Journal of Personality and Social Psychology
1984, Vol. 46, No. 2, 249–268

Inferences of Ability and Effort: Evidence for Two Different Processes

Colleen F. Surber
University of Wisconsin–Madison

Past research and theory on causal inference has not explored the effects of varying the reliability of information. In two experiments, subjects judged either ability (given performance and effort information) or effort (given performance and ability information) where both the reliability and value of the given information varied. Individual differences were found in the judged relationship between ability and effort. Some judged ability and effort to be positively related, whereas others judged ability and effort to be negatively related. These groups also differed in the way information reliability influenced their judgments. The positive group showed effects that agree with either an averaging or correlational model: the higher the reliability of one type of information, the less the effect of the other type of information. For the negative group, an increase in the reliability of one type of information actually increased the effect of the other type of information, a result that is inconsistent with the averaging model. Both an expectancy-contrast model and a correlational model can account for the results of the negative group. The different effects of information reliability for the two groups can be interpreted as evidence of two different inference processes. The results show the flexibility of human judgment strategies and the need for research considering variables that influence strategy use.

The present research deals with the issue of information credibility in causal inference, specifically with how the reliability of information affects inferences of ability and effort. Although there are many social psychological theories of causal inference (Kelley, 1972, 1973; Reeder & Brewer, 1979), there has been little attempt to formulate a theoretical representation of the way information credibility influences causal inferences. The present research tests alternative representations of the information integration process for causal inferences. Because models of information integration can be regarded as formal representations of causal schemata (Surber, 1981b), the present study can be seen as part of a research program aimed at specifying the variables that influence when different information integration rules or causal schemata are adopted for social inference tasks.

This research was supported in part by National Science Foundation Grant BNS 79-12414 and a grant from the Graduate School, University of Wisconsin.

I wish to thank Steven M. Gzesh for assistance with data analyses, and Lola L. Lopes and Barbara Mellers for useful discussions of theory.

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

AVERAGING MODEL

Anderson and Butzin (1974) suggested that ability and motivation were judged according to some type of general linear process, possibly an averaging model. The effects of the reliability of information can be incorporated in an averaging model by hypothesizing that the weight for each variable depends on information reliability:

$$\text{Ability} = \frac{W_P S_P + W_E S_E + W_O S_O}{W_P + W_E + W_O} \quad (1)$$

$$\text{Effort} = \frac{W_P S_P + W_A S_A + W_O S_O}{W_P + W_A + W_O}, \quad (2)$$

where W_P represents the weight of performance, S_P represents the value of the performance information, W_E and W_A represent the weights of effort and ability, and S_E and S_A represent the values of the effort and ability information. The term $W_O S_O$ in each equation represents the weight and scale value of the "initial impression" or impression in the absence of any information.

The averaging model makes several qualitative predictions. First, if information reliability influences the weights, then the effect of a given type of information should increase as its own reliability increases. For example, the effect of performance should increase as the reliability of performance increases if reliability influences the value of W_P. Second, Equations 1 and 2 also predict that the higher the reliability of one type of information, the *less* the effect of the other type of information. For example, as the reliability of performance increases, the effect of ability on judgments of effort is predicted to decrease. This prediction can be understood by examining the effective weight of ability: $W'_A = W_A/(W_A + W_P + W_O)$. Because W_P appears in the denominator, as the value of W_P increases the effective weight of ability (W'_A) will decrease.

Past research has typically found that adults judge ability and effort to be inversely related when performance information is given (Anderson & Butzin, 1974; Kepka & Brickman, 1971; Surber, 1981b). If all the weights have positive values, the averaging model can accommodate this finding if the scale values are defined such that when ability is described as high it is assigned a low scale value for inferring effort, and vice versa. An alternative is to assign scale values that are ordered in the same way as the verbally described levels (e.g., ability described as high is given a high scale value compared to ability described as low), but to allow W_E and W_A to have negative values, as proposed by Kepka and Brickman (1971). By manipulating the reliability of the information, the present experiments allow tests of the predictions of a negative-weight averaging model in comparison with the usual positive-weight averaging model. (The details of the negative-weight averaging model are complex and are presented in conjunction with the results below.)

CORRELATIONAL MODEL

An alternative to the weighted averaging model is a correlational model for information credibility effects proposed by Birnbaum (1976) and adapted by Surber (1981b) for judgments of ability and effort:

$$\text{Ability} = \frac{(r_{PA} - r_{AE}r_{PE})S_P}{1 - r_{PE}^2} + \frac{(r_{AE} - r_{PA}r_{PE})S_E}{1 - r_{PE}^2} + S_O \quad (3)$$

$$\text{Effort} = \frac{(r_{PE} - r_{AE}r_{PA})S_P}{1 - r_{PA}^2} + \frac{(r_{AE} - r_{PE}r_{PA})S_A}{1 - r_{PA}^2} + S_O, \quad (4)$$

where the rs represent the subjective correlations between the subscripted variables (P, A, and E refer to performance, ability, and effort, respectively), and S_A, S_E, and S_P are the subjective values of ability, effort, and performance, respectively. Equations 3 and 4 are in the tradition of "humans as intuitive statisticians" (Peterson & Beach, 1967). The correlational model differs from the averaging model in that the qualitative predictions depend on the values of rs. For example, Equation 3 can predict that either (a) ability will be judged to be directly related to effort, at constant levels of performance, or (b) ability will be judged to be inversely related to effort, at constant levels of performance. When all r values are positive, the former result is obtained if $r_{AE} > r_{PA}r_{PE}$, whereas the latter result occurs if $r_{AE} < r_{PA}r_{PE}$.

If the reliabilities of the cues influence the values of the rs, and all rs are assumed to be positive, then the correlational model can predict some of the same qualitative effects as the averaging model. When all the correlations are positive, the effect of a type of information can be predicted to increase with its own reliability and decrease with the reliability of the other given information. For example, the net effect of performance in Equation 3 should be proportional to $(r_{PA} - r_{PE}r_{AE})/(1 - r_{PE}^2)$, so it will increase as r_{PA} increases, other things being equal. The net effect of effort should be proportional to $(r_{AE} - r_{PA}r_{PE})/(1 - r_{PE}^2)$. Because a proportion of r_{PA} is subtracted from r_{AE}, the net effect of effort can decrease as the reliability of performance increases. Increasing the reliability of one cue decreases the effect of the other cue only to the extent that the subject regards the two cues as positively correlated. In contrast, no such assumption is made in the averaging model.

Another distinction between the averaging model and the correlational model is in the effect of information presented alone versus its effect when combined with other information. When only one type of information is presented, it is conventional to assume that the missing information is given zero weight in an averaging model. When all weights are positive, Equations 1 and 2 predict that the net effect of information will be greater when presented alone than when combined with other information. For example, $W_P/(W_P + W_O) > W_P/(W_P + W_E + W_O)$.

Anderson and Butzin (1974), however, proposed that ability and effort are directly related in the absence of performance but that "The ability cue takes on different values when the performance cue is specified" (p. 603). This proposal seems to require an increasing set of scale values for judgments in the absence of performance and a decreasing set for judgments in the presence of performance information. With different scale values, it is not possible to predict the relationship between the net effect of a cue when presented alone versus in combination with other information because the range of scale values may vary as well as the direction of the effect.

One way of representing the effect of information presented alone in terms of the correlational model is to assume that all the subjective correlations are zero except the one representing the judged variable and the presented cue. Under this assumption, and if all correlations are positive when both cues are presented, Equations 3 and 4 can also predict that a cue should have a greater effect when presented alone than when combined with other information. For example, the effect of performance alone on judged ability should be proportional to r_{PA}, whereas it will be proportional to $(r_{PA} - r_{AE}r_{PE})/(1 - r_{PE}^2)$ when combined with effort information.[1]

It can also be assumed that the subject infers the missing piece of information and combines the inferred value with the given information according to Equations 3 and 4, as proposed by Yamagishi and Hill (1981) for their path-analytic model.[2] In the correlational model, when given only effort information, the subject can first infer performance from the given effort value and the assumed relationship between effort and performance, represented by r_{PE}. Thus, $S_P = r_{PE}S_E + S_O$. The inferred performance value then takes the place of S_P in Equation 3, and Equation 3 can be rewritten as follows:

$$\text{Ability} = \frac{(r_{PA} - r_{AE}r_{PE})r_{PE}S_E}{1 - r_{PE}^2} + \frac{(r_{AE} - r_{PA}r_{PE})S_E}{1 - r_{PE}^2} + \text{constant}. \quad (5)$$

Equation 5 reduces to: Ability = $r_{AE}S_E$ + constant. Thus, whether or not the subject infers missing information, Equations 3 and 4 predict that judgments based on a single cue will depend on the information presented and the strength of its assumed relationship to the judged variable.

The two experiments reported below test the averaging and correlational models for inferences of ability and effort. Experiment 1 yielded the surprising finding that individuals show different slopes in using study time and IQ information, some judging IQ to increase as study time increases and others judging the variables to be negatively related. Because the individual differences were unexpected, replication of the finding was highly desirable. Experiment 2 was a replication of Experiment 1 but included a questionnaire asking subjects to report the slopes of their judgments. The questionnaire also provides an independent test of the validity of the slope classification criteria used in Experiment 1. The results of the two experiments are presented together.

EXPERIMENTS 1 AND 2

Method

Instructions

The instructions for judgments of IQ (or study time) informed the subject that he or she would receive infor-

[1] It is possible for the correlational model to predict that the effect of a single cue will be either less than or equal to the effect of a cue when combined with other information. For example, if r_{AE} is near zero and if r_{PE} is large, then $(r_{PA} - r_{AE}r_{PE})/(1 - r_{PE}^2)$ will be larger than r_{PA}. If r_{PE} and r_{AE} are both close to zero, then the effect of performance on judged ability will be the same whether performance is presented alone or in combination with effort.

[2] The path-analytic model is equivalent to the correlational model if the statistical model behind path analysis is assumed. Thus, the path-analytic model is not discussed separately.

mation about either the effort a person expended in studying for a college course (or the student's general intellectual ability, IQ), the student's performance in that course, or both. Subjects were instructed to assume that the course was of medium difficulty. The manipulations of IQ and study time information were identical to those of Surber (1981a).

