Effects of Information Reliability in Predicting Task Performance Using Ability and Effort

Colleen F. Surber
University of Wisconsin—Madison

Past research concluded that predictions of performance are a multiplicative function of ability and effort, although these studies were not designed to test between the averaging and multiplying models. The present research competitively tested these models by manipulating the reliability of information about effort and ability. The results showed that the greater the reliability of the ability information, the greater its effect on judged performance. Similarly, the greater the reliability of the effort information, the greater its effect on judged performance. In addition, the greater the reliability of one type of information, the less the effect of the other type of information. These findings are inconsistent with a multiplying model, but they are consistent with an averaging model in which the reliability of information influences its weight. Individual differences in weighting of effort and ability information were also found, and these differences were predictable from self reports of the relative importance of the variables.

In the literature on achievement motivation and achievement attributions, it has been hypothesized that predictions of performance will be a multiplicative function of effort and ability. For example, Heider (1958) stated the hypothesis as follows:

> The personal constituents, namely power and trying, are related as a multiplicative combination, since the effective personal force is zero if either of them is zero. For instance, if a person has the ability, but does not try at all, he will make no progress toward the goal. (p. 83)

Support for the multiplicative combination rule for judgments of the performance of a hypothetical other has been obtained by Anderson and Butzin (1974) and Kun, Parsons, and Ruble (1974). That is, an interaction of ability and effort was found that was concentrated in the Linear × Linear component. Anderson and Butzin, however, suggested that an averaging model in which the weights were allowed to vary with the scale values might provide a competing representation of judgments of performance based on ability and effort information. Neither Anderson and Butzin's work nor that of Kun et al. was designed to test the possibility that an averaging model might also accommodate the results. Recent work by Singh, Gupta, and Dalal (1979) concluded that in the Indian culture, the combination of ability and motivation can be represented by an averaging model. They attributed their results to a difference between the Indian and U.S. cultures. Results reported by Surber (1978, 1980) with U.S. populations also suggested that the averaging hypothesis merited further examination.

Research by Singh et al. (1979) and Surber (1978) tested one qualitative prediction of an averaging model, specifically that a piece of information presented alone should have a larger effect than when presented in combination with other information. Singh et al. (1979) and Surber (1978) asked subjects to judge performance based on infor-

This research was supported in part by Grant BNS-7912414 from the National Science Foundation and by a grant from the Graduate School, University of Wisconsin. The author is grateful to Michael H. Birnbaum whose feedback on previous work led directly to the present research. The author would also like to thank Barbara Stevens for her assistance in collecting and analyzing data.

Requests for reprints should be addressed to Colleen F. Surber, Psychology Department, University of Wisconsin—Madison, Madison, Wisconsin 53706.

Journal of Personality and Social Psychology, 1981, Vol. 40, No. 6, 977–989

mation about ability only, effort only, or ability and effort combined. The results agreed with the predictions of an averaging model. Unfortunately, there is a class of additive models that is capable of predicting effects of the number of pieces of information (T. Anderson & Birnbaum, 1976; Gollob, Rossman, & Abelson, 1973). Thus, the work of Singh et al. and Surber is not definitive.

The present experiment provides a more definitive test of the averaging model by also varying the reliability of information about ability and effort. Recent studies of source credibility by Birnbaum and his colleagues (1976; Birnbaum & Stegner, 1979; Birnbaum, Wong, & Wong, 1976) provide evidence that variation in the credibility of information can be represented by change in the weight of the information. The present experiment extends Birnbaum's analysis of source credibility effects to predictions of academic performance.

Although source credibility has long been a topic of interest to social psychologists (e.g., Cohen, 1964; McGuire, 1968), with few exceptions work on source credibility has not addressed predictions of behavior or outcomes. Attribution researchers, however, have recently begun to explore a variety of factors that influence the credibility of information such as "base rates" for making predictions. For example, sample size taken in determining a base rate (Kassin, 1979a), randomness of a sample (Hansen & Donoghue, 1977), and perceived causal relation of the base rate information to a predicted outcome (Ajzen, 1977; Tversky & Kahneman, 1977) have been examined. As noted by Kassin (1979b), these studies are somewhat atheoretical, though they do provide evidence that the credibility of such variables can be manipulated. An averaging model incorporating source credibility effects as changes in weights has the potential to provide a theoretical umbrella for such phenomena in social attribution. Examination of credibility effects on predictions of achievement can be viewed as an important step in laying the groundwork for the study of source credibility effects on attributions.

In the present experiment, subjects judged the performance of hypothetical students on a comprehensive final exam in a college course. Information about each hypothetical student's intellectual ability was given in terms of an IQ score from one of three different IQ tests described as varying in their reliability. Information about effort was given in terms of estimates of the student's study time for the exam that also varied in reliability. Based on this information, subjects predicted the students' performance on the exam.

Models for Combining Ability and Effort

An averaging model for judgments of performance can be written:

$$R = \frac{w_{IQ}s_{IQ} + w_{ST}s_{ST} + w_0 s_0}{w_{IQ} + w_{ST} + w_0}, \qquad (1)$$

where R is the judged performance; w_{IQ} and w_{ST} are weights of IQ and study time information that depend on the reliability of the information; w_0 is the weight of the initial impression; and s_0, s_{IQ}, and s_{ST} are the scale values of the initial impression (i.e., expected performance in the absence of any information), the IQ information, and the study time information, respectively.

