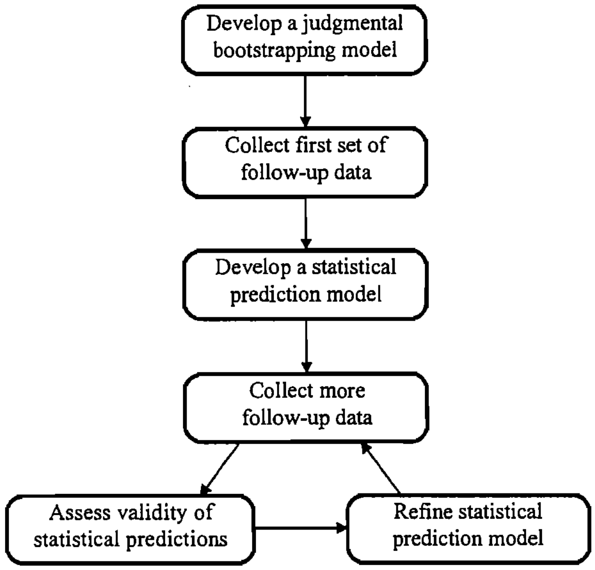

FIGURE 3: Schematic Representation of the Development and Continual Refinement of a Statistical Prediction System

models. As a cautious practitioner, you should continue the process of follow-up data collection to validate the derived equation on new sets of cases. It would be unwise to implement a decision policy based only on one set of cases because the statistical techniques used to derive equations can capitalize on chance and produce somewhat unstable weights, particularly if the original sample of cases was small. Tests of these weights on new samples of cases will allow you to assess their stability and to continually refine them as the decision policy evolves and matures.

Whereas this first methodological difference between judgmental bootstrapping and statistical prediction is practical in nature, demanding time and resources, the second difference stems from a thorny theoretical issue: It becomes necessary to define and assess an appropriate standard of child placement success. A statistical model is derived to predict a concrete outcome (criterion) measure, and the question of how to define and assess this is the most complex, difficult, and important problem posed by statistical prediction. In the case of child placement outcomes, there may be many factors contributing to your assessment of their success (e.g., physical, mental, and emotion status of the child and other involved parties; satisfaction of the child and the family; stability of the established relationships; postdecision incidence of abuse and neglect). This information may be difficult to obtain; follow-up data collection is not the norm in child welfare services. Once again, a more effective allocation of resources would dictate that agencies commit their time and money toward the collection and evaluation of follow-up information rather than toward making decisions based on the clinical method of prediction.

It is important to note, however, that the goal of statistical prediction is the same as that of clinical prediction: to make decisions that optimize outcomes. The complications that arise in defining and assessing an appropriate criterion are present regardless of which approach you employ. Caseworkers are using some criterion, however ill defined and idiosyncratic it may be, and commentators on the poor reliability of placement decisions have repeatedly called for attempts to bring the criterion into sharper focus. Whereas adherents of the clinical approach need not explicitly handle this problem, those favoring the statistical approach are forced to attack it head-on by striving to predict the most appropriate, well-specified criterion. Although there certainly is no easy way around this problem, it is safe to say that any effort to concretely define and assess the success of placement decision outcomes is likely to aid in their prediction, yielding results superior to those achieved through the prediction of implicit, vaguely specified outcomes that are not obtained through follow-up measures.

Decision-making improvement. Although this technique is more challenging to implement than judgmental bootstrapping, its results can be striking. Let us return one last time to the opening scenario to examine the impact of a statistical prediction technique on the hit rate of your decision policy. Whereas it has thus far been assumed that the validity obtained in practice is moderate, statistical prediction will necessarily improve this figure. To consider a very modest improvement, suppose that three fifths of the difference can be obtained with statistical prediction rather than the two fifths previously assumed. The hit rate now increases from 74% to 82% (see Figure 2c). This represents a 31% reduction in the error rate of judgmental bootstrapping or a full 50% reduction in the error rate of clinical prediction.