Performance Information

The information about performance was identical for judgments of IQ and study time and was given in terms of the student's score on either a 10-item quiz, a midterm, or a comprehensive final exam. The 10-item quiz was described as the indicator of overall course performance that is the most likely to be in error. The midterm score was described as much more reliable than a quiz, and the comprehensive final exam was described as the single most reliable indicator of overall course performance.

Study Time Information

When subjects judged IQ, they were given information about study time in terms of how much the student studied for the course compared to others. Subjects were told to assume that the students recorded their amounts of studying for various periods of time and that all students reported their study time truthfully. The low-reliability estimate was described as being based on the amount of time the student spent studying on one randomly selected day during the semester and as not being a very reliable estimate of overall effort in the course. The medium-reliability estimate of study time was based on recorded study time for a whole week during the semester. The high-reliability estimate of study time was described as being based on recorded study time for a whole month during the semester.

IQ Information

When subjects judged study time, they were given information about IQ described as obtained from test procedures that differed in reliability. The low-reliability IQ test scores were described as being based on a short written group-administered IQ test taking only 10 min and open to many sources of possible error. The medium-reliability IQ test scores were described as being based on an individually administered test, requiring about an hour, and were referred to as the long IQ test. The high-reliability IQ test scores were described as being based on three repeated administrations of the medium-reliability IQ test, using a different form of the test each time. This procedure was described as producing an IQ score that is "as close as you can get to the student's true IQ" and was referred to as the repeated long IQ test.

Design

The main design for judgments of IQ consisted of 144 stimuli generated by a 3(study time reliability) × 4(level of study time) × 3(performance reliability) × 4(level of performance) factorial design. The four levels of the variables were verbally described as: well below average, somewhat below average, somewhat above average, and well above average. In addition, there were 12 stimuli generated by a 3(study time reliability) × 4(level of study time) design in which performance information was not specified, and 12 stimuli generated by a 3(performance reliability) × 4(level of performance) design in which study time information was not specified. For stimuli from the main design, the study time information was printed above the performance information in each trial. The design for judgments of study time was analogous to the design for judgments of IQ, except that IQ information was given instead of study time information.

Procedure

The 168 trials for each type of judgment were randomized and printed in booklets, preceded by 30 practice trials that included some stimuli more extreme than those of the experimental trials (e.g., "extremely above average" or "extremely below average"). To decrease the likelihood of any effects of a single random order of stimuli, some subjects were orally instructed by the experimenter to answer the odd-numbered trials first, followed by the even-numbered ones, or vice versa. Subjects participated in groups of approximately 8 (Experiment 1) to approximately 25 (Experiment 2) and worked at their own paces, with most completing the experiment within 1 hour.

Rating Scales

The subjects judged either IQ or study time using integers between 1 and 19, labeled varying from 1, *extremely below average* (IQ or Study Time); to 10, *average* (IQ or Study Time); to 19, *extremely above average* (IQ or Study Time). Subjects were also given a reference page, separate from the booklet, that provided the rating scale and summarized the key points of the instructions and reliability manipulations.

Slope Judgments

Subjects in Experiment 2 also judged the truth of four statements describing the relationship between study time and IQ. The basic form of these statements for the subjects who judged IQ was, "The ___ a person's study time was, the ___ I judged IQ to be." The four possible combinations of the terms *higher* and *lower* filled in the two blanks. For the subjects who judged study time, the terms *study time* and *IQ* were reversed. The four statements were randomized and printed on the last page of the booklet. Subjects judged each statement using the integers 1 to 9, with 1 labeled *certain false*, 5 labeled *neither true nor false*, and 9 labeled *certain true*.

Subjects

The subjects were 295 undergraduate students, who participated to earn extra credit in an introductory psychology course. There were 92 subjects in Experiment 1, of whom 46 were randomly assigned to judge IQ and 46 were randomly assigned to judge study time. The 92 subjects consisted of 60 females (of whom 29 judged study time) and 32 males (of whom 17 judged study time). In Experiment 2 there were 203 subjects, of whom 102 were randomly assigned to judge IQ and 101 to judge study time. The 203 subjects consisted of 130 females (of whom 72 judged IQ) and 73 males (of whom 30 judged IQ).

Results

IQ Judgments

A 3(study time reliability) × 4(study time) × 3(performance reliability) × 4(performance) analysis of variance (ANOVA) showed no significant main effect of study time in either Experiment 1, $F(3, 135) < 1$, or Experiment 2, $F(3, 303) < 1$, although the effect of performance was quite large (both Fs > 200.0). The lack of a significant effect of study time on judged IQ was inconsistent with past research with adults (Surber, 1981b), whereas it resembled results obtained with children (Kun, 1977; Surber, 1980). An examination of the data of individual subjects showed two prevalent patterns: Judged IQ increased as the value of study time increased, and judged IQ decreased as the value of study time increased.

The data of individual subjects were classified by taking the difference between the judgments of the highest and lowest levels of study time at each level of performance, performance reliability, and study time reliability. For each of the nine Performance Reliability × Study Time Reliability combinations, the majority of the signs of the four differences was used to classify that matrix of data as showing a negative or positive slope, or an inconsistent pattern. A subject's overall slope was determined by the majority of classifications of the nine Study Time Reliability × Performance Reliability matrices. In Experiment 1, 18 subjects were classified as showing positive slopes, 23 as showing negative slopes, and 5 as showing either inconsistent or flat slopes. In Experiment 2, 35 subjects showed positive slopes, 48 negative, and 19 inconsistent or flat slopes. A chi-square test of independence showed no difference between Experiments 1 and 2 in the proportion of subjects falling in the slope categories, $\chi^2(2) = 1.52$.

Validity of Slope Grouping

Because only the two extreme study time values were used in classifying the data, the two middle values of study time can be used as an independent check on the validity of the slope classification. An ANOVA omitting the stimuli involved in the slope classification showed a significant Slope Group × Study Time interaction in both Experiment 1, $F(1, 39) = 139.53$, and Experiment 2, $F(1, 81) = 122.9$. The reported slope on the questionnaire at the conclusion of Experiment 2 also showed that the slope groups differed significantly in the predicted direction on each of the four questions (all ps < .01).

Positive Slope Group

The results of the positive slope group show the qualitative effects predicted by a weighted averaging model in which the weights depend on the reliability of the information. The left-hand panel of Figure 1 presents the interaction of Study Time × Study Time Reliability from Experiment 1, with study time plotted on the abscissa and a separate curve for each level of study time reliability. The significant crossover interaction is predicted by both the averaging and correlational models. The right-hand panel of Figure 1 plots the Performance × Performance Reliability interaction, which also has the predicted crossover form.

The left-hand panel of Figure 2 presents the Study Time × Performance Reliability interaction from Experiment 1. The averaging model predicts that as performance reliability increases, the effect of study time should decrease. The correlational model also allows this result. That the prediction holds can be seen in the left-hand panel of Figure 2 in that the steepest curve is for low performance reliability (quiz), and the flattest curve is for high performance reliability (final). Analogously, the effect of performance should be smallest when study time reliability is highest. The Performance × Study Time Reliability interaction (right-hand panel of Figure 2) was not significant, but shows a trend toward the predicted pattern, with the low-reliability study time curve showing the steepest slope and the high-reliability study time curve showing the flattest slope. Another prediction of the averaging model is that the net effect of information depends on the number of other pieces of information presented with it. Graphs of the mean judgments for the partial information designs showed that these predictions held.

The results of the positive slope group of Experiment 2 closely replicated those of Experiment 1. The interactions of Study Time ×

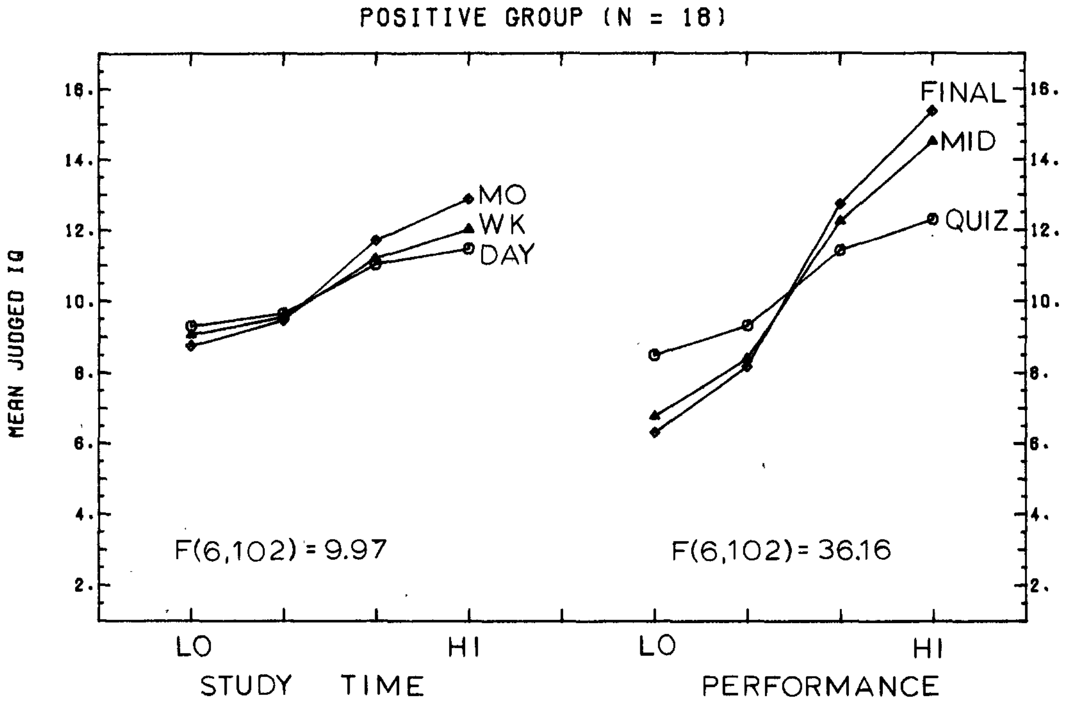

Figure 1. Mean judged IQ for the positive slope group of Experiment 1 as a function of study time and study time reliability (left-hand panel) and performance and performance reliability (right-hand panel).

