ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 48, 169–192 (1991)

The Psychology of Moderate Prediction

II. Leniency and Uncertainty

YOAV GANZACH AND DAVID H. KRANTZ

Columbia University

In this paper we demonstrate that intuitive numerical predictions can be somewhat regressive. This moderation of predictions is asymmetric: predictions are more regressive at low than at high values of the predictor. This pattern is analyzed in terms of the operation of multiple heuristics. The *representativeness* heuristic is responsible for predictions in which extremity of the predicted variable is matched to extremity of the predictor. Matching is modified by a variety of intuitions that promote moderation per se; we lump these together under the heading of *weak regressiveness.* Third is *leniency,* a heuristic suggesting that the higher the uncertainty, the more positive should be the predictions. The first experiment demonstrates leniency in isolation from the other heuristics: in multivariate prediction, inconsistent predictors yield more positive predictions. Experiment 2 demonstrates asymmetric regression in a situation where all three heuristics are assumed to have effects. The third experiment exhibits leniency in the context of explanation of regression phenomena (rather than in numerical prediction). The final experiment explores the relation between the three heuristics and experience with multiple determination (Ganzach & Krantz, in press). It demonstrates increased moderation of predictions when subjects are required to generate a predicted value of an intermediate variable, for example, when the prediction of GPA from Intelligence is made subsequent to a prediction of Motivation.

INTRODUCTION

Leniency, Uncertainty, and Multiple Heuristics in Prediction

Early work on judgmental heuristics usually isolated particular ones for study (for a summary see Kahneman, Slovic, & Tversky, 1982). Later research demonstrated that a variety of heuristics might compete in the control of judgments (Agnoli & Krantz, 1988; Nisbett, Fong, Lehmann, & Cheng, 1987). Some of the competing heuristics are nonstatistical (Kahneman & Tversky, 1972) but some are statistical heuristics—rules of thumb that resemble in some respects normative statistical rules (Nisbett, Krantz, Jepson, & Kunda, 1983). In this paper we demonstrate how three heuristics operate simultaneously in the context of numerical predictions.

Reprint requests should be sent to Yoav Ganzach, Graduate School of Business Administration, The Hebrew University, Jerusalem, Israel 91905.

0749-5978/91 $3.00

The representativeness heuristic. According to this heuristic, people base their predictions on some intuitive estimation of the dispersion of the predictor and the dispersion of the outcome. Based on this estimation, a *matching strategy* is used to make the predictions. The predicted value is chosen so that its extremity (deviation from central tendency) matches the extremity of the predictor.

The matching strategy leads to systematic differences between intuitive predictions and normative predictions. For most commonly encountered bivariate distributions, normative predictions are regressive: The position of the predicted value on the distribution of the outcome is less extreme than the position of the predictor on its distribution, and the less valid the predictor the less extreme the prediction. On the other hand, intuitive predictions are—at least to a first approximation—nonregressive. The position of the predicted value on the distribution of the outcome is roughly equal to the position of the predictor on its distribution, and the validity of the predictor is, to a large extent, ignored.

Kahneman and Tversky (1973) demonstrated the nonregressiveness of intuitive predictions in a classical experiment. They asked three groups of subjects to predict numerical GPA (Grade Point Average on a 1 to 4 scale) from 11 values, given in percentiles, of three different predictors. The predictors were the GPA itself expressed in percentile (which is a perfect predictor of numerical GPA), "Mental Concentration Test" score (which was presented to the subjects as only a partially valid predictor), or "Sense of Humor" score (which was presented as having very low validity). The results showed that there was no difference in the predictions from the first two predictors, and that predictions from Sense of Humor were only slightly less extreme than the predictions from the other two predictors.

Since the representativeness heuristic provides a good first approximation of the process of intuitive numerical prediction, not much attention was given in the prediction literature to the investigation of the details of these predictions. These details are investigated in the current paper. We show that intuitive predictions deviate systematically from predictions by representativeness. We also show that, except for representativeness, two additional heuristics—weak regressiveness and leniency—are involved in intuitive prediction. As will be discussed below, there were indications for the operation of these heuristics even in Kahneman and Tversky's (1973) original experiment. However, in the analysis and interpretation of the results, these indications were treated as "noise," since the main focus was the representativeness heuristic. In the present paper the focus is on how these three heuristics jointly operate to affect intuitive predictions.

