Journal of Consulting and Clinical Psychology
2000, Vol. 68, No. 1, 145–154

Copyright 2000 by the American Psychological Association, Inc.
0022-006X/00/$5.00 DOI: 10.1037//0022-006X.68.1.145

# The Role of Complex Thought in Clinical Prediction: Social Accountability and the Need for Cognition

John Ruscio
Brandeis University

Research shows that clinical predictions are less accurate than statistical predictions and are held with unreasonable confidence. Because there are obstacles to the implementation of statistical prediction, factors that improve clinical judgment must be identified. One hundred twelve individuals participated in an experiment investigating the role of complex thought in clinical prediction. Results revealed marked performance differences related to the amount of available clinical information. Participants' assessed need for cognition was associated with their consistency, accuracy, and cue-weighting strategies. Social accountability improved confidence performance under certain task conditions but was unrelated to accuracy. Theoretical and practical implications of these results are discussed, with emphasis on the restructuring of tasks and the selection and training of human forecasters to promote accurate and appropriately confident clinical predictions.

Prediction is an integral part of decision making in people's everyday personal and professional lives. Many predictions have significant implications for important clinical outcomes, such as predictions of future violence in patients with mental illnesses (Gardner, Lidz, Mulvey, & Shaw, 1996; Werner, Rose, & Yesavage, 1983), predictions of whether abused or neglected children will fare better if removed from their homes and placed into foster care (Ruscio, 1998b), or predictions of the level of care in clinical settings (e.g., Bickman, Karver, & Schut, 1997). In addition to the necessity of accurate predictions, it is often not only desirable but of tremendous practical importance to require an additional type of judgment quality: appropriateness of confidence.[1] Whereas a considerable body of research has examined the accuracy of clinical predictions, the confidence held in these predictions has been studied far less intensively.

The judgment process used to generate predictions can be entirely intuitive (the "clinical" approach) or aided to some extent by a statistical equation (the "statistical" or "actuarial" approach). Applying the clinical approach, a human judge evaluates available information and arrives at a prediction. This method does not necessarily involve a professional clinician; rather, the term is broadly applied whenever a human judge forecasts outcomes. The statistical approach involves an algorithmic, mechanical combination of information that maximizes the accuracy of predictions.

Although there is a long-standing debate regarding the relative efficacy of these two approaches, a substantial body of research has indicated that although human judges occasionally make predictions equal to those made by statistical formulas, they typically do worse (for reviews of the evidence, see Dawes, Faust, & Meehl, 1989; Meehl, 1954; and Sawyer, 1966). Moreover, substantial bodies of research have demonstrated that neither level of training nor degree of experience influences the quality of clinical judgments (Berman & Norton, 1985; Dawes, 1994; Faust, 1986; Faust & Ziskin, 1988; Garb, 1989; Goldberg, 1959). The superiority of statistical prediction has been attributed to two complementary sources: the desirable properties of statistical techniques and the undesirable cognitive biases of human judges (Goldberg, 1991). Compared with human judges, statistical equations are extremely effective in detecting relationships amidst considerable variation. Cognitive biases further broaden the gap in accuracy between statistical and clinical predictions.

In addition to predicting more poorly than statistical equations, human judges ordinarily hold greater confidence in their predictions than their accuracy levels merit (Alpert & Raiffa, 1969; Dawes et al., 1989; Faust & Ziskin, 1988; Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff, & Phillips, 1982; Oskamp, 1965). Inflated confidence is often mistakenly perceived by oneself and others as a gauge of accuracy. Individuals who strongly believe in the efficacy of their clinical judgment will be less likely to use decision aids, to keep abreast of research developments, and so forth. These overconfident individuals may also mislead others: 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). Although wishful thinking

Financial support for this dissertation research was provided by a grant from the Louis, Frances, and Jeffrey Sachar Fund and by the Brandeis University Psychology Department. Their support is gratefully acknowledged. I am indebted to Frank W. Weathers for providing the clinical data used as stimuli in this experiment. I also thank Theodore P. Cross, David Faust, Paul E. Meehl, Ayelet Meron Ruscio, Joan S. Tucker, and Leslie A. Zebrowitz for their insightful criticisms and suggestions at various stages of this project.

Correspondence concerning this article should be addressed to John Ruscio, who is now at the Department of Psychology, Elizabethtown College, Elizabethtown, Pennsylvania 17022. Electronic mail may be sent to rusciojp@acad.etown.edu.

[1] Because uniform overconfidence was anticipated (and confirmed by the data analysis), the term *confidence performance* was used in place of the more awkward *appropriateness of confidence* wherever possible. In this investigation, a higher score on the confidence performance measure indicated a more appropriate degree of confidence.

may inflate the perceived quality of judgments, this illusion holds little value. Accuracy and appropriateness of confidence remain the gold standards of judgmental quality. Still, despite considerable scientific evidence associating statistical methods with superior accuracy and appropriately calibrated confidence on judgment tasks, practitioners in many fields demonstrate an untenable adherence to the clinical approach to prediction (Grove & Meehl, 1996; Meehl, 1957, 1986).

Given the predominance of the clinical approach in practice, it is critical that we improve our understanding of factors that influence or improve clinical prediction, such as specific social contexts in which judgments are made and reliable individual differences in judgment abilities (Ruscio, 1998a). Previous attempts to identify such factors have found little evidence for improved decision making. Given the rarity of immediate, unambiguous, and accurate feedback in training and on-the-job experience, for example, it should come as no surprise that predictions fail to improve with training or experience (Ruscio, 1998c). However, although research has shown that judges' professional backgrounds are largely unrelated to predictive accuracy, there may yet be reliable sources of individual differences in decision-making ability. Furthermore, although little empirical work has examined aspects of real-world prediction situations (Tetlock, 1985b), there is reason to hypothesize that common social pressures may influence the quality of predictions. It is hypothesized that these unexplored social factors and potential individual differences fall partially under the rubric of complex thought and that the stimulation of complex thought may contain important clues for the improvement of clinical decision making.