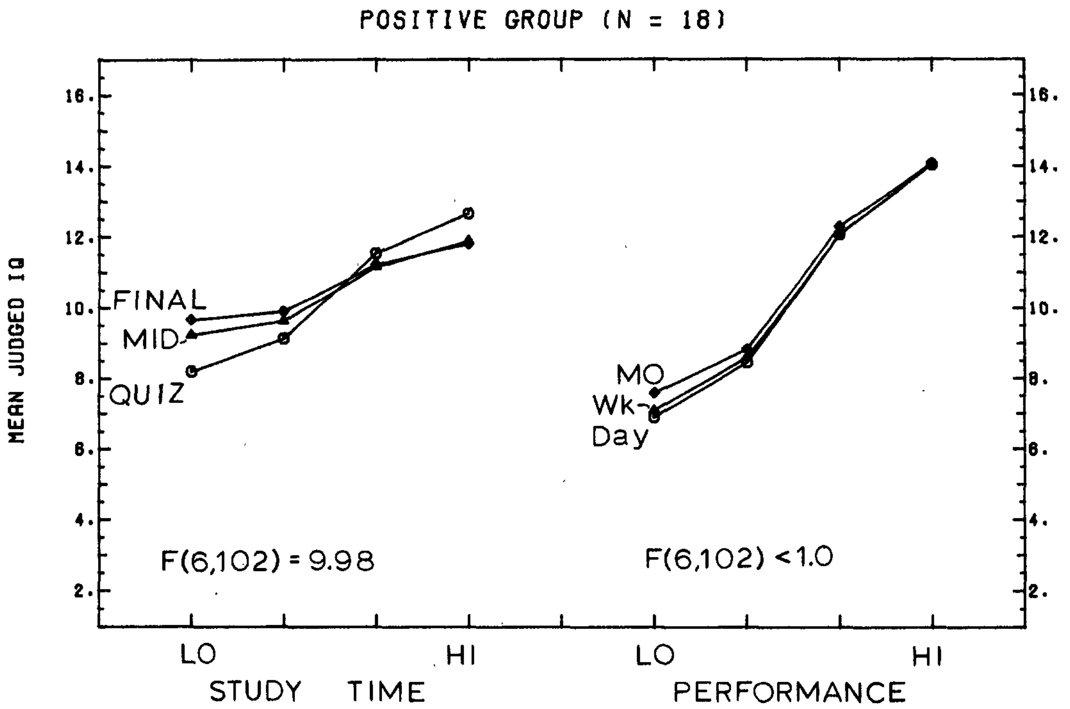

Figure 2. Mean judged IQ for the positive slope group of Experiment 1 as a function of study time and performance reliability (left-hand panel) and performance and study time reliability (right-hand panel).

Study Time Reliability, $F(6, 204) = 9.62$, Performance × Performance Reliability, $F(6, 204) = 72.03$, and Study Time × Performance Reliability, $F(6, 204) = 14.7$, all had the crossover form predicted by the averaging model. The Performance × Study Time Reliability interaction did not reach significance, $F(6, 204) = 1.4$, but a graph of this interaction showed a trend approximately like the right-hand panel of Figure 2. In addition, both study time and performance information had larger effects in the partial information conditions than when combined, as predicted by the averaging model.

Model analyses. The averaging model was fit to the mean IQ judgments of the positive slope groups using subroutine STEPIT (Chandler, 1969). The sum of squared deviations between the predictions and the means was minimized, and the value of W_O was set equal to 1.0. The values of W_P and W_E were assumed to depend on the manipulated reliabilities. Thus 15 parameters were estimated to account for 168 data points. The model provided an adequate fit, with the square root of the average squared deviation equaling 0.407 in Experiment 1 and 0.339 in Experiment 2. These compare well with the standard errors that ranged between 0.141 and 1.080. Figure 3 presents the mean judgments of Experiment 2 for the 144 stimuli of the main design. The symbols represent the means and the solid lines present the predictions of the averaging model obtained from subroutine STEPIT. An inspection of Figure 3 shows that the averaging model captures the important qualitative features of the data.

The correlational model was also fit to the data, first, by assuming that the subjective correlation between the performance and study time cues (r_{PE} in Equation 3) was constant, and second, by allowing r_{PE} to vary with both the reliability of performance and study time. When r_{PE} is a constant, Equation 3 requires estimation of 16 parameters, while allowing r_{PE} to vary requires estimating 24 parameters. With r_{PE} constant, the correlational model fit the data slightly worse than the averaging model (square roots of the average squared deviation were 0.430 and 0.360 for Experiments 1 and 2, respectively). Allowing r_{PE} to vary allowed the correlational model to fit slightly better than the averaging model (0.396 and 0.324, for Experiments 1 and 2, respectively).

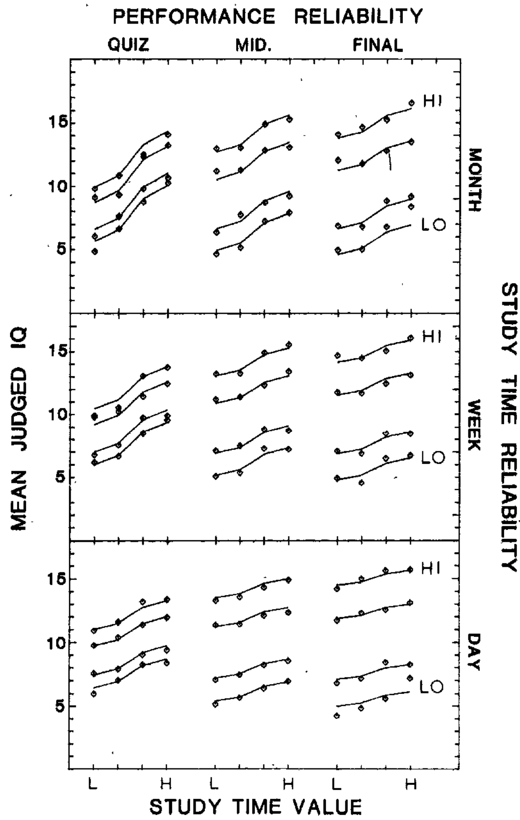

Figure 3. Mean judgments of IQ (symbols) and averaging model predictions (solid curves) for the positive slope group in Experiment 2. (Each panel represents a different combination of performance reliability and study time reliability. In each panel the abscissa variable is the value of study time and each curve is a different value of performance.)

The averaging model can be modified by allowing the weight given to study time versus performance to depend on the configuration of values on a given trial (Birnbaum & Stegner, 1979, 1981). This allows the averaging model to account for the interactions between study time and performance, something the correlational model cannot do. The following configurally weighted averaging model was also fit to the mean judgments of the positive slope groups:

$$\text{Ability} = \frac{W_P S_P + W_E S_E + W_O S_O}{W_P + W_E + W_O} + \omega |S_P - S_E| ,$$

where the parameter omega was allowed to depend on the reliability of performance and study time. The model requires estimation of 24 parameters and fit the data slightly better than the correlational model (square roots of the average squared deviation were 0.372 and 0.295 for Experiments 1 and 2, respectively). Because both the averaging and correlational models account for the major qualitative features of the data, neither model can be definitively eliminated in the present case.

Negative Slope Group

Figure 4 shows the Study Time × Study Time Reliability (left-hand panel) and Performance × Performance Reliability (right-hand panel) interactions from the main design of Experiment 1. The significant crossover interactions in both panels of Figure 4 are as predicted, with the low-reliability curves showing the flattest slopes and the high-reliability curves showing the steepest slopes. Figure 5 presents evidence that contradicts the averaging model of Equation 1, however. The left-hand panel shows the Performance Reliability × Study Time interaction, and the right-hand panel shows the Study Time Reliability × Performance interaction. In both panels the high-reliability curve shows the steepest slope and the low-reliability curve shows the flattest slope, contrary to the averaging model. The results in Figure 5 are actually the opposite of the results predicted by Equation 1 if positive weights are assumed. Figure 6 presents the results of Experiment 1 for the two partial-information designs. When study time information is presented without performance (left-hand panel), the negative slope subjects judge IQ to increase as study time increases, $F(3, 66) = 4.72$. However, a small number of subjects judged IQ to decrease as study time increased even when performance was omitted (6 out of 23 in Experiment 1 and 4 out of 48 in Experiment 2).

The results of the negative slope group in Experiment 2 also replicated the major findings of Experiment 1. The Study Time × Study Time Reliability, $F(6, 282) = 25.70$, Performance × Performance Reliability, $F(6, 282) =$

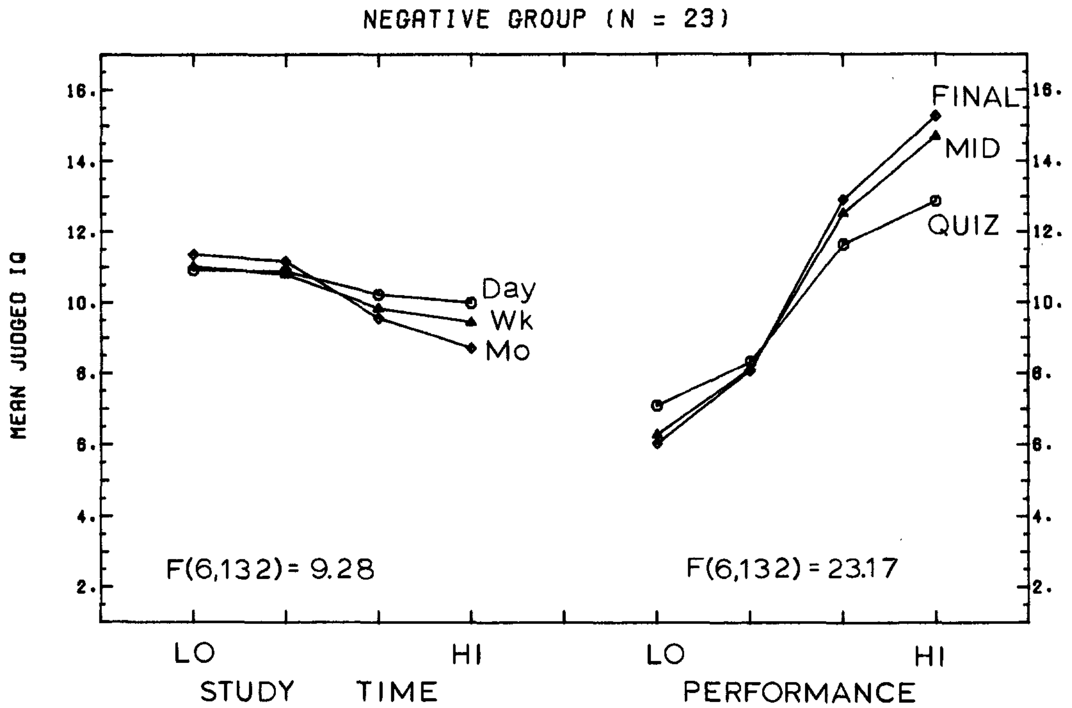

Figure 4. Mean judged IQ for the negative slope group of Experiment 1 as a function of study time and study time reliability (left-hand panel) and performance and performance reliability (right-hand panel).