Weak regressiveness. Under certain conditions intuitive predictions show some recognition of the regression principle: Experience in a certain domain leads to recognition of regression effects in that domain (Nisbett *et al.*, 1983); experience with multiple determination of an outcome leads to intuitive predictions which are regressive (Ganzach and Krantz, in press). We lump the family of heuristics that lead to such a utilization of the regression rule under the heading of weak regressiveness. Note also that while the Kahneman & Tversky (1973) experiment described above is considered a classical demonstration that people lack intuition about regression, it can be considered also a demonstration that people do possess some intuition about the regression rule, since subjects did regress the predictions from Sense of Humor against the predictions from Mental Concentration. In this sense, it depends on which half of the glass one looks.

The leniency heuristic. We use this term for an intuitive prediction rule suggesting that the higher the subjective uncertainty about the prediction (hereafter uncertainty), the more lenient should be the predictions. The expression "to give the benefit of the doubt" is a good description of this heuristic. It suggests that the higher the doubt (uncertainty), the more benefit one should give (the more lenient one should be).[1]

As an example, consider the relationship between leniency and uncertainty in a *univariate prediction task*—a prediction task that is based on a single predictor. In this type of prediction, changes in validity are the source of changes in uncertainty. According to the leniency heuristic, validity should be negatively related to leniency: the lower the validity of the predictor the more lenient the prediction. This relationship was indeed observed in Kahneman & Tversky's (1973) experiment described above: The mean over the 11 predictions of GPA from sense of humor was higher than the mean over the 11 predictions of GPA from Mental Concentration, which in turn was higher than the mean of the 11 predictions from percentile GPA.

In this paper we go beyond this example. First we show that in a *multivariate prediction task*—a prediction task that is based on more than one predictor variable—internal inconsistency between the predictors results in leniency. Since internal inconsistency is also a form of uncertainty, the relationship between leniency and validity can be understood as a particular instance of a general relationship between leniency and uncertainty. Second, we show that an accurate understanding of the influence of uncertainty on intuitive predictions in a univariate prediction

[1] This expression may be not only a description of the leniency heuristic but also a cultural reflection of intuition about how prediction should be made, as well as a prescription available to people when they make prediction.

task requires the investigation of the joint operation of leniency, representativeness and weak regressiveness, and therefore the examination of predictions from different values of the predictor, rather than the examination of their mean.

Combining Heuristics: Asymmetry between Low and High Values of the Predictor in Univariate Prediction Task

Figure 1a presents a model for the joint operation, of representativeness, weak regressiveness and leniency in a univariate predietion task. The model shows the effect of these three heuristics on intuitive predictions from low and from high values of the predietor. The first approximation of the value of the prediction is given by the representativeness heuristic. The *relative value* (i.e., value measured in *Z* seores or in percentiles) of the prediction on the outcome scale is roughly equal to the relative value of the predictor on its scale (for example, if the value of the predictor is 2 standard deviations above the mean on the predictor's scale, the prediction will be roughly equal to 2 standard deviations above the mean on the outcome seale). However, the joint influence of weak regressiveness and leniency produces a strong moderation on predictions from low values of the predictor, but weak moderation on predictions from high values. The reason is that at low value of the predictor, both weak regressiveness and leniency produce increases from the matching value generated by representativeness. At high values of the predictor, they tend to cancel each other. Thus, this model suggests that at low values of the predietor intuitive predictions are less extreme than the predictor. On the other hand, at high values of the predictor, the extremity of predictions more nearly matehes the extremity of the predictor.

Since leniency increases with uncertainty, the combined effeet of these three heuristies depends on the validity of the predietor. This is illustrated in Fig. 1b. At a low value of the predictor, an increase in uncertainty—deerease in validity—leads to less extreme predictions since both leniency and weak regressiveness tend to increase the value of predietions. At a high value of the predictor, decrease in predictor validity has little effect on the extremity of predictions, despite the fact that leniency and weak regressiveness both increase, since they operate in opposition. Our model therefore suggests an asymmetry in the influence of validity.