There are two potential ways of modifying a multiplying model to accommodate variation in the reliability of IQ and study time information. One possibility is a kind of weighted multiplying model:

$$R = (w_{IQ}s_{IQ})(w_{ST}s_{ST}). \qquad (2)$$

Although the grouping of the terms is arbitrary, they are grouped as shown to provide an intuitive interpretation of the model. Intuitively, Equation 2 can be conceptualized as a two-step integration process in which the subject first combines the weight of each type of information with the value of it (e.g., weight of IQ combined with value of IQ yields a net impression of the IQ information). The IQ impression and the study time impression are then combined multiplicatively. A second possibility is a model similar to that proposed by Einhorn (1970):

$$R = s_{IQ}^{w_{IQ}} s_{IQ}^{w_{ST}}. \qquad (3)$$

In this model, the reliability of IQ and study time information are incorporated as exponents.

The averaging model of Equation 1 and both of the multiplying models predict that as the reliability of a type of information increases, the effect of that information on the judgment should also increase. For example, the more reliable the IQ information, the greater the predicted effect of IQ. This can be seen in Equations 1 and 2 in that the weight multiplies the scale value (e.g., $w_{IQ}s_{IQ}$).

The averaging and multiplying models differ in the predicted effect of the reliability of one type of information on the impact of other information. The averaging model predicts that as the reliability of one type of information increases, the net effect of the other information decreases. For example, as the reliability of IQ increases, the effect of study time on the judgment should decrease. This can be seen by considering the relative weights of IQ and study time. The relative weight of study time, $w'_{ST} = w_{ST}/(w_{ST} + w_{IQ} + w_0)$, will decrease as the value of w_{IQ} increases because w_{IQ} appears in the denominator. In contrast, the multiplying models in Equations 2 and 3 predict that increasing the reliability of one type of information will *increase* the impact of the other information on the judgment. For Equation 2 this can be seen in that the weight of one type of information multiplies the value of the other type of information as well.

Individual Differences in Relative Importance of IQ and Study Time

In the achievement attribution literature, it has been found that individuals differ in the belief that ability as opposed to effort is responsible for one's own performance (Dweck, 1978; Nicholls, 1975; Weiner, 1974). Based on these findings, individual differences might be expected in the relative importance of IQ versus study time in judging the performance of a hypothetical other. This possibility was examined by having subjects report the relative importance of IQ versus study time for their judgments. The relationship of these self ratings to use of IQ and study time in predicting performance can then be examined.

Method

Instructions

Written instructions stated that the purpose of the experiment was to examine how people use information about a student's ability and effort to predict performance on an exam. The exam was described as a medium difficult comprehensive final in a college course.

IQ information. The instructions stated that information about a student's intellectual ability would be given in terms of an IQ score and that in different cases, the IQ score was obtained from test procedures that differed in reliability. The low reliability IQ test scores were described as based on a short, written, group administered IQ test taking only 10 minutes. The short IQ test was described as open to many sources of possible error, for example, lack of attention to the test, luck in guessing correct answers, and so forth. The instructions also stated that although the short IQ test provided some information about a student's intelligence, it was the most likely to be in error. The medium reliability IQ test scores were described as based on an individually administered test, requiring about an hour. This test was described as more likely to give a good indication of a student's true intelligence because of the larger number of items and the fact that the test was individually administered. The high reliability IQ test scores were described as based on three repeated administrations of the medium reliability IQ test, using a different form of the test each time. The instructions stated that the average of the three scores provided a highly reliable measure of true IQ because of the large variety of test items, administration of the test on 3 separate days, and so forth. This procedure was described as producing an IQ score that is "as close as you can get to the student's true IQ."

Study time information. Information about study time was given in terms of how much the student studied for the course compared to other students. This information was described as obtained by having students record their amount of studying for various periods of time. Subjects were told to assume that all students reported their study time truthfully. The low reliability study time estimate was described as based on the amount of time the student spent studying for the course for one randomly selected day during the semester. This estimate was described as not a very reliable estimate of overall effort in the course. Factors such as exams in other courses or other activities may have conflicted with the student's study effort on that day. Similarly, a high study time for a single day may not be a good indicator because the day may be atypical. The medium reliability study time estimate was described as based on recorded study time for a whole week during the semester. This procedure was described as more likely to give a reliable indicator of overall study effort than the 1-day estimate. The high reliability study time estimate was described as based on recorded study time for a whole month during the semester. This procedure was described as the most likely to give a reliable estimate of the student's overall effort in the course.

Design and Procedure

There were 144 trials generated by a 3 (reliability of IQ) × 4 (level of IQ) × 3 (reliability of study time) × 4 (level of study time) factorial design. The levels of IQ were verbally described as well below average, somewhat below average, somewhat above average, and well above average. The four levels of study time were described in the same way. In addition, there were 24 trials generated by a 3 (reliability of IQ) × 4 (level of IQ) design and a 3 (reliability of study time) × 4 (level of study time) design. These 168 trials were randomly ordered and printed in booklets. The IQ information was printed above the study time information on each trial. The experimental trials were preceded by 22 practice trials, which included some stimuli more extreme than those of the main design (e.g., "extremely above average" or "extremely below average"). After completing the practice trials, subjects were asked if they had any questions. To decrease the likelihood of any potential effects of a single random order of trials, some subjects were verbally instructed to answer the odd-numbered trials first, followed by the evens, and vice versa. Others were allowed to work straight through the booklet. Each subject worked at his or her own pace, with most completing the experiment in approximately 1 hour.

Rating Scale

The subjects judged performance using integers between 1 and 19, varying from 1 = extremely below average performance, to 10 = average, to 19 = extremely above average performance.