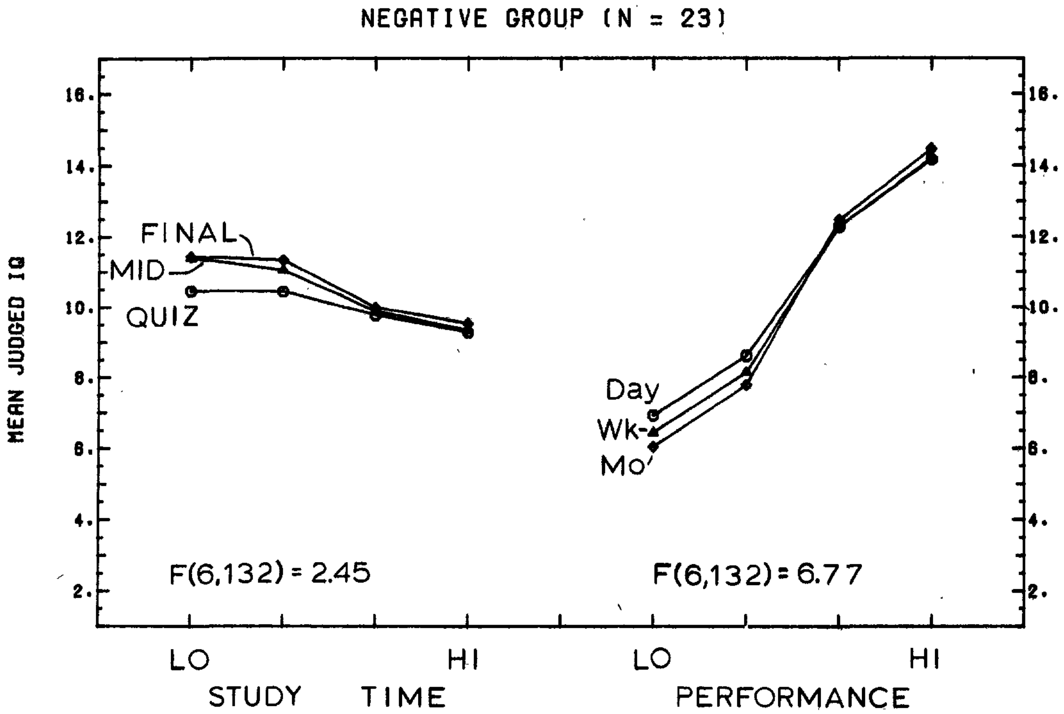

Figure 5. Mean judged IQ for the negative slope group of Experiment 1 as a function of study time and performance reliability (left-hand panel) and performance and study time reliability (right-hand panel).

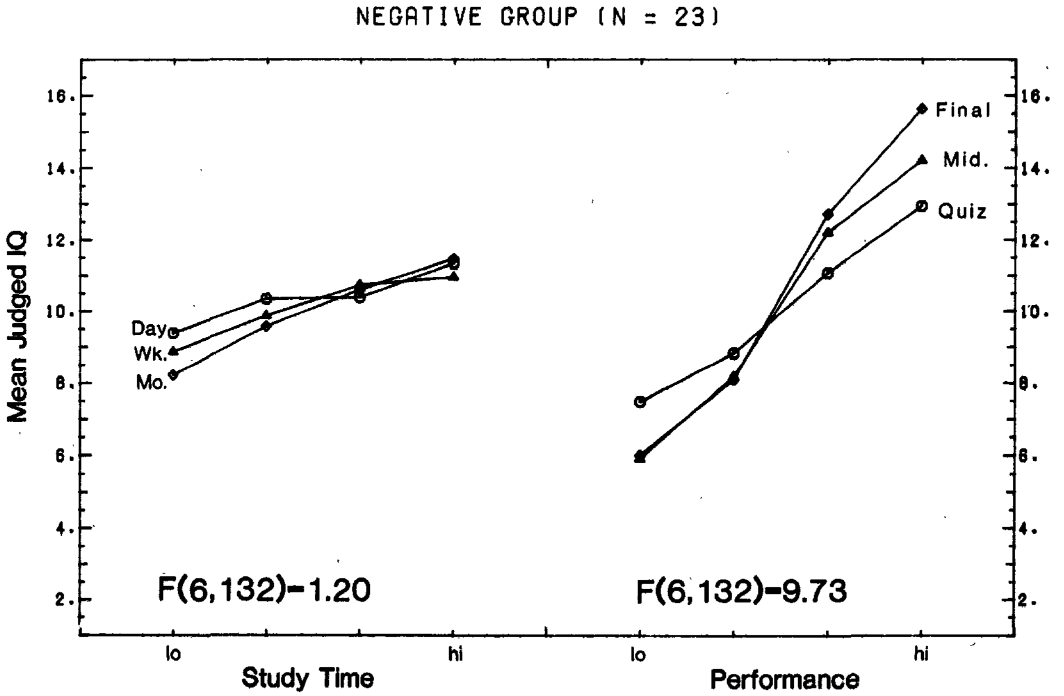

Figure 6. Mean judged IQ for the negative slope group of Experiment 1 for the Study Time × Study Time Reliability design (left-hand panel) and the Performance × Performance Reliability design (right-hand panel).

44.41, Performance Reliability × Study Time, $F(6, 282) = 7.41$, and Study Time Reliability × Performance, $F(6, 282) = 7.50$, interactions all showed crossovers similar to those shown in Figures 4 and 5. Study time was positively related to IQ when performance information was omitted, $F(3, 141) = 18.76$, and the effect of performance was larger when presented alone than when combined with study time.

Assuming that the weights are all positive, the data of the negative slope groups contradict two qualitative predictions of the positive-weight averaging model: (a) As the reliability of study time (or performance) increases, the net effect of performance (or study time) should decrease, but it actually increases, and (b) the direction of the effect of study time when performance information is omitted should be the same as when combined with performance, but it is the opposite. Further model analyses for the negative slope groups are deferred until after the results of the study time judgments.

Study Time Judgments

A 3(IQ reliability) × 4(IQ) × 3(performance reliability) × 4(performance) ANOVA showed surprisingly small, though statistically significant, main effects of IQ in Experiment 1, $F(3, 135) = 5.64$, and Experiment 2, $F(3, 300) = 15.45$. Subjects were classified according to the same criteria used for the IQ judgments, but with the IQ variable replacing the study time variable. In Experiment 1, 10 were classified as showing positive slopes, 34 as showing negative slopes, and 2 as showing inconsistent or flat slopes. For Experiment 2 the numbers of subjects were 21, 77, and 3, respectively. The ANOVAS, excluding the two extreme IQ values, showed significant Slope Group × IQ interactions—Experiment 1, $F(1, 42) = 189.56$, Experiment 2, $F(1, 96) = 295.4$—providing evidence that the grouping reflects genuine differences in use of IQ information. A chi-square test of independence shows no difference between Experiments 1 and 2 in the proportion of subjects in each slope category, $\chi^2(2) < 1.0$. The reported slope on the questionnaire at the conclusion of Experiment 2 showed that the slope groups differed significantly in the predicted direction on all four questions (all $ps < .01$).

Positive Slope Group

The results of the positive slope groups for the study time judgments show the qualitative effects predicted by the averaging model. Figure 7 shows the IQ × IQ Reliability (left-hand panel) and Performance × Performance Reliability (right-hand panel) interactions in Experiment 1. In both panels, the high-reliability curve is steepest and the low-reliability curve is flattest. Figure 8 shows the IQ × Performance Reliability (left-hand panel) and Performance × IQ Reliability (right-hand panel) interactions in Experiment 1, which also show the crossover form predicted by the averaging model (the high-reliability curves are flattest and the low-reliability curves are steepest). Graphs of the results of the partial-information conditions also showed crossover interactions. In addition, the effects of performance and IQ presented alone were clearly greater than when combined, as predicted by the averaging model.

In Experiment 2 the interactions of IQ × IQ Reliability, $F(6, 120) = 20.05$, Performance × Performance Reliability, $F(6, 120) = 31.99$, IQ × Performance Reliability, $F(6, 120) = 14.40$, and Performance × IQ Reliability, $F(6, 120) = 10.67$, all showed the crossover form predicted by the averaging model and were similar to the results for Experiment 1 shown in Figures 7 and 8. In addition, both IQ and performance had much larger effects when presented alone than when combined, as predicted by the averaging model.

The averaging model of Equation 2 was fit to the mean study time judgments of the positive slope group with the value of W_O set equal to 1.0. The square root of the average squared deviation was 0.599 in Experiment 1 and 0.413 in Experiment 2. The standard errors ranged from 0.203 to 1.237. Figure 9 presents the mean judged study time (symbols) for the 144 stimuli of the main design of Experiment 2 and the predictions of the averaging model obtained from STEPIT (solid lines).