The lengths of the arrows in Fig. 1 are the same for low and high values, as though leniency did not interact with value of the predictor. On a bounded outcome scale, this can scarcely be true: the higher the prediction based on representativeness, the less room there is for leniency to operate. Conceivably, the size of the leniency effect might be some fixed proportion of the difference between the prediction based on representa-

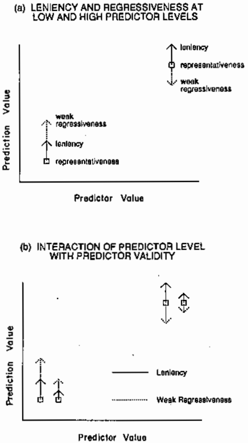

FIG. 1. Combined effects of leniency and weak regressiveness. (a) Synergistic effects at a low value of the predictor, opposing effects at a high value. (b) The left (large) arrows depict the effects for a predictor with low perceived validity; the small arrows, for high perceived validity.

tiveness and the ceiling. Although such an interaction does not affect the qualitative pattern of mean predictions derived from Fig. 1, it will affect the details, including the anticipated variability of predictions. Variability of predictions depends on the variability of application of each of the three heuristics, and the variability of each is apt to vary monotonically with the mean effect. In particular, where there is little room for leniency, near the ceiling of an outcome scale, there is also little room for variability in leniency. On the other hand, at low value of the predictor, leniency could vary considerably across subjects. Therefore, we expect high (low) variability of predictions at low (high) values of the predictor.

EXPERIMENT 1

Internal Inconsistency and Leniency

In a multivariate prediction task subjective uncertainty may depend not only the kind of predictor and outcome variables in the problem but also on the internal inconsistency of the predictors (Slovic, 1966). For example, predicting GPA for a student whose SAT percentile is 90 and whose Achievement Tests percentile is 10 entails more uncertainty then predicting GPA for a student whose SAT and Achievement tests percentiles are 55 and 45, respectively. Therefore, uncertainty is varied in this experiment by varying the internal inconsistency of the information. We test if such variations affect leniency of prediction.

The data analyzed here are unpublished data from the first paper in this series (Ganzach & Krantz, in press). In Experiment 1 of this paper, 55 subjects, students at Columbia University, were given 44 fictional vignettes, allegedly describing Columbia University students. On the basis of these vignettes, subjects predicted the students' numerical GPAs at graduation. The vignettes were presented in a fixed order for all subjects and there was no time limit and no feedback.

The format of the vignettes is illustrated by the following example:

> "J. is 18. He is Catholic. He has one sister. His parents are divorced. His mother has 11 years of education and his father has 12 years of education. His family's income is $70,000 a year. His personal Statement was evaluated as average. His SAT score is in the 23 percentile of the Columbia population and his Achievement Test score is in the 59 percentile."

In the vignettes, the Achievement Test (ACH) percentile and the SAT percentile were drawn randomly and independently from a uniform distribution over (1, 2, . . . , 100), and their correlation in the vignettes did not differ significantly from zero ($p > .8$). The personal details were also random and independent. For example, there were three levels of Personal Statement (PS), low, average, and high, each appeared in about a third of the vignettes, and there was no association between this variable and SAT or ACH (one way ANOVA with PS as independent variable revealed p values higher than .4 and .8 for SAT and ACH, respectively).

The vignettes were constructed so that two variables—SAT and ACH—would have a high perceived validity as predictors of GPA, and the inconsistency between them would serve as a measure for uncertainty; a third variable—PS—would have moderate perceived validity, and the rest of the variables would have very low perceived validity. An examination of the subjects' predictions revealed that these were indeed the perceived validities of the variables. By regressing the mean predictions (over subjects) of the vignettes on these variables it was found that the incremental variances explained by SAT, ACH, and PS were 30.2, 25.9, and 9.7%, respectively. The variance explained by the other variables was negligible.

As an index for internal inconsistency we used the absolute difference between SAT and ACH. The hypothesis that there is a positive relationship between leniency and uncertainty was examined by estimating for each subject the following model for the 44 predictions,

$$\mathrm{GPA} = b_0 + b_1 * \mathrm{SAT} + b_2 * \mathrm{ACH} + b_3 * \mathrm{DEV} + b_4 * X_1 + b_5 * X_2,$$

where GPA is the dependent variable (prediction), SAT and ACH are the percentiles in the vignette, and DEV, the index of internal inconsistency, is $|\mathrm{SAT} - \mathrm{ACH}|$. X_1 and X_2 were two dummy variables representing low PS ($X_1 = 1, X_2 = 0$), average PS ($X_1 = 0, X_2 = 0$), and high PS ($X_1 = 0, X_2 = 1$).