Thought is complex to the extent that it involves high levels of effortful cognitive processes (Cacioppo & Petty, 1982). It is hypothesized that complex thought is composed of at least two broadly conceived components that jointly influence prediction. The first component entails the degree to which one considers a wide range of available information before reaching a judgment. An individual who considers a number of informational sources before predicting an outcome can be seen as engaging in more complex thought than an individual who considers fewer informational sources. This aspect of complex thought is referred to as the scope of the judgment process. The focus of the judgment process, on the other hand, consists of the degree to which an individual relies solely on the information most useful to prediction and discards the rest.

In describing successful prediction strategies, Dawes and Corrigan (1974) stated, "The whole trick is to know what variables to look at and then to know how to add" (p. 105). The success of statistical prediction may thus be ascribed to the procedure's broad scope and tight focus: Valid cues are identified through research, and these (and only these) cues are included in the prediction equation. Thus, the concoction of complex explanations based on wide arrays of information may in fact work to the detriment of human judges, and the stimulation of clinical judgment that is both broad in scope and narrow in focus may improve predictive accuracy (Meehl, 1954).

How, then, can we stimulate complex thought in the context of clinical decision making? Such thought may be stimulated through both situational factors and individual differences in cognitive style. One powerful, prevalent, and often neglected situational motive shown to lead to complex thought is social accountability (Tetlock, 1983a). Research conducted primarily by Tetlock and his colleagues indicates that we are responsive to the social importance of a task when making decisions. Accountability has been shown to improve recall of evidence and eliminate primacy effects in a legal decision-making task (Tetlock, 1983b), to reduce the overattribution effect in an essay-attribution paradigm (Tetlock, 1985a), to promote better calibrated confidence levels in a personality prediction task (Tetlock & Kim, 1987), and to attenuate the carryover of anger to attributions and decisions for punishment (Lerner, Goldberg, & Tetlock, 1998).

Despite these benefits, accountability seems to magnify the dilution effect, or the tendency to give weight to all available sources of information (Tetlock & Boettger, 1989). Studies have shown that accountable individuals weaken the quality of their judgments by underweighting diagnostic cues, giving weight to both nondiagnostic and diagnostic cues in a misguided effort to consider every piece of information, regardless of its relevance to the predicted outcome. Thus, research suggests that accountability only improves the scope of judgment, not its focus.

Complex thought is also motivated by the need for cognition, a propensity to engage in and enjoy complex cognitive processing (Cacioppo & Petty, 1982). A recent conceptual review indicates that individuals high in need for cognition engage in more effortful information-processing activities and are more willing to expend the necessary effort to overcome cognitive and motivational judgment biases than are individuals low in need for cognition (Cacioppo, Petty, Feinstein, & Jarvis, 1996). For example, high-need-for-cognition individuals are more responsive than low-need-for-cognition individuals to the quality of an argument (Cacioppo, Petty, & Morris, 1983) but are less responsive to the mere number of arguments in a message (Petty & Cacioppo, 1984).

Research regarding high-need-for-cognition individuals suggests that they use broad judgmental scope: They generate complex attributions for behavior (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986), pursue new experiences that stimulate thinking (Venkatraman, Marlino, Kardes, & Sklar, 1990), and exhibit greater intrinsic motivation for seeking and engaging in challenging activities (Amabile, Hill, Hennessey, & Tighe, 1994). Other research also suggests that these individuals use a narrow judgmental focus by devoting attentional processes exclusively to an ongoing cognitive task (Osberg, 1987) and forming beliefs on the basis of empirical information and rational considerations (Leary, Sheppard, McNeil, Jenkins, & Barnes, 1986).

In addition to the need for cognition and social accountability, a well-structured clinical decision-making task also serves to facilitate judgment. For example, individuals may be provided with few or many pieces of information—or cues—and each cue may be of high or low relevance to the clinical predictions. The number and relevance of available cues influences the way in which information is considered and combined to arrive at predictions, and highly complex thinkers will likely process this information differently than less complex thinkers.

The purpose of the present experiment was to identify reliable individual differences and social and task factors that influence the accuracy and confidence of clinical predictions. First, high need for cognition was expected to yield high predictive accuracy by broadening the scope and narrowing the focus of judgment, thus approximating the process of statistical decision making. This broadened scope may also result in improved confidence performance, as research suggests that greater consideration of alternatives is one of the best methods available to improve the calibration of confidence (Arkes, 1991; Koriat, Lichtenstein, & Fischhoff, 1980). Second, judges held socially accountable for their predictions were expected to exhibit improved appropriateness of confidence, given their increased awareness and consideration of a wide range of alternative explanations. Social accountability was expected to have little impact on accuracy, perhaps impairing it by increasing the scope of judgment without sufficiently narrowing its focus, thereby magnifying the dilution effect. Finally, a task structure featuring exclusively strong cues, or pieces of information highly relevant to the clinical decisions at hand, was expected to promote the accuracy of clinical prediction. The addition of weak cues was expected to increase dilution, consequently reducing accuracy. However, it was hypothesized that a moderate amount of dilution enhances predictive accuracy by offsetting insufficiently regressive clinical predictions (Kahneman & Tversky, 1973).

## Method

### *Design*

A three-way between-subjects factorial design, including one nonmanipulated and two manipulated factors, was used. Participants' total score on the Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984) served as the nonmanipulated factor. Participants were randomly assigned to an accountable or an unaccountable condition and one of four cue conditions: four strong cues (4S), two strong cues (2S), two strong and three weak cues (2S–3W), or two strong and six weak cues (2S–6W).

### *Participants*

One hundred twelve undergraduate students participated in the study, some in voluntary fulfillment of an introductory psychology course's experimental participation requirement and others in exchange for a payment of $10. Given the factorial design used and the typically high reliability of repeated measurements in a judgment analysis experiment of this type (Cooksey, 1996), this sample size afforded powerful statistical comparisons (Cohen, 1988).

### *Materials*

The 18-item short form of the Need for Cognition Scale was used to assess each participant's tendency to engage in and enjoy complex thought. This face-valid scale yields measurements with excellent psychometric properties. Estimates of internal consistency from several dozen independent investigations have ranged from .81 to .97 (see Cacioppo et al., 1996), with test–retest correlations of .88 over a 7-week period (Sadowski & Gulgoz, 1992) and .66 over an 8-month period (Verplanken, 1991). The construct validity of this scale has also been demonstrated (for a review, see Cacioppo et al., 1996).