The correlational model of Equation 4 was also fit to the mean study time judgments of the positive slope groups. Assuming r_{PA} to be a constant, the correlational model fit slightly worse than the averaging model (square roots of average squared deviation equaled 0.629

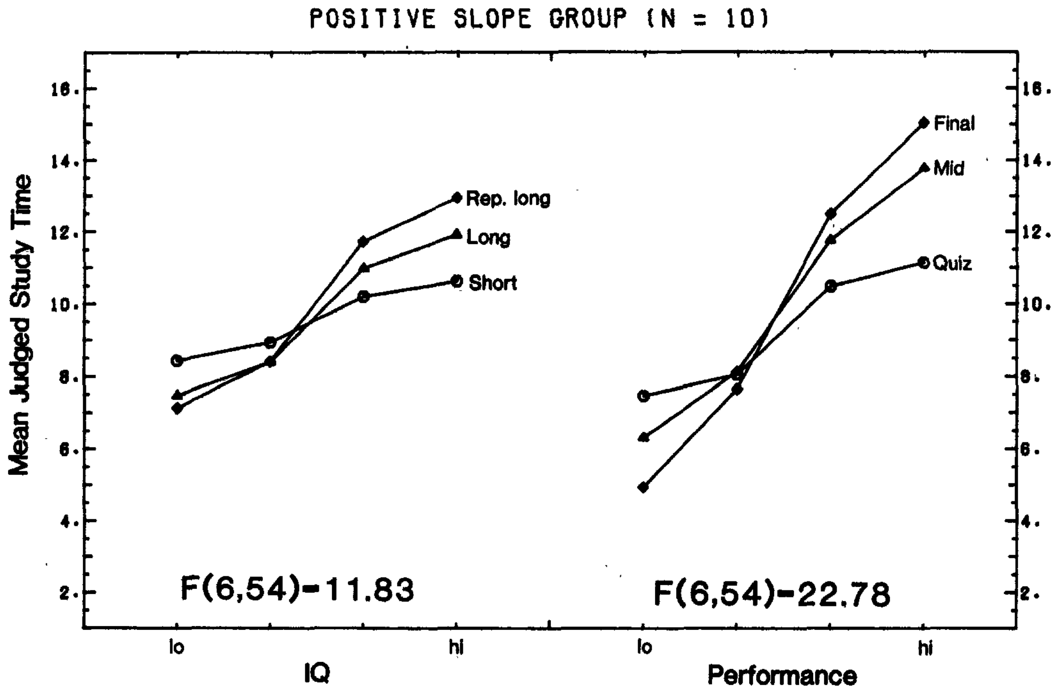

Figure 7. Mean judged study time for the positive slope group of Experiment 1 as a function of IQ and IQ reliability (left-hand panel) and performance and performance reliability (right-hand panel).

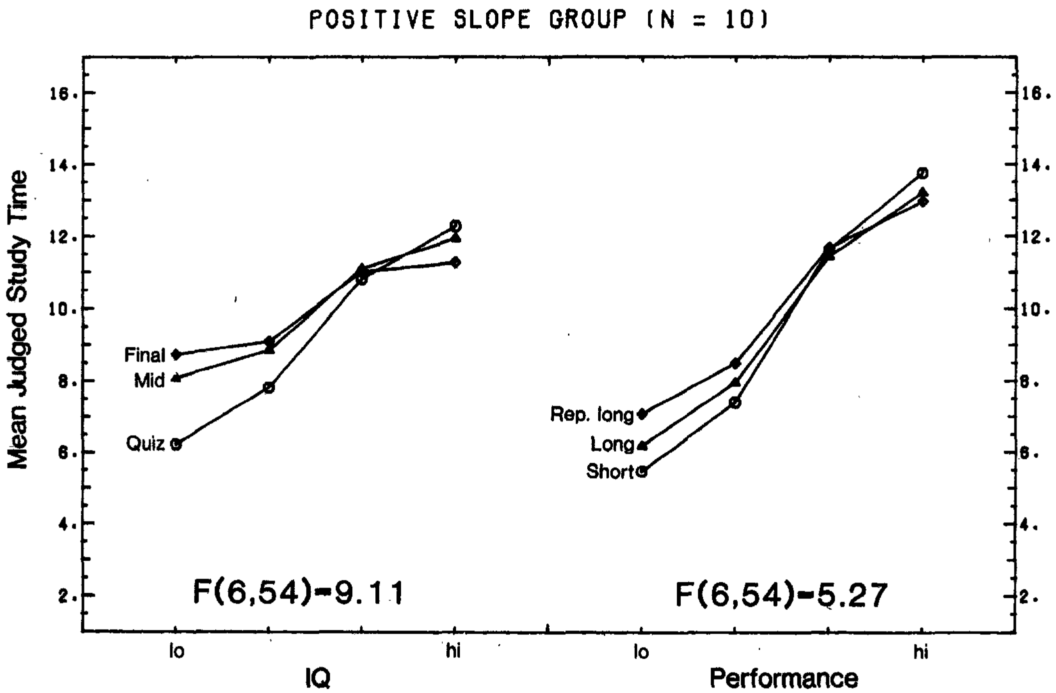

Figure 8. Mean judged study time for the positive slope group of Experiment 1 as a function of IQ and performance reliability (left-hand panel) and performance and IQ reliability (right-hand panel).

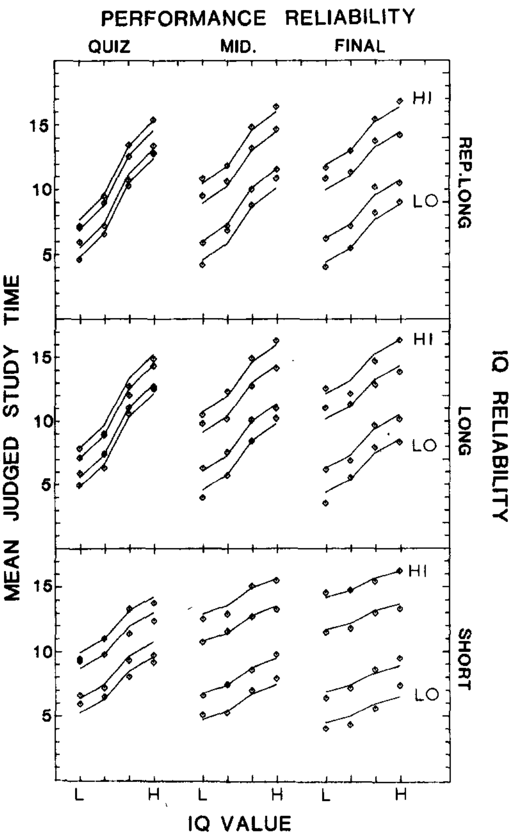

Figure 9. Mean judgments of study time (symbols) and averaging model predictions (solid curves) for the positive slope group of Experiment 2. (Each panel represents a different combination of performance reliability and IQ reliability. In each panel the abscissa variable is the value of IQ and each curve is a different value of performance.)

and 0.436 for Experiments 1 and 2, respectively). Allowing r_{PA} to vary with information reliability allowed the correlational model to fit slightly better than the averaging model of Equation 2 (0.600 and 0.388 for Experiments 1 and 2, respectively). The configural-weight averaging model, which requires estimation of 24 parameters, fit slightly better than the correlational model (0.536 and 0.389 for Experiments 1 and 2, respectively).

Negative Slope Group

Figures 10 through 12 show the study time judgments of the negative slope group of Experiment 1, which are similar to the results of the negative slope IQ groups. The left-hand panel of Figure 10 shows the IQ × IQ Reliability interaction, which has the expected crossover form. The right-hand panel of Figure 10 presents the Performance × Performance Reliability interaction, in which the high-reliability performance curve shows the steepest slope and the low-reliability performance curve shows the flattest slope. Figure 11 presents evidence that eliminates the positive-weight averaging model of Equation 2. The left-hand panel shows the Performance Reliability × IQ interaction, and the right-hand panel shows the IQ Reliability × Performance interaction. In both panels the high-reliability curve shows the steepest slope and the low-reliability curve shows the flattest slope, contrary to Equation 2.

Figure 12 presents the results of the two partial-information designs in Experiment 1. In the left-hand panel, the IQ curves show an upward slope, $F(3, 99) = 10.08$, an effect that is the opposite of that shown in Figure 10. As in the IQ judgments, a few subjects judged study time to decrease as IQ increased (8 out of 34 in Experiment 1 and 15 out of 77 in Experiment 2) in the partial-information conditions. The results of the Performance × Performance Reliability design (right-hand panel) shows a crossover interaction of the expected form, and the ordinate variation is greater than in Figure 10 where performance is combined with IQ.

In Experiment 2, the four two-way interactions replicated Experiment 1: IQ × IQ Reliability, $F(6, 456) = 85.57$; Performance × Performance Reliability, $F(6, 456) = 87.19$; IQ × Performance Reliability, $F(6, 456) = 26.41$; and Performance × IQ Reliability, $F(6, 456) = 13.59$. As in Experiment 1, the crossover for the IQ × Performance Reliability and Performance × IQ Reliability interactions contradicted the predictions of the averaging model. In the partial-information conditions, IQ was positively related to judged study time, $F(3, 228) = 24.16$, and the effect of performance was larger than when combined with IQ information.