Relative Importance Ratings

After completing the experimental trials, subjects filled out a brief questionnaire in which they were asked to judge the relative importance of each type of study time estimate (1 day, 1 week, 1 month) compared to each type of IQ estimate (short test, long test, repeated long test). In addition to the nine relative importance judgments generated in this way, subjects also reported the relative importance of each type of study time estimate compared to the other study time estimates and of each IQ estimate compared to the others. The rating scale consisted of the integers 1 to 11 printed between each pair of stimuli. One and 11 were labeled "very very much more important," and 6 was labeled "equally important."

Subjects

The subjects were 65 undergraduate students at the University of Wisconsin who participated for extra credit in an introductory psychology course. There were 16 males and 49 females.

Results

Test of the Averaging Model

The left-hand panel of Figure 1 presents the effects of IQ and IQ reliability on judged performance (averaged across study time and study time reliability). As predicted by both the multiplying and averaging model, as IQ reliability increases, the effect of the level of IQ increases. This is also true for the effect of study time and study time reliability, which are presented in the right-hand panel of Figure 1 (averaged over the levels of IQ and IQ reliability). The IQ × IQ Reliability interaction was significant, $F(6, 384) = 120.94$, as was the Study Time × Study Time Reliability interaction, $F(6, 384) = 108.59$.

Figure 2 presents the evidence that distinguishes the averaging from the multiplying model. The left-hand panel of Figure 2 presents the mean judgments of performance as a function of the level of IQ (abscissa), with a different curve for each level of study time reliability. It can be seen that the higher the reliability of study time, the lower the effect of the level of IQ. This finding is predicted by the averaging model if study time reliability influences the value of w_{ST}. As the value of w_{ST} increases, the slope of the IQ curve should decrease, since it is proportional to $w_{IQ}/(w_{IQ} + w_{ST} + w_0)$. Analogously, the higher the IQ reliability, the lower the observed effect of study time (see the right-hand panel of Figure 2). Both the Study Time Reliability × IQ and IQ Reliability × Study Time interactions were significant, $Fs(6, 384) = 13.41$ and 20.41, respectively.

Figure 3 presents the mean judgments of exam performance for the complete 3 × 4 × 3 × 4 design. The 16 points in each panel are the 4 × 4 combinations of IQ and study time for one level of IQ reliability combined with one level of study time reliability. In each panel, IQ is on the abscissa, and there is a separate curve for each level of study time. The panels in the top row are the mean judgments for the low level of IQ reliability, the middle row for medium IQ reliability, and the bottom row for high IQ reliability. The level of study time reliability increases across the panels from left to right.

The data of Figure 3 can be seen to agree with the predictions of the averaging model. As the level of study time reliability increases (as one moves from the left panel to the right panel within each row), the spread of the curves increases. This follows from the fact that the spread of the curves in each panel should be related to the relative weight of study time. Similarly, the effect of IQ reliability can be seen by examining the change in slope within each column. The curves are steeper in the bottom row than in the top row. The effect of IQ reliability

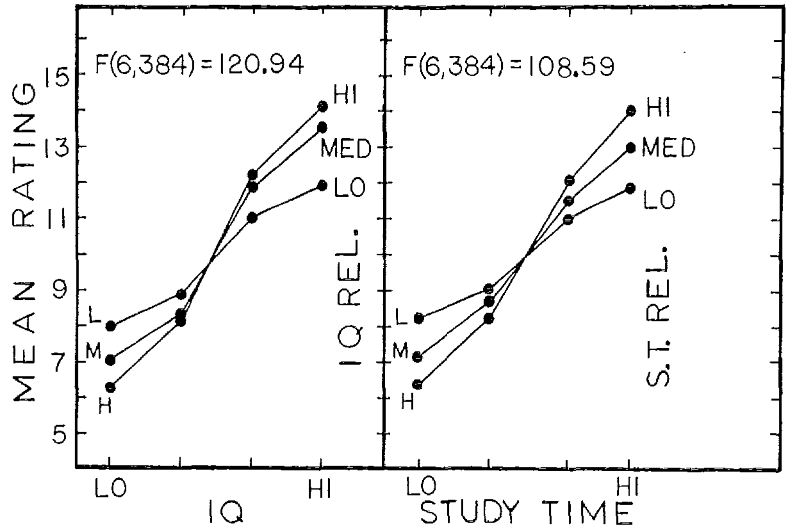


Figure 1. Mean judgments of exam performance as a function of IQ and IQ reliability (left-hand panel) and study time and study time reliability (right-hand panel).

can be seen to decrease the effect of study time by noting that within each column of panels, the steeper the slope the *smaller* the spread of the curves. This follows from the averaging model, since increasing the absolute weight of IQ (w_{IQ}) should decrease the relative weight of study time, $w_{ST}/(w_{ST} + w_{IQ} + w_0)$. The need for the initial impression in Equation 1 can be seen by examining the panels in the diagonal of Figure 3. In the upper left corner, where the reliability of both cues is low, neither the slope nor spread is very great. In contrast, in the lower right panel, where the reliability of both cues is high, both the slope of the curves and the spread of the curves are great. This is predicted nicely by the relative weight averaging model, since the relative weight of the initial impression, $w_0/(w_{IQ} + w_{ST} + w_0)$, should decrease as the values of either w_{IQ} or w_{ST} increase.