The prediction task presented a decision-making situation with variables that were familiar and easily interpretable, using data collected for a study at the Boston Veterans Administration Medical Center. Cues were selected from a wide array of clinical data on the basis of several criteria. All cues, as well as the criterion, were continuous and normally distributed. All cues, but not the criterion, were on the same scale of measurement (*T* scores computed through the standard Minnesota Multiphasic Personality Inventory—2 [MMPI-2] scoring system; Hathaway & McKinley, 1989) and had approximately the same means and standard deviations. The strong cues (family problems, psychopathic deviate, anger, and criminality) were moderately correlated with the criterion, an Antisocial Behavior Inventory (ASBI; Weathers, 1992) score (*r*s ranged from .34 to .43). The weak cues (cynicism, depression, fears, health concerns, hypochondriasis, and obsessiveness) were only weakly correlated with the criterion (*r*s ranged from .11 to .15; see Table 1 for intercorrelations between the criterion and all cues).

### *Procedure*

Participants completed a number of background questions before they were introduced to the experimental task. These included questions about participants' age, sex, and the number of mathematics and statistics courses completed, as well as the short form of the Need for Cognition Scale described above.

Participants were then randomly assigned to one of the two accountability conditions. One half of the participants, composing the no-accountability condition, were told that their responses would be kept completely confidential and that even the experimenter would not be able to identify the responses of individual participants. The other half of the participants, composing the accountability condition, were told that the experimenter would conduct an audiotaped discussion subsequent to the prediction task, in which each participant would be required to justify

Table 1
*Correlations Between Criterion, Strong Cues, and Weak Cues*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. ASBI score[a] | — | | | | | | | | | | |
| 2. Psychopathic deviate[b] | .43 | — | | | | | | | | | |
| 3. Family problems[b] | .40 | .62 | — | | | | | | | | |
| 4. Anger[b] | .34 | .48 | .67 | — | | | | | | | |
| 5. Criminality[b] | .40 | .27 | .50 | .55 | — | | | | | | |
| 6. Depression[c] | .15 | .58 | .47 | .50 | .17 | — | | | | | |
| 7. Hypochondriasis[c] | .13 | .39 | .39 | .37 | .07 | .72 | — | | | | |
| 8. Obsessiveness[c] | .12 | .38 | .55 | .62 | .42 | .56 | .33 | — | | | |
| 9. Cynicism[c] | .11 | .27 | .61 | .62 | .66 | .31 | .27 | .59 | — | | |
| 10. Fear[c] | .15 | .38 | .41 | .35 | .21 | .33 | .21 | .52 | .34 | — | |
| 11. Health concerns[c] | .14 | .39 | .55 | .53 | .26 | .73 | .90 | .54 | .46 | .35 | — |

*Note.* Correlations above $r = .19$ are statistically significant at $\alpha = .05$; correlations above $r = .25$ are statistically significant at $\alpha = .01$. ASBI = Antisocial Behavior Inventory (Weathers, 1992).
[a] Criterion variable. [b] Strong cue. [c] Weak cue.

his or her judgment strategy in front of his or her peers.[2] Accountable participants were asked to sign a form consenting to have the discussion audiotaped for future data analysis; all participants signed this form. This accountability manipulation was modeled after a procedure that has been used extensively in previous research (Tetlock, 1983b, 1985a; Tetlock & Boettger, 1989; Tetlock & Kim, 1987).

Participants were told that their task was to predict how antisocial each of a series of men was likely to be. All the participants were given two strong cues, family problems and psychopathic deviate, with which to predict the criterion. The amount of additional information given to participants varied according to randomly assigned cue conditions. One group of participants was given two additional strong cues: anger and criminality.[3] The second group was given no additional information beyond the two original strong cues. The third group was given three relatively weak cues. These cues included either depression, hypochondriasis, and obsessiveness or cynicism, fears, and health concerns, with the particular set of weak cues determined randomly and counterbalanced across accountability and no-accountability conditions. The final group was given all six weak cues. Random assignment to the four cue conditions was counterbalanced across the accountability and no-accountability conditions. Previous research has suggested that participants can process up to eight cues with negligible fatigue (Cooksey, 1996).

The relative predictive strengths of the cues was made explicit to participants to circumvent problems stemming from inaccurate perceptions of cue validities. Strong cues were described as relatively good predictors of antisocial behavior, whereas all other cues were said to be routinely collected through clinical assessment in spite of their relatively poor ability to predict antisocial behavior (see the Appendix for a table summarizing cue data).

Participants sequentially judged a total of 100 cases of information,[4] taking breaks whenever needed. For each case, participants were asked to make a point prediction of ASBI scores, as well as to construct a 75% confidence interval around that prediction. The point predictions—ranging along the 0-to-19 scale of ASBI scores—were used to assess accuracy, whereas the confidence intervals were used to assess confidence performance. The 75% value for confidence intervals was selected after consideration of several factors. First, this value was relatively easy for participants to understand. They simply needed to construct intervals so that they were likely to be correct three-fourths of the time. Second, because variation around 75% was possible, this value was sensitive to both overconfidence and underconfidence. Third, the 75% value apportioned greater sensitivity to discriminating variations in overconfidence than in underconfidence. Given that underconfidence is relatively rare, this seemed appropriate.

After completing the prediction task, participants were asked several questions regarding their subjective cue consideration and use, as well as their understanding of and memory for the task instructions. These posttask questions were posed primarily to prepare accountable participants for subsequent discussion of their judgment strategies and not for analytic purposes.

Finally, after completion of the prediction task, participants in the accountability condition engaged in an audiotaped discussion with other participants from their session. With minimal prompting from the experimenter, each participant, in turn, described and attempted to justify his or her judgment strategy. This discussion was conducted both to convince future participants who might have heard about the experiment that they would be audiotaped and to uncover clinical prediction strategies that might inform data analysis. At the conclusion of the experimental session, all the participants were fully debriefed, thanked for their time, and excused.

## Results

### *Data Screening and Collapsing*

Several criteria were established for the inclusion of participants' responses in data analysis: Each participant had to make judgments using the full response scale, complete at least two thirds of the cases so that his or her judgments could be assessed and compared with those made by others, and show a reasonable level of judgmental consistency across predictions to demonstrate that he or she was taking the task seriously. Six participants' data were deemed unsuitable for subsequent analysis, resulting in a final sample size of $N = 106$.