Model Analysis for Negative Slopes

The finding that an increase in the reliability of one type of information enhanced the effect of the other information in the judgments of the negative slope subjects clearly eliminates

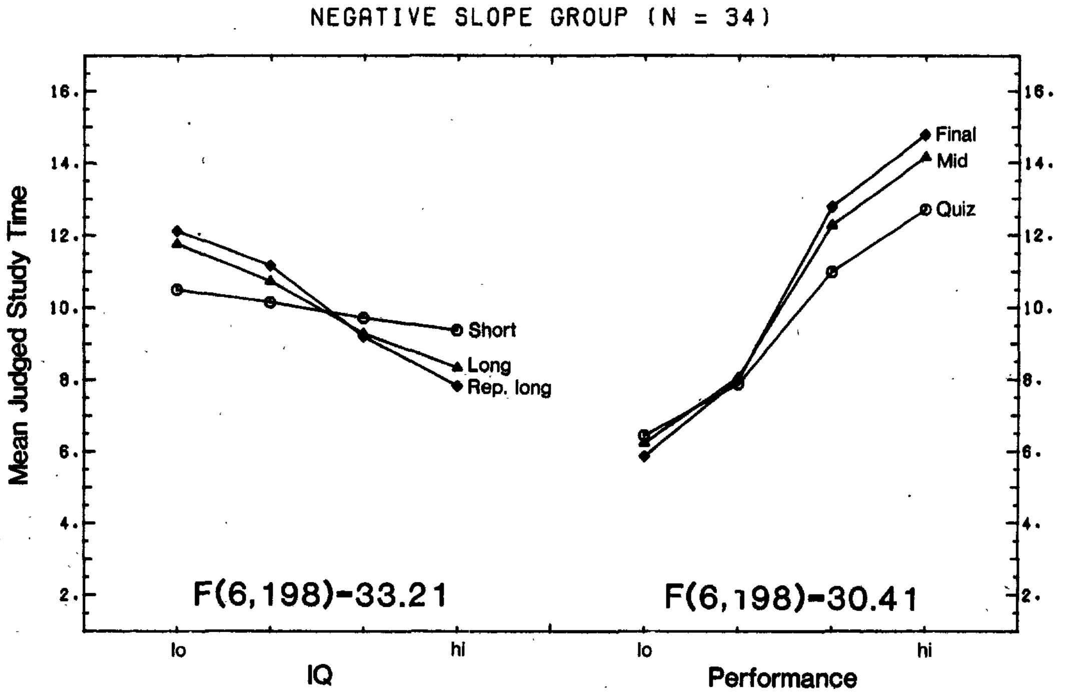

Figure 10. Mean judged study time for the negative slope group of Experiment 1 as a function of IQ and IQ reliability (left-hand panel) and performance and performance reliability (right-hand panel).

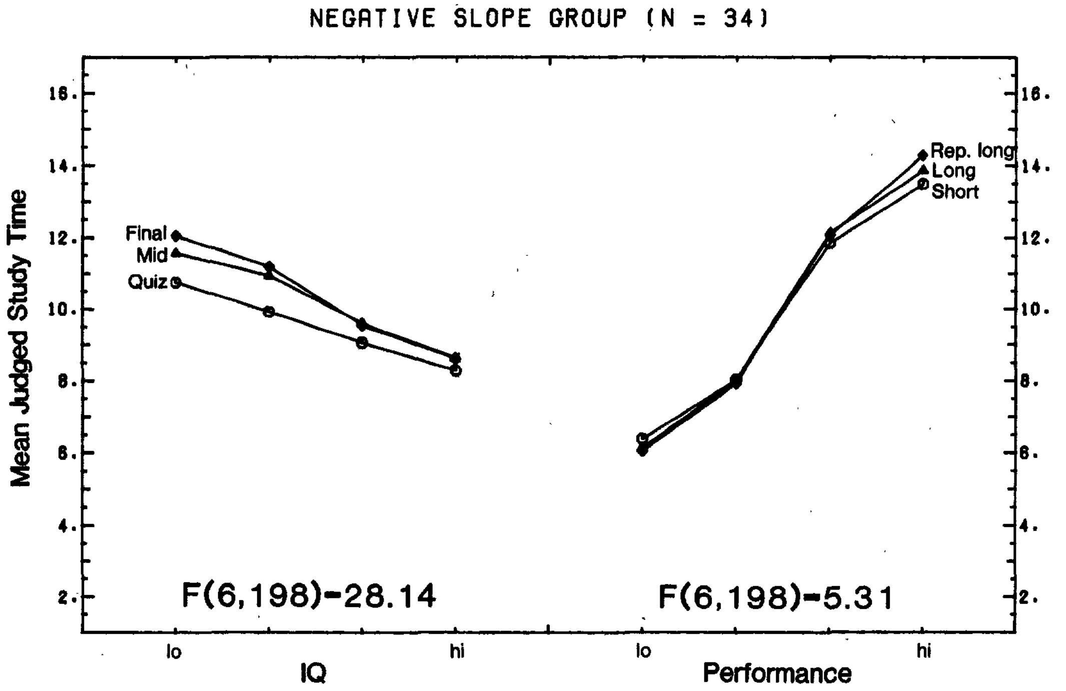

Figure 11. Mean judged study time for the negative slope group of Experiment 1 as a function of IQ and performance reliability (left-hand panel) and performance and IQ reliability (right-hand panel).

only the positive-weight averaging model. This result has potential to provide insight into the different inference processes used by negative and positive slope subjects and requires detailed consideration of the models and their predictions.

Negative-Weighted Averaging for Negative Slopes?

As hypothesized by Kepka and Brickman (1971), it may be possible to predict the results of the negative slope groups by allowing negative weights for IQ and study time. Assume that the scale values of ability and effort increase as the verbally described levels increase, that the values of W_A and W_E in Equations 1 and 2 are negative, and that whenever performance is also presented the sum of the weights is positive. Under these assumptions, the averaging model can predict negative slopes because the effective weights of ability and effort (e.g., $W_E/[W_E + W_P + W_O]$) will be negative. The negative-weight averaging model also makes two other correct predictions. First, as the reliability of IQ or study time increases, the net effect of performance is predicted to increase (right-hand panels of Figures 5 and 11) if the absolute values of W_A and W_E increase as the reliability of ability and effort increase. Thus, the value of the denominator (e.g., $W_A + W_P + W_O$) will decrease as the reliability of ability or effort increases, and the net weight of performance (e.g., $W_P/[W_A + W_P + W_O]$) will increase. Second, the negative-weight averaging model can also predict positive slopes when performance information is omitted if the absolute values of W_A and W_E are greater than W_O. If this is true, then when performance is omitted the effective weight of effort or ability will be positive because both the numerator and denominator will be negative.

Unfortunately, the negative-weight averaging model also makes three incorrect predictions. First, it predicts that the net effect of study time or IQ will decrease as the reliability of performance increases. As the value of W_P increases, the denominator sum will increase, decreasing the effect of the numerator variable. The interactions in the left-hand panels of Figures 5 and 11 contradict the prediction. Sec-

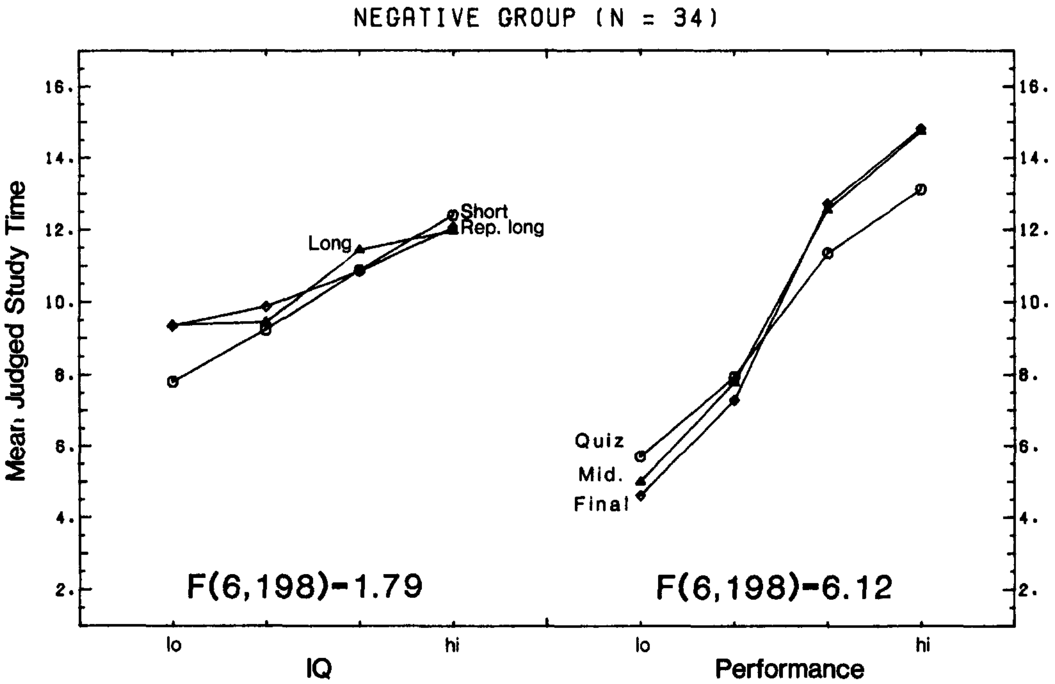

Figure 12. Mean judged study time for the negative slope group of Experiment 1 for the IQ × IQ Reliability design (left-hand panel) and Performance × Performance Reliability design (right-hand panel).

ond, the negative-weight averaging model predicts that when performance is omitted, the effect of study time on IQ (and vice versa) should be greatest when its own reliability is lowest. This prediction is both counterintuitive and contrary to the data shown in the left-hand panels of Figures 6 and 12. Third, the negative-weight averaging model predicts that the net effect of performance will decrease as the reliability of performance increases, contrary to the results in the right-hand panels of Figures 4 and 10.

The latter two predictions of the negative-weight averaging model can be derived from the limits of the effective weight terms. For example, when performance is omitted, the effective weight of ability is $W_A/(W_A + W_O)$. When both W_A and $W_A + W_O$ are negative, but W_O is greater than zero, the value of $W_A/(W_A + W_O)$ must be greater than 1. The limit of $W_A/(W_A + W_O)$ as W_A approaches negative infinity is 1.0. Therefore, the value of $W_A/(W_A + W_O)$ and the effect of ability will decrease toward 1.0 as $|W_A|$ increases. Analogous logic can be used to show that $W_P/(W_P + W_A + W_O)$ also decreases toward 1.0 as W_P increases. Thus, neither a negative-weight nor a positive-weight averaging model can provide an adequate account of the data of the negative slope groups, and so no attempt was made to fit either model to the data.

Correlational Model for Negative Slopes?