A significant IQ × Study Time interaction, $F(9, 576) = 5.84$, was also found. This interaction is due to the fact that averaged

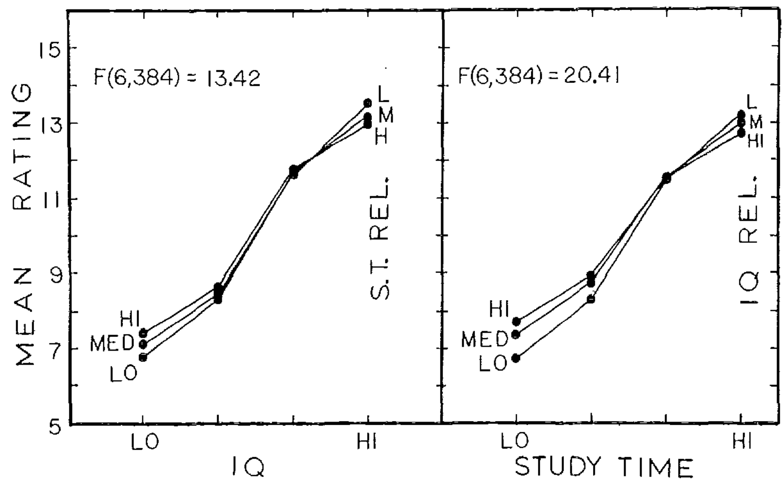
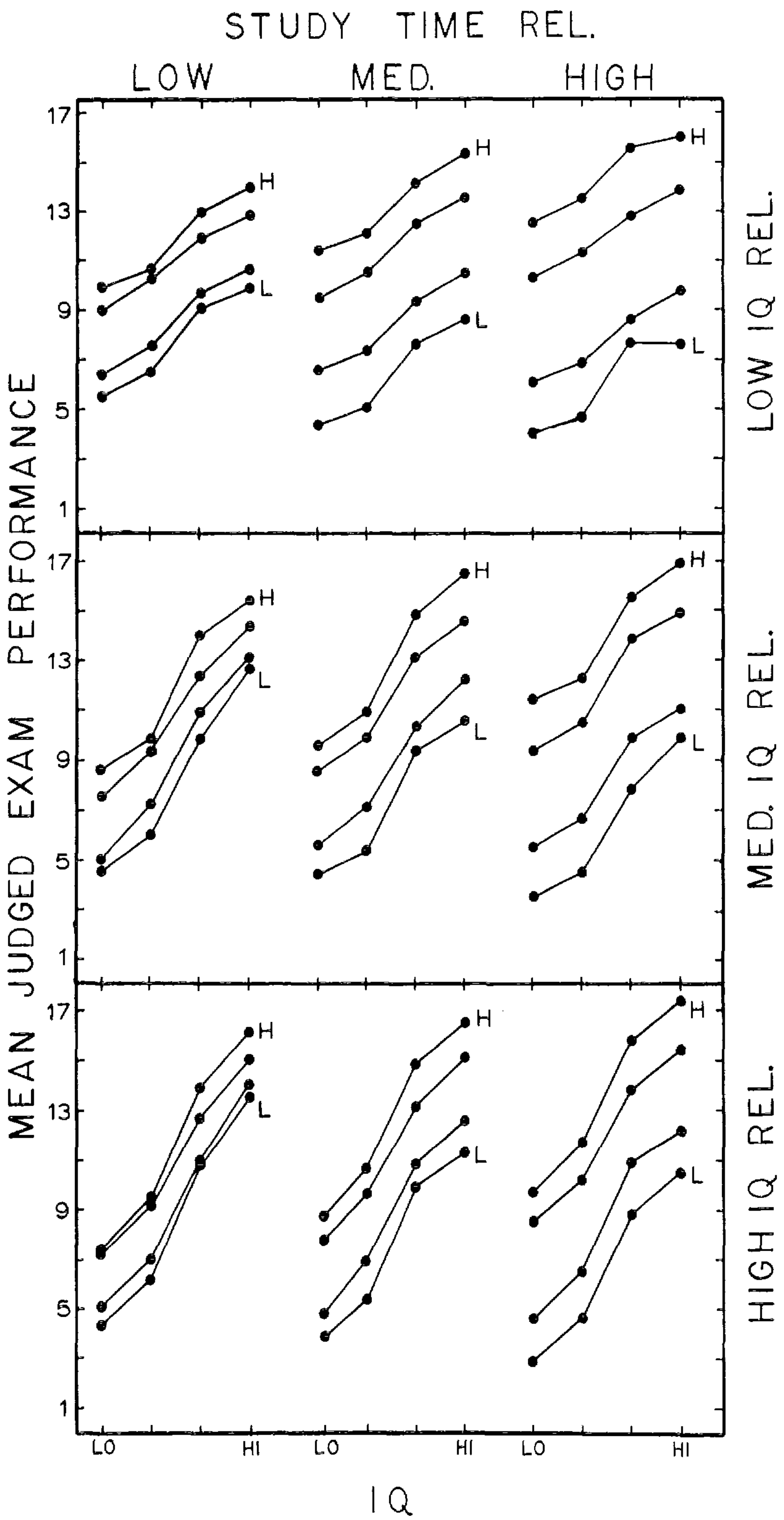


Figure 2. Mean judgments of exam performance as a function of IQ and study time reliability (left-hand panel) and study time and IQ reliability (right-hand panel). (Note that the order of curves in each panel of Figure 2 is the reverse of the order in Figure 1, as predicted by an averaging model.)


STUDY TIME REL.
LOW
MED.
HIGH
MEAN JUDGED EXAM PERFORMANCE
LOW IQ REL.
MED. IQ REL.
HIGH IQ REL.
17
13
9
5
1
H
L
LO
HI
IQ

over the levels of IQ reliability and study time reliability, the curves converge slightly as the level of IQ increases. This interaction differs from Anderson and Butzin's (1974) and Kun et al.'s (1974) results and is inconsistent with a multiplying model. The multiplying model predicts that the curves in each panel of Figure 3 should diverge toward the right as the level of IQ increases. The present findings are consistent with other findings, however (Singh et al., 1979; Surber, 1978). An averaging model can account for the interaction in the present experiment if the weights are allowed to vary with the scale values (see Birnbaum & Stegner, 1979, for a discussion of configural versus differentially weighted averaging models). There was also a significant interaction of IQ Reliability × IQ × Study Time Reliability, $F(12, 768) = 2.97$. This interaction was small and appeared to be due to variations in the size of the interaction of study time reliability with IQ across levels of IQ reliability. These effects did not appear to be systematic or serious enough to merit further consideration.