Differences on the primary dependent measures of predictive accuracy and confidence, as well as on the three experimental variables of accountability condition, need for cognition, and cue condition, were examined across the background variables of sex, age, and number of mathematics and statistics courses completed to determine whether data could safely be combined for analysis. Only two relationships emerged: Need for cognition was correlated with age, $r(102) = .28, p = .004$, and with math background, $r(102) = .26, p = .008$. Students with a high need for cognition would be expected to remain in college at a higher rate and therefore have the opportunity to take more math classes, so these correlations served only to support the construct validity of the Need for Cognition Scale. Having uncovered no reason to analyze data separately across sex, age, or math background, I collapsed data for all subsequent analyses.

### *The Quality of Clinical Prediction*

Several analyses were conducted to determine the overall quality of clinical predictions. Judgmental consistency was determined by regressing the available cues on each participant's predictions and examining the $R$ value of this "judge model" regression equation (Cooksey, 1996). Consistency values were high, ranging from .62 to .99, with a median $R$ of .91 (see Table 2 for all ranges, medians, and statistical comparison values presented here).

Each participant's predictions were correlated with the criterion—corresponding ASBI scores—to determine the accuracy of prediction (or, in Brunswikian terms, the participant's "achievement"; Cooksey, 1996). These accuracy scores were generally high, ranging from .09 to .57, with a median $r$ of .40. Although high, clinical predictions ($r = .40$) were inferior to those of optimal statistical models ($r = .52$). In addition, clinical predictions were inferior to those of unit-weighted ($r = .48$) and cross-validated ($r = .43$) statistical models, neither of which capitalizes

[2] Past research found that only a preexposure accountability condition, in which participants were made aware of their accountability before the experimental manipulation(s), influenced the quality of subsequent judgments (Tetlock, 1983b, 1985a; Tetlock & Kim, 1987). Judgments made in a postexposure accountability condition, in which participants were made aware of their accountability after the experimental manipulation(s), did not differ from those of participants in a no-accountability condition. Therefore, only the preexposure accountability and no-accountability conditions are used in the present experiment.

[3] The name of this scale was changed from Antisocial Practices to reduce its surface similarity to the criterion (ASBI score). The *criminality* label reflects the key elements of the Antisocial Practices Scale as described in the scoring manual (Hathaway & McKinley, 1989, p. 43).

[4] A total of 123 cases were judged, with outcome feedback provided on the final 23 cases. Because of space limitations, the impact of this feedback is not discussed here, and all results are based on analyses involving the first 100 cases of information.

Table 2
*Lowest, Median, Highest, and Statistical Values for Dependent Variables*

| Variable | Lowest | *Mdn* | Highest | Statistical |
|---|---|---|---|---|
| Consistency | .62 | .91 | .99 | 1.00 |
| Accuracy | .09 | .40 | .57 | .52[a] |
| Confidence (%) | 14 | 38 | 62 | 75 |
| Mean width of CI | 1.49 | 3.34 | 5.39 | 9.79[b] |
| Accuracy–confidence correlation | −.24 | −.01 | .25 | .04[b] |
| Mean $\beta$ | | | | |
| Strong | 0.08 | 0.41 | 0.57 | 0.31[a] |
| Weak[c] | 0.01 | 0.10 | 0.25 | 0.12[a] |
| Standard deviation of predictions | 1.95 | 3.65 | 4.92 | NA |

*Note.* CI = confidence interval; NA = not applicable.
[a] Mean of optimal models. [b] Based on the simplest optimal model (two strong cues). [c] Cue conditions were two strong cues–three weak cues and two strong cues–six weak cues only.

on chance (Dawes, 1979), $t(105) > 5.97$, $p < .001$, for each comparison. This rank ordering of accuracy levels (optimal model > unit-weighted model > cross-validated model > clinical judgment), which is consistent with previous research, was also observed separately within each cue condition.

Participants were asked to construct a range of values for each case so that they were 75% confident that the criterion value fell within the range. The percentage of these intuitively derived confidence intervals that actually included the criterion value—henceforth referred to as each participant's confidence score—ranged from 14% to 62%, with a median of 38%. That every participant's confidence score fell far short of 75% strongly corroborates the expectation of widespread overconfidence. Intuitively derived confidence intervals were much too narrow. Calculated within participants, mean width of the confidence intervals ranged from 1.49 to 5.39 units along the ASBI scale (which runs from 0 to 19), with a median of 3.34. These intuitively derived intervals ($M$ = 3.39, $SD$ = 0.18) were much more narrow than their statistically derived counterparts ($M$ = 9.79, $SD$ = 0.08), $t(99) = 383.01$, $p < .001$. The narrow nature of intuitively derived intervals cannot simply be ascribed to the narrow sample interval provided in the task instructions (width = 4) because participants' intervals were more narrow than the sample as well, $t(99) = 33.29$, $p < .001$.

Within participants, the normatively appropriate link between confidence and accuracy was examined. The correlations across cases between the width of the interval and the (absolute) residual in prediction ranged from $r = -.24$ to $r = .25$, with a median of $r = -.01$. Thus, there did not appear to be any systematic relationship between accuracy and confidence.

Although no a priori hypothesis was made, these data permitted the examination of an interesting issue regarding the extremity of predictions. One feature of statistically derived confidence intervals is that they become wider as predictions deviate from the mean criterion value. A quadratic relationship emerged ($\beta = -0.25$, $p = .008$) after a linear relationship between the participants' (centered) predicted value and the mean width of corresponding confidence intervals across cases was controlled for. That is, more extreme predictions were surrounded by narrower confidence intervals, an intriguing reversal of the normatively correct state of affairs.

Cue usage was evaluated by examining both the weights assigned to strong and weak cues (in terms of the standardized regression coefficient $\beta$, the most widely accepted unit; Cooksey, 1996) and the variation in participants' predictions across cases. The mean weight assigned by participants to the strong cues ranged from .08 to .57, with a median $\beta$ of 0.41. The mean weight assigned by participants to the weak cues ranged from .01 to .25, with a median $\beta$ of 0.10. Overall, participants assigned cue weights that were similar to those assigned by statistical models.