Cue intercorrelation fixed. First, the correlational model is examined when the correlation between the cues is assumed to be fixed over variation in reliability (i.e., r_{PE} in Equation 3 or r_{PA} in Equation 4). If one assumes that the scale values of study time and IQ increase as their verbally described levels increase, then it is necessary to assume that $r_{AE} < r_{PA}r_{PE}$. When $r_{AE} < r_{PA}r_{PE}$ the net effect of ability or effort will *increase* as the reliability of performance increases (assuming that r_{PA} or r_{PE} increases as performance reliability increases) because the absolute value of $r_{AE} - r_{PE}r_{PA}$ will increase. Thus, Equations 3 and 4 can predict the interactions in the left-hand panels of Figures 5 and 11, interactions that cannot be predicted by either the negative-weight or positive-weight averaging model.

It is also necessary for a successful model to predict that the effect of study time will increase as study time reliability increases (left-hand panel of Figure 4) and that the effect of IQ will increase as IQ reliability increases (left-hand panel of Figure 10). This prediction necessitates values or r_{AE} that *decrease* as the reliability of IQ or study time increases because smaller r_{AE} values result in larger absolute values of $r_{AE} - r_{PE}r_{PA}$. The decreasing values of r_{AE} also allow the correlational model to predict correctly the interactions of Study Time Reliability × Performance (right-hand panel of Figure 5) and IQ Reliability × Performance (right-hand panel of Figure 11). For example, performance is predicted to have a larger effect as study time reliability increases because the value of r_{AE} decreases, thus increasing the value of $r_{PA} - r_{AE}r_{PE}$.

Unfortunately, the correlational model also makes a qualitatively incorrect prediction for judgments when performance information is omitted. Specifically, if the values of r_{AE} decrease as the reliability of study time or IQ increases but remain greater than zero, then the effect of study time and IQ will decrease as their reliability increases when performance information is omitted. This was not the case.

In summary, the correlational model with fixed cue intercorrelation accounts for two important effects in the judgments of the negative slope groups: (a) The effect of information increases as its own reliability increases, and (b) the effect of information increases as the reliability of other information increases. These effects cannot both be predicted by either a positive- or a negative-weight averaging model. Equations 3 and 4 were fit to the 168 mean IQ and study time judgments of the negative slope groups. The model requires estimation of 16 parameters. The scale values were assumed to increase as their verbally described levels increased, and the cue intercorrelation was estimated as a constant. The square root of the average squared deviations were 0.472 and 0.466 for IQ judgments and 0.628 and 0.570 for the study time judgments of Experiments 1 and 2, respectively. Nevertheless, the model makes qualitatively incorrect predictions in the partial-information conditions in which performance information is omitted.

Cue intercorrelation variable. If the cue intercorrelation (r_{PE} in Equation 3 and r_{PA} in Equation 4) is assumed to vary with the manipulations of reliability, the predictions of the correlational model are more complex. In order to predict negative slopes, it is still necessary to assume that $r_{AE} < r_{PA}r_{PE}$. The net effect of ability or effort can be increased by either increasing the reliability of performance or increasing the cue intercorrelation. Because the cue intercorrelation influences the net effect of ability or effort through both the numerator and denominator, it is possible to predict that the effect of ability or effort will increase with its own reliability without the values of r_{AE} decreasing with increasing reliability. This means that the correlational model with a variable cue intercorrelation can predict the major qualitative effects in the judgments of the negative slope groups, including the judgments when performance information is omitted.

The correlational model was fit to the data of the negative slope groups, allowing the cue intercorrelation parameter to vary with the reliability of both given cues. The model requires estimation of 24 parameters and fit the data better than the correlational model with fixed cue intercorrelation (square roots of the average squared deviation were 0.382 and 0.347 for IQ judgments and 0.602 and 0.496 for study time judgments in Experiments 1 and 2, respectively).

The estimated values of r_{AE} increased slightly with the reliability of IQ or study time, as expected. The estimated cue intercorrelations increased with increases in the reliability of IQ or study time, but decreased with increases in the reliability of performance. This pattern of estimated cue intercorrelations seems intuitively implausible but is required for the model to account for the data. The model could be more completely tested in a future experiment by having subjects make judgments that would help constrain the cue intercorrelation parameters. For example, subjects could be asked to judge IQ, given study time and performance that vary in reliability, as in the present experiment. In addition, the same subjects could be asked to judge (a) performance on quiz, midterm, and final given study time information that varies in reliability and (b) 1-day, 1-week and 1-month study time, given performance information that varies in reliability. According to the correlational model, these judgments should be predictable from the cue intercorrelations.

Expectancy-Contrast Model for Negative Slopes?

Lopes (1972) proposed that in social attribution subjects may form an expectancy for the value of one variable based on the other given information. The expectancy is then compared with the given value of the variable, and the contrast or discrepancy influences the judgments. For example, in judging IQ the subject may form an expectancy for study time based on the performance information. The expected study time information is then compared with the given value of study time, and the discrepancy influences the subject's judgment of IQ. This type of expectancy-contrast process can yield a negative slope for judgments as a function of IQ or study time (see also Birnbaum, 1975).

Lopes proposed a particular model in which the discrepancy between the expected and actual value of a cue is weighted and averaged with the directly given information. In this model, the weight given to the discrepancy would be expected to increase as the reliability of either variable involved in the discrepancy term increases. For example, the discrepancy between expected and actual study time should have more influence when one is more confident in the discrepancy, and confidence in the discrepancy should increase with the reliability of either performance or study time.

The general form of Lopes's model for the present case can be written as follows:

Ability =

$$[W_P S_P + W_{ST} S_{ST} + W_D(\hat{S}_{ST} - S_{ST}) + W_O S_O]/(W_P + W_{ST} + W_O) \quad (6)$$

Study Time =

$$[W_P S_P + W_{IQ} S_{IQ} + W_D(\hat{S}_{IQ} - S_{IQ}) + W_O S_O]/(W_P + W_{IQ} + W_O), \quad (7)$$

where $\hat{S}_{IQ}$ and $\hat{S}_{ST}$ are the anticipated values of IQ and study time, and W_D is the weight given to the discrepancy between anticipated and actual study time or IQ, and the other terms are as defined for Equations 1 and 2. The values of $\hat{S}_{ST}$ and $\hat{S}_{IQ}$ are assumed to depend on the corresponding values of S_P, and the values of W_D are assumed to increase monotonically with both W_P and W_{ST} (or W_{IQ}). Under these assumptions, the model can predict all the qualitative effects observed in the negative slope groups.

In order to fit the model to the negative slope group data, the terms $W_{ST}S_{ST}$ and $W_{IQ}S_{IQ}$ were omitted from the numerators, the values of $\hat{S}_{ST}$ and $\hat{S}_{IQ}$ were assumed to equal S_P, W_D was assumed to equal a constant multiplied by the product of W_P and W_{ST} (or W_{IQ}), and W_O was set equal to 1.0.[3] The model uses 16 parameters, fit the mean judgments of the negative slope groups better than the correlational model with fixed intercorrelation, fit as well or better than the correlational model with variable cue intercorrelations, yielded intuitively plausible parameters (estimated weights increased with reliability), and made no serious qualitative errors (square roots of the average squared error were 0.390 and 0.380 for IQ and 0.601 and 0.498 for the study time judgments for Experiments 1 and 2, respectively). The predictions of the expectancy-contrast model for the negative slope groups of Experiment 2 are presented in Figures 13 (IQ judgments) and 14 (study time judgments) along with the actual means (symbols).

The expectancy-contrast model is interesting because it provides a possible mechanism for the unexpected effects of information reliability in the negative slope groups and for the fact that the majority judge IQ and study time to be positively related when performance is omitted (there is no explicit information on which to base an expectancy for study time or IQ in these cases). The minority who judge IQ and study time to be negatively related when performance is omitted might be supposed to form an expectancy for those variables, perhaps using the value of the initial impression (S_O) for their expectancy. Because the correlational model with varying cue intercorrelation and the expectancy contrast model fit the data of the negative slope groups almost equally well, neither model can be definitely eliminated in the present context.

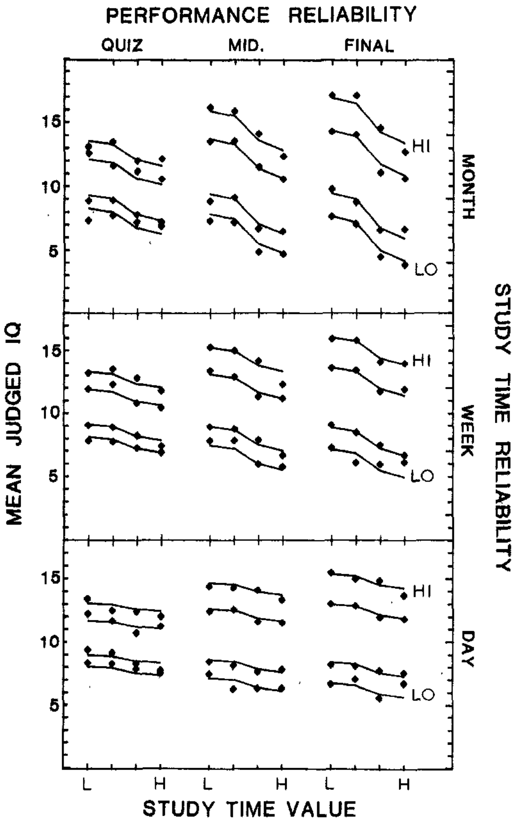

Figure 13. Mean judgments of IQ (symbols) and expectancy-contrast model predictions (solid curves) for the negative slope group of Experiment 2. (Each panel represents a different combination of performance reliability and study time reliability. In each panel the abscissa variable is the value of study time and each curve is a different value of performance.)

Discussion

The difference between groups in the judged relationship between ability and effort shows two distinctly different ways of using information to make inferences of ability and effort. The positive and negative slope groups differ not only in the judged relationship between IQ and Study Time but also in the way these variables are combined with performance. For the positive slope groups, increasing the reliability of one cue decreases the effect of the other cue. In contrast, increasing the reliability of one cue *increases* the effect of the other cue for the negative slope groups. The effect of information reliabilities in the negative slope groups was surprising, but seems plausible when considered in terms of the expectancy-contrast model of inference.