Table 1 shows the mean and standard error of the mean (across the 55 subjects) for each of the raw regression coefficients, b_0 to b_5. The hypothesis is supported by results. The value .00194 for b_3 means that a difference of 50 percentiles between SAT and ACH produced an average "leniency increment" of over 0.1 on the 1–4 GPA scale. This is not a huge effect (the average incremental variance explained by DEV is 1.6%), but it amounts to one-fourth to one-third of the size of the prediction increment for an increase of 50 percentiles in SAT or ACH.

In addition to this model we examined, post hoc, the two-way interaction terms. This analysis revealed complex interaetions between DEV with SAT or with PS. The interactions were sueh that in low PS there was a positive linear relationship between DEV and GPA. In average and high PS there was interaction between SAT and DEV, such that the "typical" relationship between GPA and DEV was positive, but its exact value was dependent on the value of SAT. The details of the post hoc analysis are given in the Appendix.

These interactions are consistent with the leniency heuristic but suggest that the influence of uncertainty on prediction cannot be adequately de-

TABLE 1

SUMMARY OF INDIVIDUAL MULTIVARIATE PREDICTION EQUATIONS WITH UNCERTAINTY (DEV) AS AN ADDITIVE FACTOR

Variable	Coefficient	Mean	Standard error
SAT	b_1	.00719	.00058
ACH	b_2	.00593	.00056
DEV	b_3	.00194	.00026
X_1	b_4	−1.17	.14
X_2	b_5	2.00	.22

scribed as a simple main effect. People do not simply increase the value of their predictions in face of uncertainty. It appears that the leniency heuristic interacts with other heuristics that determine "how much" leniency is exercised in each prediction problem. However, in the rest of this paper we will not try to determine the manner by which the leniency heuristic operates in the multivariate prediction task used in this experiment. Rather, the next three experiments will examine how leniency operates in the context of a univariate prediction task. We demonstrate that in this task, as well, the influence of uncertainty on predictions cannot be described by a simple main effect. Rather, the leniency heuristic interacts with weak regressiveness to determine when uncertainty will result in lenient predictions.

EXPERIMENT 2

Validity and Leniency

The prediction task in this experiment is a univariate prediction task. Uncertainty is varied by varying the validity of the predictor. To examine the model presented in Fig. 1 we repeated Kahneman and Tversky's (1973) experiment described in the introduction with two major changes: (1) The use of percentile GPA rather than numerical GPA as an outcome variable. This change had two purposes. First, to rule out the possibility that the asymmetry in predictions stems from the fact that the actual (and subjective) distribution of numerical GPA is negatively skewed. This could possibly account for an asymmetry in regression. Second, to permit the comparison of the relative value of the predictor with the relative value of the prediction. (2) We compared predictions from Mental Concentration to predietions from motivation towards academie success, which was presented as a quite valid, but not perfect, predictor. Kahneman and Tversky compared predictions from Mental Concentration to predictions from percentile GPA, a perfect predictor.[2]

Method

Subjects. Sixty-five undergraduates at Columbia University participated, to fulfill a course requirement, 32 in the Mental Concentration condition and 33 in the Motivation condition.

Procedure. The experimental task was to predict percentile GPA based on percentile score in Motivation Towards Academic Success or percentile score in Mental Concentration. Subjects received a short explanation about percentile and then were given three percentile scores—25, 50, and

[2] Another change was the use of 3 levels of the predictor rather than 11 levels. This leads to higher tendency to make regressive predictions (Slovic *et al.*, 1979).

75—(representing the Motivation or Mental Concentration scores of three Columbia students). Motivation was presented as "... not a perfeet predietor of GPA, but known to be highly correlated with academie achievement." Mental Concentration was presented in the same way as in Kahneman and Tversky (1973), (i.e., "Students with high GPA tend to seore high on the Mental Concentration test and viee versa. However, performanee on the Mental Concentration test was found to depend on the mood and the mental state of the person . . .").