Variation within each participant's predictions served as an index of dilution. The standard deviation of participants' predictions ranged from 1.95 to 4.92, with a median of 3.65. Standard deviations were correlated with judgmental consistency, $r(104) = .40$, $p < .001$, and accuracy, $r(104) = .27$, $p = .005$, but not with confidence scores, $r(103) = -.10$, *ns*. To test the hypothesis that accuracy peaks with a moderate amount of dilution, I tested a curvilinear trend. After I controlled for a linear relationship between accuracy and the (centered) standard deviation of predictions, a quadratic relationship emerged ($\beta = -0.20$, $p = .034$). Accuracy was in fact greatest when predictions were moderately diluted.

Finally, participants' prediction strategies were examined. Potential prediction strategies submitted for evaluation were conceived a priori and generated from participants' self-reports of their judgment processes. This resulted in a total of six possible strategies: (a) optimal weights assigned to all cues, (b) equal weights assigned to all cues, (c) equal weights assigned to strong cues only, (d) equal weights assigned to strong cues and their interaction, (e) weights assigned only to the highest cue value for each case, and (f) weights assigned only to the highest strong cue value for each case. Strategies involving low cue values were not considered because no participant spontaneously commented on their use. Predictions were computed from a linear equation representing each of the six strategies and then correlated with the predictions made by each participant to compute an index of fit. The best fit strategy for each participant was the one with predictions that correlated most highly with those of the participant. Table 3 shows the range of fit values for each strategy and identifies the strategies used most frequently in each cue condition.

### *Regression Analyses*

Hypothesis-testing analyses addressed differences on the dependent measures attributable to the primary experimental variables:

Table 3
*Best Fitting Strategies by Cue Condition*

| Cue condition | Optimal weight | Unit weight, all cues | Unit weight, strong cues | Unit weight + interaction | Highest cue | Highest strong cue |
|---|---|---|---|---|---|---|
| 4S | 0 | 23[a] | 23[a] | 0 | 3 | 0 |
| 2S | 14 | 11[a] | 11[a] | 1 | 1 | 0 |
| 2S–3W | 0 | 12 | 15 | 0 | 0 | 0 |
| 2S–6W | 0 | 7 | 19 | 0 | 0 | 0 |
| Lowest fit | .26 | .60 | .55 | .44 | .51 | .50 |
| Median fit | .76 | .83 | .88 | .76 | .75 | .81 |
| Highest fit | .98 | .98 | .99 | .89 | .95 | .91 |

*Note.* Values in the upper half of the table represent the number of participants' best fit by each strategy model within each cue condition. Values in the lower half of the table represent correlations between participants' predictions and those of each strategy's model. 4S = four strong cues; 2S = two strong cues; 2S–3W = two strong and three weak cues; 2S–6W = two strong and six weak cues.

[a] Only strong cues were present; therefore, the fit of the unit weight model for all cues was equal to the fit of the unit weight model for strong cues.

need for cognition, accountability, and cue condition. Regression analysis was used instead of analysis of variance (ANOVA) procedures to preserve the continuity of the need for cognition variable, thus avoiding median splits and other arbitrary grouping procedures that reduce statistical power (Cohen, 1983). The four cue conditions were represented by three contrast variables (Cohen & Cohen, 1983). The first contrast compared the two conditions containing exclusively strong cues (4S and 2S, coded as 1) with the two conditions that contained additional weak cues (2S–3W and 2S–6W, coded as −1). The second contrast compared cue conditions 4S (coded as 1) and 2S (coded as −1), and the third contrast compared cue conditions 2S–3W (coded as 1) and 2S–6W (coded as −1).

Hierarchical multiple regression analyses were performed using the three experimental variables, with cue conditions entered through the three contrasts, followed by all possible interaction terms. All three experimental variables were entered simultaneously on the first step of each hierarchical regression analysis, with their interactions entered on subsequent steps (3 two-way interactions on Step 2, 1 three-way interaction on Step 3). Because of the sheer number of predictors stemming from a three-factor analysis, those that did not even marginally predict the dependent variable were removed from the model in a trimmed regression analysis.

*Need for cognition.* Individuals high in need for cognition were hypothesized to achieve greater accuracy levels, and perhaps superior confidence performance, than individuals low in need for cognition. Although confidence scores did not differ across levels of need for cognition, the results of several regression analyses revealed an effect on accuracy. Need for cognition predicted the consistency of participants' predictions, $\beta = 0.19$, $p = .047$. In addition, need for cognition interacted with the contrast between cue conditions 2S–3W and 2S–6W to predict accuracy levels ($\beta = 1.13$, $p = .031$) and the weight assigned to strong cues ($\beta = 1.11$, $p = .008$). Analyses of separate regression lines within cue conditions revealed that higher need for cognition predicted increased accuracy and weight in the 2S–3W condition and decreased accuracy and weight in the 2S–6W condition. High-need-for-cognition individuals performed well when given a few weak cues and poorly when given many weak cues. This interpretation is also consistent with two additional interactions—those for consistency ($\beta = 1.07$, $p = .063$) and the weight assigned to weak cues ($\beta = -1.38$, $p = .091$)—that reached only marginal levels of statistical reliability.

*Accountability.* Holding participants accountable for their judgments was predicted to improve confidence performance and, perhaps, impair accuracy. Although there was no effect of accountability on accuracy, it did have the anticipated effect on confidence. Because neither need for cognition nor any of its higher order interactions predicted confidence scores and because both accountability and cue condition were categorical, the trimmed analysis simplified to an ANOVA. This analysis uncovered a main effect of accountability, $F(1, 97) = 8.61$, $p = .004$, that was qualified by an interaction with cue condition, $F(3, 97) = 2.95$, $p = .037$. Within the exclusively strong cue conditions (4S and 2S), accountable participants had more appropriate confidence scores than did unaccountable participants. This was not the case within the conditions that included weak cues (2S–3W and 2S–6W; see Figure 1). In summary, holding participants accountable for their predictions had the hypothesized beneficial effect on confidence but only with exclusively strong cues.