SUMMARY AND CONCLUSION

Workers in the field of child maltreatment are required to make many complex and far-reaching decisions each day. As Mandel et al. (1995) have noted, child abuse and neglect

> investigators must operate in a complex and often ambiguous decision-making environment that poses

> multiple trade-offs and considerations. They face the arduous task of balancing their roles as farsighted planners and myopic doers so that in the end they emerge as effective practical reasoners. (p. 920)

The present method for making these decisions—clinical prediction—has yielded unsatisfactory results, primarily in terms of the severely constrained reliability of child welfare workers and the questionable validity of their decision policies. These psychometric constraints place severe limits on the number of correct case decisions that child welfare workers can make. It is highly doubtful that additional training or experience with clinical prediction will lead to any practical improvement. Decision makers are clearly in need of tools to increase their predictive efficacy, a need made all the more pressing when considering the high stakes involved in child placement. Child welfare workers, the children and families that they serve directly, and the taxpayers who support these agencies demand and deserve the most effective services available. A system fundamentally based on clinical decision making cannot offer such services.

A shift to the more ethically defensible statistical mode of prediction would confer several noteworthy benefits. Statistical techniques completely eliminate reliability problems while potentially boosting validity, thereby improving service for children and their families through an increased hit rate for placement decisions. Furthermore, these techniques would profoundly transform the job of child welfare case workers by streamlining the decision-making process. Social and institutional pressures would become more manageable if a worker's time was used more efficiently. Moreover, all parties involved in the child welfare system would reap the benefits as workers spend their time performing functions for which they are well-suited, such as providing services to children and families, performing investigations and collecting information, or conducting empirical research to better understand the complex interrelationships of factors influencing child welfare, rather than functions for which they are not well-suited, such as clinically generating predictions that influence momentous decisions. The tools of statistical decision making are available to anyone who wishes to provide more effective and efficient services in the field of child welfare.

APPENDIX

Availability Heuristic

This demonstration is taken from Plous (1993).

> For each pair, circle the cause of death that is *most common* in the United States:
> 1. (a) Diabetes or (b) Homicide
> 2. (a) Car accidents or (b) Stomach cancer
> 3. (a) Tornado or (b) Lightning
> 4. (a) Falling airplane parts or (b) Shark attack

Due to disproportionate media coverage, some types of events seem more common than they actually are, whereas others seem less common than they actually are. You may be surprised to learn that the correct answers to these four comparisons are, respectively, (a), (b), (b), (a).

Representativeness Heuristic

This demonstration is taken from Tversky and Kahneman (1982).

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations. Which of the following is the *most likely* alternative:
>
> a. Linda is a bank teller
>
> b. Linda is a bank teller and is active in the feminist movement

Due to a perceived similarity of Linda's personality with the prototypical feminist, you may be surprised to learn that the correct answer is in fact (a). This alternative includes all bank tellers, regardless of whether they happen to be active in the feminist movement or not, whereas alternative (b) is a subset excluding all bank tellers that are not active in the feminist movement.

Anchoring Heuristic

Consider these two questions:

> 1. How many people must be present in order to be 50% certain that at least two of them share a birthday?
> 2. How many people must be present in order to be 50% certain that at least one of them was born on July 4th?

As a first approximation for each question, 50% certainty seems to involve half of the days in the year, or a starting point of about 180. Most people make an adjustment that lies in the proper direction (i.e., either greater than or less than 180) but is of insufficient magnitude (i.e., not far enough from 180). You may be surprised to learn that the answer to the first question is as low as 23 and that the answer to the second question is as high as 254.

REFERENCES

Alpert, W., & Raiffa, H. (1969). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Alter, C. F. (1985). Decision-making factors in cases of child neglect. *Child Welfare, 64,* 99-111.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486-498.

Baird, C. (1997, March). *Comparative study of the use and effectiveness of different risk-assessment models in CPS decision-making.* Paper presented at the Research Grantees' meeting of the National Center on Child Abuse and Neglect, Bethesda, MD.

Berger, D. K., Rolon, Y., Sachs, J., & Wilson, B. (1989). Child abuse and neglect: An instrument to assist with case referral decision-making. *Health and Social Work, 14*, 60-73.

Berman, J. S., & Norton, N. C. (1985). Does professional training make a therapist more effective? *Psychological Bulletin, 98*, 401-407.