The expectancy-contrast model for the negative slope results deserves further research.

[3] Initially, a separate value of W_D was estimated for each of the nine information reliability combinations. When graphed, the estimated values appeared to be an approximately multiplicative function of W_P and W_{ST} (or W_{IQ}).

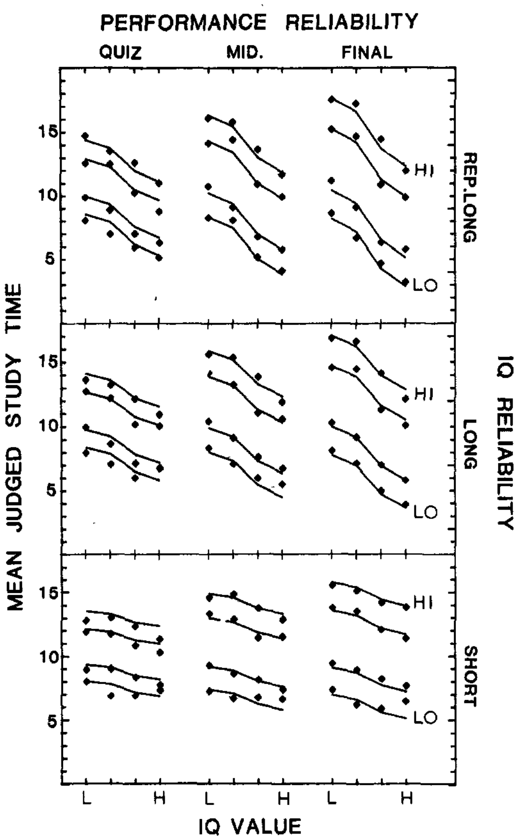

Figure 14. Mean judgments of study time (symbols) and expectancy-contrast model predictions (solid curves) for the negative slope group of Experiment 2. (Each panel represents a different combination of performance reliability and IQ reliability. In each panel the abscissa variable is the value of IQ and each curve is a different value of performance.)

First, it captures some of the essential aspects of traditional attribution theories (Lopes, 1972). Second, although it is intuitively plausible that the weight given to the discrepancy term should increase as the reliabilities of the variables on which it is based increase, there appears to be no logical necessity for this weighting pattern. It would be interesting to know whether the interactions of information reliability obtained here would occur for other types of causal inferences. Third, one might wonder whether the expectancy-contrast processing step is a feature of all causal inferences.

Individual Differences in Slope

Past research on inferences of ability and effort has found that only a small portion of subjects in college age samples did not show negative slopes, although children commonly show individual variation in slope (Surber, in press-a, in press-b). Surber (1980) reported that 88% and 83% of adults showed negative slopes in judging effort and ability, respectively. Anderson and Butzin (1974) did not report any variation in slope (though they may not have examined the individual data with this in mind), and Karabenick and Heller (1976) found that 88% and 91% of adults showed negative slopes in judging effort and ability. In contrast, the present study found 75% and 48% negative slope subjects for judgments of effort and ability. These between-study differences require explanation.

Processing Load Hypothesis

One possible source of the between-experiment differences in slope is that the addition of the reliability information may leave less cognitive work space for formulating one's impression of IQ or study time. Shatz (1978) hypothesized that when the processing load is increased, subjects may shift to using more routinized strategies in order to perform the whole task adequately, and Payne (1982) recently reviewed evidence showing that increasing task demands (as indexed by number of alternatives, number of dimensions, or time pressure) causes decision makers to change strategies. Using this idea, it is necessary to hypothesize that judging two variables to be positively related is somehow simpler than judging them to be negatively related. This is plausible for two reasons. First, studies of intuitive prediction have generally shown that negative relationships between variables are more difficult to learn than positive relationships (Slovic, 1974). Second, the "cue revaluation" hypothesis of Anderson and Butzin (1974) and the expectancy-contrast model of Lopes (1972) both imply that there are more mental steps involved in judging ability and effort to be negatively related than positively related. The correlational model is equivalent for negative and positive slopes (except for the parameter values) and so does not provide a rationale for expecting negative slope judgments to be more complex.

The processing load hypothesis does not, however, explain the differences in proportion of negative slopes between judgments of ability and effort (48% vs. 75%). A second difficulty with the processing load hypothesis is that it does not predict which subjects will show positive slopes. What is needed in order to test this hypothesis is some way of manipulating and/or independently measuring the processing load imposed by use of a judgment strategy for individual subjects.

Two Concepts of Ability

A second hypothesis for the individual differences in slope is that the positive and negative slope subjects are using different concepts of ability. Dweck and Elliott (in press) proposed that individuals can conceptualize ability either as a stable trait or entity versus as a continuously growing set of skills. If ability is conceptualized as a growing set of skills, then ability and effort should be judged to be positively related. For example, if a person has high ability he or she must have worked hard to acquire that ability. Conversely, if a person works hard, he or she must be acquiring high ability. In contrast, the trait concept of ability should lead to negative slope judgments. If ability is a fixed trait in each individual, then it must vary inversely with effort for individuals of equal performance. This explanation also requires specifying what individuals will use which concept of ability and when (see Dweck & Elliott, in press, for some suggestions).

Inference of Missing Information?

The results of the two experiments also have relevance to Yamagishi and Hill's (1981) hypothesis that missing information is inferred on partial information trials, at least in the context of judgments of ability and effort. Consider first the negative slope group judgments of performance on trials when only performance information was given. If these subjects had inferred the missing information (either study time or IQ) to be congruent with the given performance value, we would expect the judgments of the negative slope group to be much less extreme than those of the positive slope group because the inferred value of IQ or study time should have opposite effects for the two groups. Comparison of the relevant graphs showed that the judgments based on performance alone were highly similar across slope groups. In ANOVAs of the performance-only judgments the only significant effect was a Slope Group × Performance × Performance Reliability interaction in the study time judgments of Experiment 1, $F(6, 252) = 4.21$, that did not replicate in Experiment 2. For the performance-only trials it seems reasonable to conclude that the subjects did not infer missing information.

Second, consider the judgments of the negative slope group when performance information is omitted. The majority of the negative slope subjects judged IQ and study time to be positively related when they were not given performance information. If the negative slope subjects had inferred the missing performance value, then we would expect them to show negative slopes just as when performance was presented. This did not occur for the majority, but inference of missing information remains a possibility for the minority groups.

For the positive slope groups taken alone, it is possible that inference of the missing information occurs. As shown in the introduction, however, the correlational model is incapable of distinguishing between judgments based on implicit inference versus the simple correlation between the given variable and the variable to be judged. For the present data, the hypothesis that missing information is inferred is not very parsimonious at best and inconsistent with the data at worst.

CONCLUSIONS

The present data show that there are two qualitatively different ways of making inferences of ability and effort. This finding adds to the accumulating evidence that judgment and decision-making strategies show variation and can be contingent on task demands. Second, the manipulation of information reliability and the different effects of information reliability on the positive and negative slope groups is consistent with the possibility that the negative slope inferences involve a processing step not involved in the positive slope inferences (specifically that an expectancy-contrast component may be involved). Further research on variables that influence when positive or negative slope inferences are made is needed. Such research should advance our knowledge of social inference and the processes on which it is based.

References

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

Birnbaum, M. H. (1975). Expectancy and judgment. In F. Restle, R. Shiffrin, N. J. Castellan, H. Lindman, & D. Pisoni (Eds.), *Cognitive theory: Vol. 1* (pp. 107–118). Hillsdale, NJ: Erlbaum.

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

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

Birnbaum, M. H., & Stegner, S. E. (1981). Measuring the importance of cues in judgment for individuals: Subjective theories of IQ, heredity, and environment. *Journal of Experimental Social Psychology, 17,* 159–182.

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

Dweck, C. S., & Elliott, E. S. (in press). Achievement motivation. In E. M. Hetherington (Ed.), *Carmichael's manual of Child Psychology.* New York: Wiley.

Karabenick, J. D., & Heller, K. A. (1976). A developmental study of effort and ability attributions. *Developmental Psychology, 12,* 559–560.

Kelley, H. H. (1972). *Causal schemata and the attribution process.* New York: General Learning Press.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28,* 107–128.

Kepka, E., & Brickman, P. (1971). Consistency versus discrepancy as clues in the attribution of intelligence and motivation. *Journal of Personality and Social Psychology, 20,* 223–229.

Kun, A. (1977). Development of the magnitude-covariation and compensation schemata in ability and effort attributions of performance. *Child Development, 48,* 862–873.

Lopes, L. L. (1972). A unified integration model for "Prior Expectancy and Behavioral Extremity as Determinants of Attitude Attribution." *Journal of Experimental Social Psychology, 8,* 156–160.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92,* 382–402.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29–46.

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

Shatz, M. (1978). The relationship between cognitive processes and the development of communication skills. In C. B. Keasey (Ed.), *Nebraska symposium on motivation: 1977,* (Vol. 25, pp. 1–42). Lincoln: University of Nebraska Press.

Slovic, P. (1974). Hypothesis testing in the learning of positive and negative linear functions. *Organizational Behavior and Human Performance, 11,* 368–376.

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

Surber, C. F. (1981a). Effects of information reliability in predicting task performance using ability and effort. *Journal of Personality and Social Psychology, 40,* 977–989.

Surber, C. F. (1981b). Necessary versus sufficient causal schemata: Attributions for achievement in difficult and easy tasks. *Journal of Experimental Social Psychology, 17,* 569–586.

Surber, C. F. (in press-a). The development of achievement-related judgment processes. In J. Nicholls (Ed.), *The development of achievement motivation.* Greenwich, CT: JAI Press.

Surber, C. F. (in press-b). Developmental changes in inverse compensation in social and nonsocial attributions. In S. Yussen (Ed.), *The development of reflection.* New York: Academic Press.

Yamagishi, T., & Hill, C. T. (1981). Adding versus averaging models revisited: A test of a path-analytic integration model. *Journal of Personality and Social Psychology, 41,* 13–25.

Received February 21, 1983
Revision received August 4, 1983 ■