*Cue conditions.* The cue profile provided to participants was predicted to have an impact on the strategy used to generate predictions and therefore also on accuracy and confidence performance. The hypothesis that strong cue profiles would promote the adoption of strategies that better mimic statistical prediction was clearly supported. There was a strong association between cue condition and best fitting strategy, $\chi^2(15, N = 106) = 149.98$, $p < .001$. In the 4S cue condition, nearly all the participants equally weighted the four strong cues. In the 2S cue condition, participants were split fairly evenly between the optimal ($n = 14$) and equal ($n = 11$) weighting strategies. In the 2S–3W and 2S–6W cue conditions, participants were split between equal weighting of all cues (*ns* = 12 and 7, respectively) and equal weighting of only the strong cues (*ns* = 15 and 19, respectively). Very few participants appeared to incorporate the interaction between strong cues or rely primarily on high cue values; these strategies were the best fitting for only 5 participants.

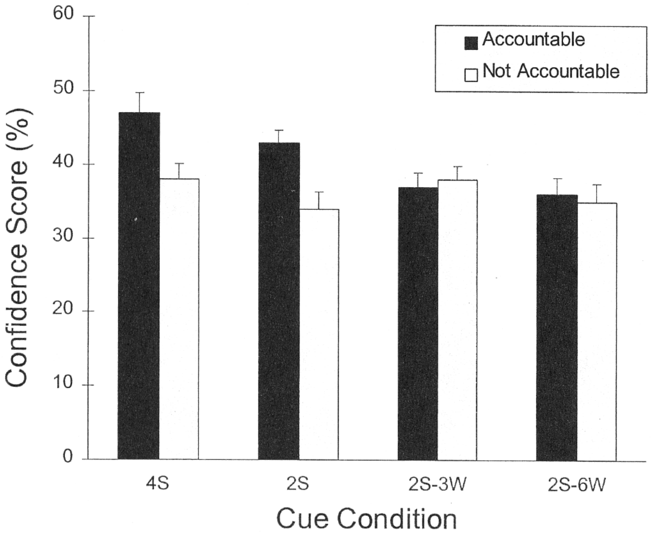

*Figure 1.* Mean confidence score (%) across accountability and cue conditions. Higher scores represent more appropriate degrees of confidence. Each error bar extends to +1 *SEM*. 4S = four strong cues; 2S = two strong cues; 2S-3W = two strong and three weak cues; 2S-6W = two strong and six weak cues.

Cue profiles had a strong impact on accuracy levels. In the trimmed regression model, the contrast between exclusively strong cue conditions (4S and 2S; $M = .42$, $SD = .06$) and those that included weak cues (2S-3W and 2S-6W; $M = .34$, $SD = .09$) was a strong predictor of accuracy ($\beta = 0.48$, $p < .001$). This finding supported the hypothesis that participants given only strong cues would markedly outpredict those given additional weak cues.

Finally, cue profiles influenced confidence performance. In the trimmed analysis (described above) that simplified to an ANOVA, there was a main effect for cue condition, $F(3, 97) = 4.00$, $p = .010$. The pattern of means revealed indicated that confidence performance was positively associated with the strength of the information provided (4S > 2S > 2S-3W > 2S-6W). Restricting access to more valid sets of information did help to calibrate participants' confidence levels.

## Discussion

In many situations, difficult decisions about future events can be addressed either clinically or statistically. Given that extremely important decisions often depend on these predictions, it is critical that we learn when to trust the accuracy of clinical predictions and the confidence with which they are made. In addition, it is important that we determine how best to structure decision-making tasks and select human judges to maximize judgmental quality.

Participants in the present study predicted with an impressive degree of consistency (median $R = .91$), near the upper boundary of what one would expect in a typical judgment study ($R = .70$ to $.90$; Cooksey, 1996). Consistency was particularly high when participants better used available cues, weighting strong cues more heavily than weak cues. Although the predictive accuracy of clinical judgment fell well below that of optimal, unit-weighted, and cross-validated statistical models, it was far above chance levels (mean $r = .40$). There was evidence of a dilution effect in the restricted range evident in some participants' predictions. As hypothesized, accuracy peaked with a moderate degree of dilution, suggesting that it did offset the insufficiently regressive nature of clinical predictions to some extent.

Participants' subjective confidence regarding their predictions was inappropriately high and not associated with accuracy. Interestingly, participants placed greater confidence in their extreme predictions than in those falling closer to the average criterion value. There are at least two potential interpretations of this unexpected result. First, whereas statistically derived confidence intervals grow wider as predictions become more extreme, it may be that extreme clinical predictions are actually guided by confidence levels. High levels of subjective confidence may prompt decision makers to generate extreme predictions. Second, participants may have perceived less room for expressing uncertainty around extreme predictions. Participants tended to construct symmetric confidence intervals around predictions, and the relatively small room for error beyond an extreme prediction may have precluded intervals as wide as those possible around a more average prediction. No matter what its source, this error in judgment undermines the predictive validity of judges' confidence. Future research could investigate the reasons for this deviation from appropriate confidence interval construction, as well as develop methods to correct the problem.

Although participants endorsed several different cue usage strategies, the similarities in their strategies were more striking than the differences. In general, participants gave more weight to strong cues than to weak cues, and models that fit their strategies best reflected either an equal weighting of all cues, an equal weighting of only strong cues, or an optimal weighting of all cues. The latter two strategies, both highly valid methods of cue combination, were the best fitting models for 67 of 106 (63%) participants. Thus, with the exception of confidence performance, the present study rendered a generally positive evaluation of participants' clinical judgment.

Both of the factors hypothesized to stimulate aspects of complex thought—social accountability and need for cognition—produced a mixture of interesting and surprising results. The general hypothesis regarding need for cognition was that it would stimulate an increase in the scope of judgment along with a corresponding tightening of focus. There was some evidence that this led to the expected increase in accuracy levels. Although there was no evidence for improved confidence performance among high-need-for-cognition individuals, these individuals were more accurate with a moderate number of weak cues—but less accurate with a large number of weak cues—than individuals low in need for cognition. On the basis of these results and an array of parallel findings regarding consistency, cue weighting, and strategy use, it seems that the judgments of complex thinkers improved with a bit of weak information. However, when faced with too much of it, complex thinkers succumbed more to the same human weakness that often compels people to incorporate all of the available information—gold and garbage alike—into their predictions.