An averaging model also predicts that the relative weight of information depends on the number of other pieces of information presented with it. This prediction follows if a missing piece of information is assigned a weight of zero. These predictions of the averaging model of Equation 1 can be tested in the present experiment by comparing the effect of study time information presented alone with its effect when combined with IQ (and vice versa). Figure 4 presents the mean judgments for the IQ × IQ Reliability and the Study Time × Study Time Reliability designs. According to the averaging model, the ordinate variation in each panel of Figure 4 should be greater than the ordinate variation in the corresponding panels of Figure 1 (see Birnbaum et al., 1976, Experiment II). This can be shown by a comparison of the relative weights of the information presented alone, e.g., $w_{IQ}/(w_{IQ} + w_0)$, versus in combination with other information, $w_{IQ}/(w_{IQ} + w_{ST} + w_0)$. Comparison of Figure 4 with Figure 1 shows that these predictions of the relative weight averaging model hold for the present experiment.

The averaging model of Equation 1 was fit to the mean judgments using subroutine STEPIT (Chandler, 1969) to minimize the sum of squared deviations. The weight of the initial impression was set to 1.0. The overall root mean squared error was .290 across the 168 data points. The estimated weights of IQ for the three levels of IQ reliability were .397, .793, and 1.040, respectively, compared with .317, .564, and .819, respectively, for the weights of the three levels of study time reliability.

Individual Differences—Group Analyses

The relative importance judgments were used to divide the sample into two groups: those who reported IQ to be more important than study time ($n = 33$) and those who reported study time to be more important than IQ ($n = 28$). A small number ($n = 4$) were excluded from these analyses because the nine IQ versus study time importance ratings balanced out to be perfectly equal. A t test of the sum of the nine importance ratings confirmed that this procedure resulted in groups that differed significantly in judged importance of IQ versus study time (Ms = 45.27 and 62.93 for the IQ group and study time group, respectively), $t(59) = 9.98$. Speculating from the literature on sex differences in achievement attribution, we might expect to find significant sex differences in the importance ratings. No such differences materialized, however.

Two questions can be addressed by considering the data of the two importance rating groups. First, is judged importance of a variable related to use of the variable in predicting performance? Those who judge IQ to be more important than study time should place more emphasis on IQ in predicting performance and vice versa. Second, if the groups do differ in use of information, how can those differences be represented? One possibility is that the weights of IQ versus study time may vary across groups.

Figure 3. Mean judgments of exam performance as a function of study time and IQ information. (Each row of panels represents a different IQ reliability; each column of panels represents a different study time reliability. In each panel, each solid curve is a different level of study time. IQ levels are on the abscissa.)

Importance rating group was adjoined as an additional factor in the main four-way design. In analysis of variance, all previously significant effects remained significant. The interesting effects are those involving the grouping variable. As can be seen in Figure 5, the group reporting IQ to be more important showed a greater effect of IQ, whereas the group reporting study time to be more important showed a greater effect of study time. These effects were significant as shown by the Group × IQ and Group × Study Time interactions, $Fs(3, 177) = 26.73$, 12.10, respectively.

Three other interactions were also significant: Group × IQ Reliability × IQ, $F(6, 354) = 3.77$; Group × IQ Reliability × Study Time, $F(6, 354) = 3.76$; and Group × IQ × Study Time Reliability, $F(6, 354) = 3.94$. The first two interactions appeared to be due to the fact that the study time importance group showed smaller effects of IQ reliability than the IQ importance group. The smaller interactions involving IQ reliability can be predicted by assuming that the study time group placed less weight on IQ at all levels of IQ reliability. If IQ reliability influences the weight of IQ, then the interactions of IQ reliability with study time and with IQ would be predicted to be smaller for the study time group. The significant Group × IQ × Study Time Reliability interaction appeared to be due to the smaller interaction of IQ with study time reliability in the judgments of the study time importance group.

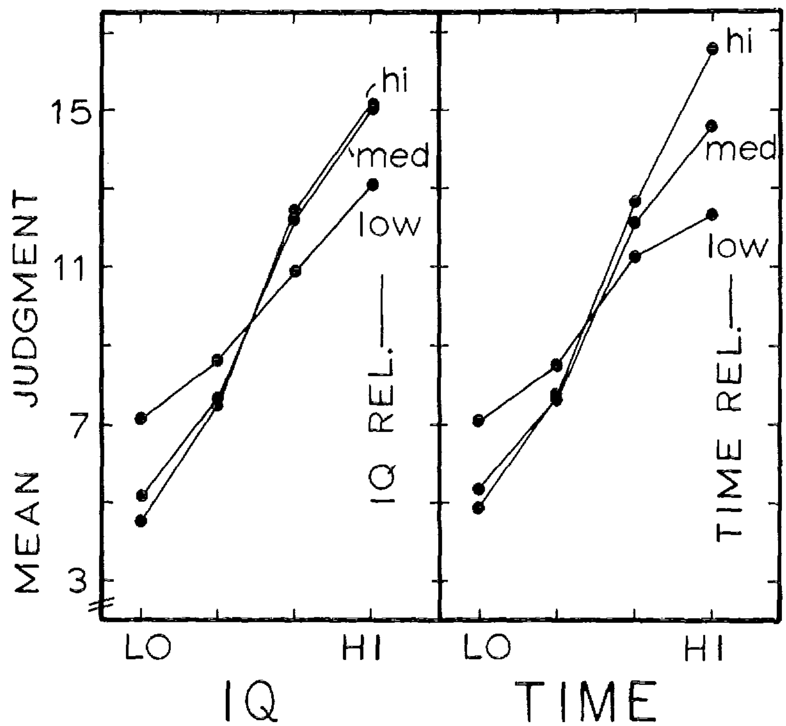


Figure 4. Mean judgments of exam performance for the IQ × IQ Reliability design (left-hand panel) and the Study Time × Study Time Reliability design (right-hand panel).