Camerer, C. (1981). General conditions for the success of bootstrapping models. *Organizational Behavior and Human Performance, 27*, 411-422.

Chapman, L. J., & Chapman, J. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology, 72*, 193-204.

Chapman, L. J., & Chapman, J. (1971, November). Test results are what you think they are. *Psychology Today*, 18-22, 106-110.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

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

Corby, B., & Mills, C. (1986). Child abuse: Risks and resources. *British Journal of Social Work, 16*, 531-542.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90*, 272-292.

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist, 26*, 180-188.

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

Dawes, R. M. (1989). Experience and validity of clinical judgment: The illusory correlation. *Behavioral Sciences and the Law, 7*, 457-467.

Dawes, R. M. (1991). Probabilistic versus causal thinking. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1). Minneapolis: University of Minnesota Press.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York: Free Press.

Dawes, R. M. (1995). The nature of human nature: An empirical case for withholding judgment—perhaps indefinitely. *Political Psychology, 16*, 81-97.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95-106.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.

Dawes, R. M., Faust, D., & Meehl, P. E. (1993). Statistical prediction versus clinical prediction: Improving what works. In G. Keren & C. Lewis (Eds.), *Handbook for data analysis in the behavioral sciences: Methodological issues.* Hillsdale, NJ: Lawrence Erlbaum.

DePanfilis, D., & Scannapieco, M. (1994). Assessing the safety of children at risk of maltreatment: Decision-making models. *Child Welfare, 73*, 229-245.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50*, 387-395.

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*, 420-430.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241*, 31-35.

Fischhoff, B. (1988). Judgment and decision making. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought.* Cambridge, UK: Cambridge University Press.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387-396.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reasoning in everyday life.* New York: Free Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295-314.

Gleeson, J. P. (1987). Implementing structured decision-making procedures at child welfare intake. *Child Welfare, 66*, 101-112.

Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt test. *Journal of Consulting Psychology, 23*, 25-33.

Goldberg, L. R. (1965). Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs, 79*, 1-28.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23*, 483-496.

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73*, 422-432.

Goldberg, L. R. (1986). Some informal explorations and ruminations about graphology. In B. Nevo (Ed.), *Scientific aspects of graphology.* Springfield, IL: Charles C. Thomas.

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1). Minneapolis: University of Minnesota Press.

Guilmette, T. J., Faust, D., Hart, K., & Arkes, H. R. (1990). A national survey of psychologists who offer neurological services. *Archives of Clinical Neuropsychology, 5*, 373-392.

Haskett, M. E., Wayland, K., Hutcheson, J. S., & Tavana, T. (1995). Substantiation of sexual abuse allegations: Factors involved in the decision-making process. *Journal of Child Sexual Abuse, 4*, 19-47.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57*, 116-131.

Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Jones, L. (1993). Decision-making in child welfare: A critical review of the literature. *Child and Adolescent Social Work Journal, 10*, 241-262.

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

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

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

Kleinmuntz, B. (1990). Why we still use our heads instead of the formulas: Toward an integrative approach. *Psychological Bulletin, 107*, 296-310.

Kurtz, R. M., & Garfield, S. L. (1978). Illusory correlation: A further exploration of Chapman's paradigm. *Journal of Consulting and Clinical Psychology, 46*, 1009-1015.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 26*, 149-171.

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.* New York: Cambridge University Press.

Lindsey, D. (1991). Factors affecting the foster care placement decision: An analysis of national survey data. *American Journal of Orthopsychiatry, 61*, 272-281.

Lindsey, D. (1992). Reliability of the foster care placement decision: A review. *Research on Social Work Practice, 2*, 65-80.

Mandel, D. R., Lehman, D. R., & Yuille, J. C. (1994). Should this child be removed from home? Hypothesis generation and information seeking as predictors of case decisions. *Child Abuse and Neglect, 18*, 1051-1062.

Mandel, D. R., Lehman, D. R., & Yuille, J. C. (1995). Reasoning about the removal of a child from home: A comparison of police officers and social workers. *Journal of Applied Social Psychology, 25*, 906-921.