The present study hypothesized that accountability would stimulate an increase in the scope of judgment without a corresponding tightening of focus. This was expected to result in better confidence performance among participants held accountable for their predictions, although perhaps a decrease in accuracy levels because of dilution. Here, too, there were mixed results. This study revealed no negative effect of accountability on accuracy. Confidence performance did differ as expected, although the main effect was qualified by an interaction with cue condition: Improved confidence performance of accountable participants was only observed with exclusively strong cues.

In addition to these results, it should be noted that accountability had little impact on several other aspects of judgment. The consistency, cue weighting, and strategy use of accountable participants did not differ from those of unaccountable participants. Therefore, although social accountability may lead to improved confidence performance, it is by no means a panacea for all of the shortcomings of clinical prediction.

The four types of cue profiles had a considerable impact on clinical prediction. Among participants given only strong cues, those given fewer cues assigned them greater weight than those given more cues. Predictions were more accurate when participants were given only strong cues than when they were provided with additional weak cues. This appeared to result from a shift in cue-weighting strategies. Participants given a mix of strong and weak cues—regardless of the number of weak cues—either weighted all cues equally or weighted only the strong cues and disregarded the weak cues. Participants weighting only the strong cues achieved greater accuracy levels than those who equally weighted all of the cues, although no differences were found between the groups on confidence performance. These results indicate that individuals provided with exclusively strong cue profiles adopted strategies that were closer to those of normative statistical models.

Cue profiles had a lesser impact on confidence performance than on accuracy levels. There was a main effect of cue condition on confidence, but it was qualified by the interaction with accountability that was described above. Finally, one predicted effect was conspicuously absent: There was no support for the hypothesis that the availability of additional, nondiagnostic information would lead to dilution. As the dilution effect has only been examined in the context of single-prediction tasks, it may be eliminated through the generation of multiple predictions.

Three major limitations of this experiment must be considered. First, the sheer number of cues provided across conditions precluded a direct comparison between exclusively strong cue conditions and conditions including some weak cues. The argument that the overarching group difference could be attributed solely to the number of available cues should be reflected in a linear trend across the number of cues (see Figure 2a). In contrast, a steplike function across the number of cues would support the notion that exclusively strong cue profiles produced judgments qualitatively different from profiles including some weak cues (see Figure 2b). When the actual data are plotted for the contrast between strong and weak cue conditions on accuracy levels, the steplike model—and its associated interpretation of the data—is unambiguously supported (see Figure 2c).

Second, the accountability manipulation may not have been effective. Every effort was made to maximize the strength of this widely used manipulation, but it is unclear how much of an impact the impending audiotaped posttask discussion had on participants. Possible manipulation checks were considered, but each seemed too transparent, reactive, or devoid of meaning to implement. Consequently, there is no direct evidence to demonstrate the strength of this manipulation, and one can only speculate as to why accountability had so few effects on judgment.

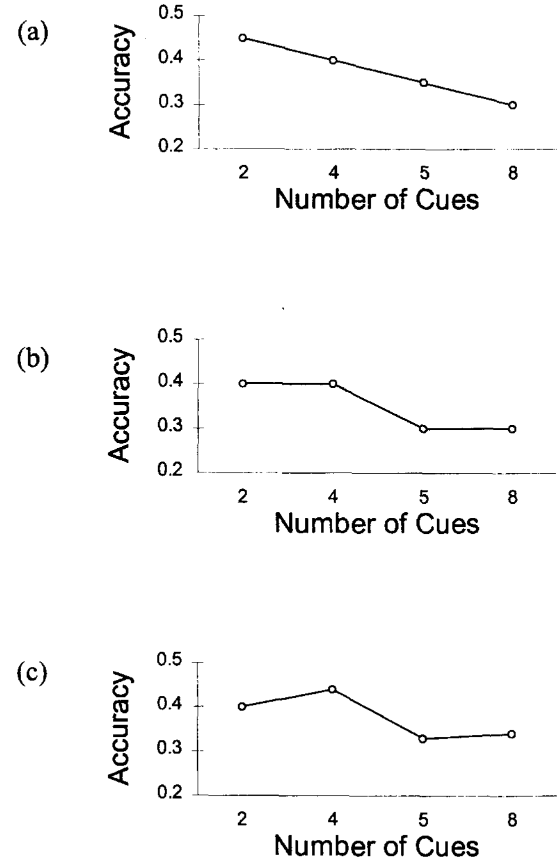

*Figure 2.* Models used to help interpret the cue condition contrast for accuracy levels, in correlational units: Linear function depicts a drop-off in accuracy with increasing numbers of cues (a), steplike function depicts a qualitative difference between strong and weak cue conditions (b), and actual accuracy data supports the steplike model (c).

Third, participants' high predictive accuracy may appear to contradict the relatively poor performance of clinical judgment found in past investigations, thereby limiting the generalizability of these results. However, this apparent inconsistency can be used to highlight important factors that simplified the task and enhanced the efficacy of clinical prediction. To achieve some control over extraneous factors and to ensure that participants were capable of understanding and completing the task, judges were provided with detailed instructions indicating the strength and distributional characteristics of all available cues. All cues were expressed in the same units of measurement (*T* scores) and were positively correlated with the criterion. Given the uniformly positive cue–criterion correlations, the sizable interrelationships among the cues guaranteed substantial accuracy, even when suboptimal weighting schemes were used. Previous research has identified each of these characteristics as favorable to clinical prediction. Therefore, these results provide further support for the ways in which tasks can be structured to facilitate human judgment.

A related concern is that the elevation of performance caused by task simplification may have resulted in a ceiling effect, obscuring the impact of the experimental variables. Perhaps the effects of complex thought on clinical prediction would have been more dramatic in the context of a more challenging task. Examination of judgment quality within additional task environments is therefore necessary before results—particularly null results, if there was indeed a ceiling effect—can confidently be generalized beyond the parameters of this experiment.

Bearing in mind these limitations, the results of the present investigation do have several important theoretical and practical implications. The examination of complex thought has provided some positive leads in the search for reliable individual differences and social influences on clinical prediction, two factors conspicuously absent in most theories of human judgment. The need for cognition, particularly as it interacted with different task structures, had consistent consequences across many facets of clinical judgment. Furthermore, holding judges accountable for their predictions reduced overconfidence, particularly within certain task structures. This experiment is a first step in a new direction, and its results constitute encouraging evidence that complex thought might be an important variable in clinical prediction.