The averaging model of Equation 1 was fit to the mean judgments of each of the two importance groups using STEPIT and setting the weight of the initial impression in both groups equal to 1.0. The weights of IQ and study time were estimated separately for each group, while the scale values were constrained to be equal. The root mean square error across the two groups was .353. The estimated weights are given in Table 1.

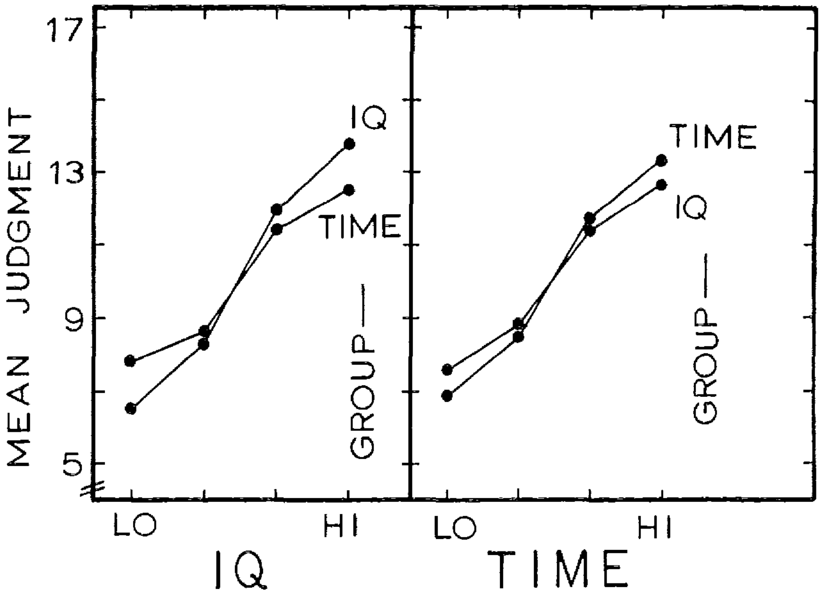


Figure 5. Mean judgments of exam performance as a function of importance group and level of IQ (left-hand panel) and importance group and level of study time (right-hand panel). (Notice in the left panel that the curve for the IQ importance group is steeper than the study time importance group curve, whereas in the right panel, the curve for the study time importance group is steeper than the IQ importance group curve.)

Several aspects of the estimated weights are worthy of note. First, all three IQ weights of the IQ importance group are greater than the IQ weights of the study time group. Second, all three study time weights for the study time importance group are greater than the study time weights of the IQ group. Third, for the IQ importance group, the weight of IQ for each level of reliability is higher than the weight of study time at the corresponding reliability level. The opposite relationships hold in the study time importance group. These results are as one would expect based on the importance ratings.

The data of the two groups were also fit with the averaging model by allowing *both* separate weights and scale values for the groups. The addition of separate scale values did not improve the fit noticeably (root mean square error = .342). Table 2 shows that the estimated weights of study time were larger than the estimated weights of IQ for the study time group, and vice versa for the IQ importance group. Thus, the pattern of weights for the two variables within each group does not depend on constraining the scale values to be equal across groups. It appears that changes in the weights are needed to account for the group differences. The estimated scale values across the two groups did not differ greatly, although the range of scale values was slightly greater for the study time importance group than for the IQ importance group.

Individual Differences—Single Subject Analyses

The group analyses provide encouraging evidence that people can report the relative importance of information in their judgments. Stronger evidence for the validity of the importance ratings would be provided if they could be shown to be related to relative weights at the individual level. In order to examine the predictability of the importance ratings at the individual level, Equation 1 was fit to the 168 judgments of each individual. The scale values were fixed at the best fit values obtained from the model fit of the total sample, and the value of w_0 was set to 1.0. The root mean squared error over all individuals was considerably higher than for any of the group fits reported above (1.630), but it should be remembered that each individual judged only one replication of the experiment.

Table 1
Estimated Weights of IQ and Study Time for Two Importance Groups

Importance group	IQ reliability			Study time reliability		
	Low	Medium	High	Low	Medium	High
Study time	.281	.534	.678	.343	.602	.824
IQ	.421	.885	1.188	.277	.503	.744

Note. Scale values were estimated in common across groups, and $w_0 = 1.0$.

The estimated weights obtained for each individual were converted to relative weights for each of the nine IQ Reliability × Study Time Reliability combinations. The differences between the relative weights (e.g., the relative weight of IQ is $w_{IQ}/(w_{IQ} + w_{ST} + w_0)$) for each of the nine IQ Reliability × Study Time Reliability conditions were correlated with the nine importance judgments across all individuals. The differences between relative weights seemed to agree with what the importance ratings intuitively represent, since subjects were asked to report the "relative importance" of each source of information compared with each other source. The correlation was significant and accounted for more than half of the variance ($r = .800$). This provides some evidence that the importance ratings are related to the estimated weights at the individual level as well.