Matarazzo, J. D. (1986). Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma. *American Psychologist, 41*, 14-24.

Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4,* 268-273.
Meehl, P. E. (1959). A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *Journal of Counseling Psychology, 6,* 102-109.
Meehl, P. E. (1967). What can the clinician do well? In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment.* New York: McGraw-Hill.
Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370-375.
Meehl, P. E. (1992). Cliometric metatheory: The actuarial approach to empirical, history-based philosophy of science. *Psychological Reports, 71,* 339-467.
Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194-216.
Miller, L. B., Fisher, T., & Sinclair, I. (1993). Decisions to register children as at risk of abuse. *Social Work and Social Sciences Review, 4,* 101-118.
Murdach, A. D. (1994). Avoiding errors in clinical prediction. *Social Work, 39,* 381-386.
Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. (1976). Popular induction: Information is not necessarily informative. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, NJ: Lawrence Erlbaum.
Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.
Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29,* 261-265.
Oskamp, S. (1967). Clinical judgment from the MMPI: Simple or complex? *Journal of Clinical Psychology, 23,* 411-415.
Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). Fort Worth, TX: Holt, Rinehart & Winston.
Plous, S. (1993). *The psychology of judgment and decision making.* Philadelphia: Temple University Press.
Quinsey, V. L., & Maguire, A. (1986). Maximum security psychiatric patients: Actuarial and clinical prediction of dangerousness. *Journal of Interpersonal Violence, 1,* 143-171.
Sarbin, T. R. (1944). The logic of prediction in psychology. *Psychological Review, 51,* 210-228.
Sarbin, T. R. (1986). Prediction and clinical inference: Forty years later. *Journal of Personality Assessment, 50,* 362-369.
Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66,* 178-200.
Schuerman, J. (1997, March). *Understanding child maltreatment decisions and those who make them.* Paper presented at the Research Grantees' meeting of the National Center on Child Abuse and Neglect, Bethesda, MD.
Schuerman, J. R., & Vogel, L. H. (1986). Computer support of placement planning: The use of expert systems in child welfare. *Child Welfare, 65,* 531-543.
Schwab, A. J., Bruce, M. E., & McRoy, R. G. (1985). A statistical model of child placement decisions. *Social Work Research and Abstracts, 21,* 28-34.
Schwab, A. J., Bruce, M. E., & McRoy, R. G. (1986). Using computer technology in child placement decisions. *Social Casework: The Journal of Contemporary Social Work, 67,* 359-368.
Schwab, A. J., & Wilson, S. S. (1989). The continuum of care system: Decision support for practitioners. *Computers in Human Services, 4,* 123-140.
Sicoly, F. (1989). Prediction and decision making in child welfare. *Computers in Human Services, 5,* 43-56.
Slovic, P. (1966). Cue-consistency and cue-utilization in judgment. *American Journal of Psychology, 79,* 427-434.
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.
Slovic, P., & MacPhillamy, D. (1974). Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance, 11,* 172-194.
Spence, D. P. (1982). *Narrative truth and historical truth.* New York: Norton.
Stein, T. J., & Rzepnicki, T. (1983). *Decision making at child welfare intake: A handbook for practitioners.* New York: Child Welfare League of America.
Stewart, T. R. (1988). Judgment analysis: Procedures. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view.* Amsterdam: North-Holland Elsevier.
Turk, D. C., & Salovey, P. (1988). *Reasoning, inference, and judgment in clinical psychology.* New York: Free Press.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207-232.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.
Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.
Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371-384.
Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83,* 213-217.
Wick, J., & Schoech, D. (1988). Computerizing protective services intake expertise: Preliminary research. *Children and Youth Services Review, 10,* 233-252.
Wiggins, J. S. (1981). Clinical and statistical prediction: Where are we and where do we go from here? *Clinical Psychology Review, 1,* 3-18.
Wormith, J. S., & Goldstone, C. S. (1984). The clinical and statistical prediction of recidivism. *Criminal Justice and Behavior, 11,* 3-34.
Ziskin, J., & Faust, D. (1988). *Coping with psychiatric and psychological testimony.* Venice, CA: Law and Psychology Press.

John Ruscio is in the Department of Psychology at Brandeis University.