In addition to differences in need for cognition and the manipulation of social accountability, how else might the two aspects of complex thought be stimulated? Factors that have been found to induce individuals to follow the central route to persuasion (Petty & Cacioppo, 1981), such as personal involvement with an issue, may also promote the adoption of more effective clinical prediction strategies. This seems particularly likely if involved individuals are open-minded in their evaluation of issue-relevant information. Kunda (1990) has argued that motivation influences reasoning through the selection of strategies for accessing, constructing, and evaluating beliefs. When our reasoning is driven by predetermined goals or conclusions, people are often strongly biased in favor of deciding as we wish, but when driven by accuracy goals, people are more likely to engage in the type of reality-testing cognitive processes that mimic statistical prediction.

Under what general conditions might it be fruitful to actively suppress complex thought? Results indicate that exposure to largely nondiagnostic cue information hinders predictive performance of clinical decision makers. Ideally, access to nondiagnostic cues should be sharply restricted until human forecasters become more willing and better able to disregard them. When it is impossible to do so, actions taken to reduce complexity of thought may be beneficial.

Finally, what might prompt an individual to abandon a less effective strategy in favor of a more effective alternative? Although changes in strategy use were not assessed in this experiment because of the limited number of cases judged, results suggest that judgment strategies may benefit from intervention. Future research should examine differential responsiveness to feedback among judges exhibiting different levels of complex thought.

## References

Alpert, W., & Raiffa, H. (1969). *A progress report on the training of probability assessors.* Unpublished manuscript.

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology, 66,* 950–967.

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

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

Bickman, L., Karver, M. S., & Schut, L. J. A. (1997). Clinician reliability and accuracy in judging appropriate level of care. *Journal of Consulting and Clinical Psychology, 65,* 515–520.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119,* 197–253.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48,* 306–307.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

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

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

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

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, March, 31). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

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, July 1). The expert witness in psychology and psychiatry. *Science, 241,* 31–35.

Fletcher, F. J. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual difference measure. *Journal of Personality and Social Psychology, 51,* 875–884.

Gardner, W., Lidz, C. W., Mulvey, E. P., & Shaw, E. C. (1996). Clinical versus actuarial predictions of violence in patients with mental illnesses. *Journal of Consulting and Clinical Psychology, 64,* 602–609.

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. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 173–184). Minneapolis: University of Minnesota Press.

Hathaway, S. R., & McKinley, J. C. (1989). *MMPI-2 manual for administration and scoring.* Minneapolis: University of Minnesota Press.

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

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498.

Leary, M. R., Sheppard, J. A., McNeil, M. S., Jenkins, T. B., & Barnes, B. D. (1986). Objectivism in information utilization: Theory and measurement. *Journal of Personality Assessment, 50,* 32–43.

Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin, 24,* 563-574.

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.

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. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Osberg, T. (1987). The convergent and discriminant validity of the Need for Cognition Scale. *Journal of Personality Assessment, 51,* 441–450.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29,* 261–265.

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches.* Dubuque, IA: William C. Brown.

Ruscio, J. (1998a, November) Applying what we have learned: Understanding and correcting biased judgment [11 paragraphs]. *Psycoloquy* [On-line serial], *9*(69). Available: http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?9.69

Ruscio, J. (1998b). Information integration in child welfare cases: An introduction to statistical decision making. *Child Maltreatment, 3,* 143–156.

Ruscio, J. (1998c). The perils of post-hockery. *Skeptical Inquirer, 22,* 44–48.

Tetlock, P. E. (1983a). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45,* 74–83.

Tetlock, P. E. (1983b). Accountability and the perseverance of first impressions. *Social Psychology Quarterly, 46,* 285–292.

Tetlock, P. E. (1985a). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48,* 227–236.

Tetlock, P. E. (1985b). Accountability: The neglected social context of judgment and choice. In L. L. Cummings & B. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 297–332). Greenwich, CT: JAI Press.

Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology, 57,* 388–398.

Tetlock, P. E., & Kim, J. J. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52,* 700–709.

Venkatraman, M. P., Marlino, D., Kardes, F. R., & Sklar, K. B. (1990). Effects of individual difference variables on response to factual and evaluative ads. *Advances in Consumer Research, 17,* 761–765.

Verplanken, B. (1991). Persuasive communication of risk information: A test of cue versus message processing effects in a field experiment. *Personality and Social Psychology Bulletin, 17,* 188–193.

Weathers, F. W. (1992). *The Antisocial Behavior Inventory.* (Available from F. W. Weathers, Department of Psychology, Auburn University, Auburn, AL 36830)

Werner, P. D., Rose, T. L., & Yesavage, J. A. (1983). Reliability, accuracy, and decision-making strategy in clinical predictions of imminent dangerousness. *Journal of Consulting and Clinical Psychology, 51,* 815–825.

## Appendix

### Summary of Cue Data

The following information describes the lowest, average, and highest value on each cue. Additionally, the range of values in which the middle 50% of all cases fall is provided. This range has no special clinical significance and is provided simply to give a sense for the degree of variation present in this sample of cases.

| Cue | Average | Low | 50% range | High |
|---|---|---|---|---|
| Cynicism | 60.9 | 32 | 49–74 | 83 |
| Depression | 73.4 | 34 | 64–85 | 100 |
| Family problems[a] | 64.2 | 33 | 52–77 | 97 |
| Fears | 58.0 | 35 | 48–67 | 90 |
| Health concerns | 70.7 | 33 | 56–83 | 103 |
| Hypochondriasis | 67.9 | 31 | 54–81 | 105 |
| Obsessiveness | 60.7 | 33 | 47–73 | 87 |
| Psychopathic deviate[a] | 69.6 | 40 | 59–79 | 97 |
| Anger[a] | 64.4 | 32 | 53–74 | 86 |
| Criminality[a] | 59.6 | 30 | 51–69 | 90 |
| Antisocial Behavior Inventory | 8.6 | 0 | 6–11 | 19 |

[a] Cues are relatively strong predictors of antisocial behaviors. All other cues are relatively weak predictors.

Received February 22, 1999
Revision received May 21, 1999
Accepted June 6, 1999 ■