Discussion

The data of the present experiment provide evidence in favor of the averaging model as a representation of the way ability and effort information are combined. It appears that Singh et al. (1979) may have been too hasty in concluding that the Indian and American cultures differ in how they view ability and effort as determining performance. In both the present experiment and in Singh et al.'s study, there was no evidence of a multiplicative or diverging interaction of ability and effort. The present results extend those of Singh et al. by showing that the predictions of the averaging model hold when information reliability is manipulated.

The fact that the interaction of ability and effort in the present work (and in Singh et al.'s) was not the multiplicative or diverging pattern found by both Anderson and Butzin (1974) and Kun et al. (1974) needs to be explained, however. One possible source of the difference is that Anderson and Butzin described the tasks as extremely difficult. The instructions for judging graduate school performance stated, "A disturbingly large number of graduate students do not last beyond the first year of study" (Butzin, Note 1). In the present experiment, the difficulty of the test was purposely described as medium so that the results would be representative of college students' views of performance in college courses. Results similar to the present experiment have been obtained in other experiments in which college students judged academic performance (Surber, 1978, in press). Singh et al.'s work, which produced results closely resembling those of the present experiment, included no special instructions pertaining to difficulty. Based on this analysis, task difficulty may influence the way ability and effort are subjectively combined to predict performance

Table 2
Estimated Weights of IQ and Study Time for Two Importance Groups

Importance group	IQ reliability			Study time reliability		
	Low	Medium	High	Low	Medium	High
Study time	.240	.455	.571	.300	.522	.709
IQ	.458	.971	1.316	.306	.559	.833

Note. Scale values were estimated separately for each group, and $w_0 = 1.0$.

(Kun & Weiner, 1973; Surber, in press). It is possible, for example, that a multiplying model will provide a better representation of judgments of performance on a high difficulty task. Future research could profitably examine this issue.

Individual Differences and Self Reports

Researchers examining information integration in social judgment have begun to place more emphasis on the ways in which individuals differ in their strategies for combining information (Birnbaum & Stegner, in press; Lopes, 1976; Ostrom & Davis, 1979; Ostrom, Werner, & Saks, 1978). The present work provides a substantive contribution to the literature on individual differences in information integration by demonstrating that (a) the relative importance placed on ability versus effort varies across individuals, (b) individual differences in the relative importance of ability versus effort can be predicted from the subjects' own reports, and (c) these differences can be adequately represented by changes in the values of the weights of IQ and study time. The fact that self reports predicted the use of information in the present experiment provides evidence contradicting Nisbett and Wilson's (1977) claim that people cannot report on their own subjective weights but instead report some cultural standards for how the information is to be used.

Further research on individual differences in achievement judgments is needed. It is possible, for example, that the judgments of individuals who differ in the importance of ability versus effort will differ in ways other than the relative weight of the cues. In actual achievement settings, individuals who report that ability is a predominant factor in their own failures tend to show less persistance at tasks (Dweck, 1978). It is possible that those who emphasize ability in attributing their own failures believe that *both* high ability and high effort are necessary for high performance. Such a belief might be expected to yield a diverging interaction in predicting performance. In contrast, those who attribute their own performance to effort might be less likely to show the diverging pattern. This hypothesis could be examined by collecting multiple replications of predictions of performance from each individual, relative importance judgments, and a more traditional attribution measure. Whether beliefs about one's own performance generalize to judgments of hypothetical others is an important, albeit largely unexamined, topic. The present work points to intriguing avenues for research on this topic.

Implications for Heuristics of Judgment in Attribution

Recently, Ross (1977) discussed the topic of "attributional biases in prediction," employing a variety of heuristic concepts such as representativeness, availability, anchoring and adjustment, concrete versus abstract information, correlation error, regression error, conservatism and nonconservatism. The approach of the present study suggests an alternative to enumerating judgmental heuristics in predicting outcomes.

As pointed out by Birnbaum (1976), an averaging model of source reliability effects can predict judgments that others might describe in terms of a variety of heuristics. For example, anchoring and adjustment refers to adjusting predictions that were initially anchored at some salient value. In the present experiment, the effects of information reliability on predictions could be discussed in terms of more firm anchoring at the higher reliability values in combination with less adjustment for low reliability information. Similarly, the finding that the judgments based on only a single type of information (Figure 4) are more extreme than judgments based on the same information when it is combined with other information (Figures 1 and 3) could be described as nonconservative in Ross's terminology. This might be viewed as an underutilization of the mean value when making predictions using a single source of information. Ross calls this form of nonconservatism "regression error." It would obviously be awkward to attempt to describe the full pattern of results of the present experiment in terms of such heuristics. Happily, all the results can be parsimoniously described by the averaging model. In the current context, the model can provide a unifying theoretical framework for predicting when the effects described by the various heuristics will occur.

By extending models of source credibility effects to predictions of achievement, the present research suggests a variety of experiments on source credibility in attribution. The most immediate extension would be to examine the effects of information reliability on attributions of ability and effort in an experiment analogous to the present one. For example, the reliability of information about performance and study time might be manipulated while asking for attributions of IQ. A common assumption of attribution theories is that how causes are regarded as determining an effect has an influence on attributions of the effect (Kelley, 1972; Reeder & Brewer, 1979; Zuckerman & Mann, 1979). Based on this assumption (albeit, a questionable one), one might expect to find effects of information reliability on ability and effort attributions that parallel the present ones.

The results of the present experiment relate directly to the finding by Kassin (1979a) that sample size influences the impact of base rate information. The manipulation of study time reliability was essentially a manipulation of sample size (the number of days during the semester on which study time was sampled) that can be represented as change in weight in an averaging model. This raises the possibility that a number of the factors that influence use of base rate information may be described by parameters of an averaging formulation.

Proposing that an averaging model has the potential to describe a variety of source credibility effects in attribution does not mean that it necessarily will be successful. Application of algebraic models of judgment to the effects of variables such as concreteness–abstractness of information, the perceived causal relation of information to outcome, randomness and/or size of samples represented by base rate information, etc., can serve several purposes. First, it will provide a theoretical context for unifying a set of phenomena in attribution. Second, it will help to discover and define the boundaries of algebraic models of attribution. Third, such research will provide enriched empirical interpretations for the parameters in the models and by doing so should stimulate research into the cognitive processes behind the models (cf. Graesser & Anderson, 1974; Slovic, Fischoff, & Lichtenstein, 1977; Lopes & Ekberg, Note 2).

Reference Notes

1. Butzin, C. Personal communication, July 1977.
2. Lopes, L., & Ekberg, P. S. *Comparison of analog and computational strategies in multiplicative judgment.* Unpublished manuscript, University of Wisconsin, Madison, 1981.

References

Ajzen, I. Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, 1977, *35*, 303–314.

Anderson, N. H., & Butzin, C. A. Performance = Motivation × Ability: An integration-theoretical analysis. *Journal of Personality and Social Psychology*, 1974, *30*, 598–604.

Anderson, T., & Birnbaum, M. H. Test of an additive model of social inference. *Journal of Personality and Social Psychology*, 1976, *33*, 655–662.

Birnbaum, M. H. Intuitive numerical prediction. *American Journal of Psychology*, 1976, *89*, 417–429.

Birnbaum, M. H., & Stegner, S. E. Source credibility in social judgment: Bias, expertise and the judge's point of view. *Journal of Personality and Social Psychology*, 1979, *37*, 48–74.

Birnbaum, M. H., & Stegner, S. E. Measuring the importance of cues in judgment for individuals: Subjective theories of IQ, heredity, and environment. *Journal of Experimental Social Psychology*, in press.

Birnbaum, M. H., Wong, R., & Wong, L. K. Combining information from sources that vary in credibility. *Memory & Cognition*, 1976, *4*, 330–336.

Chandler, J. D. Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, *14*, 81–82.

Cohen, A. R. *Attitude change and social influence.* New York: Basic Books, 1964.

Dweck, C. S. Achievement. In M. E. Lamb (Ed.), *Social and personality development.* New York: Holt, Rinehart & Winston, 1978.

Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 1970, *73*, 221–230.

Gollob, H. F., Rossman, B. B., & Abelson, R. P. Social inference as a function of the number of instances and consistency of information presented. *Journal of Personality and Social Psychology*, 1973, *27*, 19–33.

Graesser, C., & Anderson, N. H. Cognitive algebra of the equation: Gift Size = Generosity × Income. *Journal of Experimental Psychology*, 1974, *103*, 692–699.

Hansen, R. D., & Donoghue, J. M. The power of consensus: Information derived from one's own behavior. *Journal of Personality and Social Psychology*, 1977, *35*, 294–302.

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

Kassin, S. M. Base rates and prediction: The role of sample size. *Personality and Social Psychology Bulletin*, 1979, *5*, 210–213. (a)

Kassin, S. M. Consensus information, prediction, and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology*, 1979, *37*, 1966–1981. (b)

Kelley, H. H. *Causal schemata and the attribution process*. Morristown, N.J.: General Learning Press, 1972.

Kun, A., Parsons, J., & Ruble, D. Development of integration processes using ability and effort information to predict outcome. *Developmental Psychology*, 1974, *10*, 721–732.

Kun, A., & Weiner, B. Necessary versus sufficient causal schemata for success and failure. *Journal of Research in Personality*, 1973, *7*, 197–207.

Lopes, L. L. Model-based decision and inference in stud poker. *Journal of Experimental Psychology: General*, 1976, *105*, 217–239.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 3). Reading, Mass.: Addison-Wesley, 1968.

Nicholls, J. G. Causal attributions and other achievement-related cognitions: Effects of task outcome, attainment value, and sex. *Journal of Personality and Social Psychology*, 1975, *31*, 379–389.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, *84*, 231–259.

Ostrom, T. M., & Davis, D. Idiosyncratic weighting of trait information in impression formation. *Journal of Personality and Social Psychology*, 1979, *37*, 2025–2043.

Ostrom, T. M., Werner, C., & Saks, M. J. An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, 1978, *36*, 436–450.

Reeder, G. D., & Brewer, M. B. A schematic model of dispositional attribution. *Psychological Review*, 1979, *86*, 61–79.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10). New York: Academic Press, 1977.

Singh, R., Gupta, M., & Dalal, A. K. Cultural difference in attribution of performance: An integration-theoretical analysis. *Journal of Personality and Social Psychology*, 1979, *37*, 1342–1351.

Slovic, P., Fischoff, B., & Lichtenstein, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, *28*, 1–39.

Surber, C. F. Organization in social inference: Is there a schema for judgments of ability, effort, and performance. Unpublished doctoral dissertation, University of Illinois, 1978.

Surber, C. F. The development of reversible operations in judgments of ability, effort and performance. *Child Development*, 1980, *51*, 1018–1029.

Surber, C. F. Necessary versus sufficient schemata: Attributions for achievement in difficult and easy tasks. *Journal of Experimental Social Psychology*, in press.

Tversky, A., & Kahneman, D. Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Weiner, B. *Achievement attribution and attribution theory*. Morristown, N.J.: General Learning Press, 1974.

Zuckerman, M., & Mann, R. W. The other way around. Effects of causal attributions on estimates of consensus, distinctiveness, and consistency. *Journal of Experimental Social Psychology*, 1979, *15*, 582–597.

Received June 4, 1980
Revision received October 27, 1980 ■