Results and Discussion

The predictions in both eonditions showed no systematie departure from linearity. The mean eorrelation between the predictor and the outcome was .99 in the Motivation condition and .93 in the Mental Concentration condition.[3] The mean predictions from each value of the predictor are graphed in Fig. 2. In addition, two relevant summary statistics are presented by eondition in Table 2: the average mean prediction and the average regression eoefficient (regressing percentile GPA on each predictor).

The relation between uncertainty and leniency is demonstrated in two ways: First, the mean prediction from each predictor is higher than the mean value of the predietor (50); for Mental Concentration, the 95% confidence interval is 57.6 ± 3.8, while for Motivation it is 55.4 ± 3.3. Second, the mean prediction from the lower validity predietor, Mental Concentration, was somewhat higher than that from the higher validity predictor, Motivation. Though the t test for the differenee was not statistically reliable, the effect does turn out reliable in Experiment 4.

A close look at Fig. 2 reveals that the influence of validity on the mean value of predictions is the result of its asymmetric influence on predictions from the low and high values of the predictor. At the low value of the predictor, the 25 percentile, the mean predietions from Mental Concentration and from Motivation exceed 25. The respective 95% confidence intervals are 40.2 ± 5.6 and 32.4 ± 4.6. Also, at this value the predictions from the lower validity predictor, Mental Concentration, are higher than those from Motivation (The 95% confidence interval for the difference is 7.8 ± 7.2). On the other hand, at the high value, the 75 percentile, the predictions from the two predictors are close to each other and close to

[3] Two subjects in the Mental Concentration condition gave the same percentile, 50th, for all three levels of the predictor. Although these responses represent linear predictions, they result in undefined correlation. For these subjects, therefore, the correlation was defined as 1.0. For one subject in the Mental Concentration condition the correlation between the predictor and the prediction was −1.0.

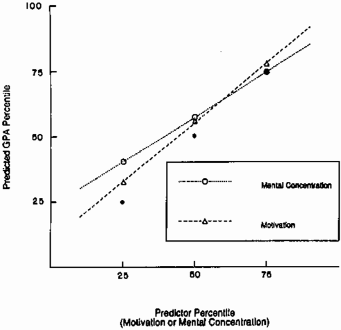

FIG. 2. Predictions of Percentile GPA from predictors that differ in their perceived validity. Open circles/triangles are the group means of predictions from the 25, 50, and 75 percentiles. The standard errors of these means are respectively 2.3, 2.1, and 1.3 for prediction from Motivation and 2.8, 2.3, and 2.5 for prediction from Mental Concentration. Lines are fit by least squares to the three points for each group. Solid circles are predictions by matching.

75. The mean predictions are 74.7 ± 4.8 and 77.8 ± 2.6 from Mental Concentration and Motivation, respectively.

Note that the variability of the predictions is higher at the low value than at the high value of the predictor. This can be seen from the confidence intervals just noted, especially for Motivation, where the halfwidth of the 95% interval is 4.6 for the low value and 2.6 for the high. Since

TABLE 2
AVERAGE PREDICTION EQUATION FOR TWO PREDICTORS

	Motivation		Mental concentration	
	Mean	Standard error	Mean	Standard error
Mean prediction	55.4	1.7	57.6	1.9
Slope	.908	.042	.690	.076

sample size is the same, this directly reflects population standard deviation. A similar pattern of variability was observed in our previous paper (Ganzach & Krantz, in press, Fig. 1). As noted in the introduction, this difference in variability derives from the fact that there is much room for individual differences in leniency at the low value, whereas leniency is limited in its possible extent at the high value.

In comparing the results of this experiment to the results of Kahneman and Tversky's experiment, two things are worthwhile noting. First, Kahneman and Tversky's finding that predictions from Sense of Humor are somewhat regressed against predietions from Mental Concentration is due primarily to differences in predictions from low values of the predictors (Fig. 3 of their paper). Thus the pattern of predietions in their experiment is in accord with the model presented in Fig. 1. Second, Kahneman and Tversky found that predictions from Mental Concentration were not regressed against predictions from percentile GPA, a perfect predictor, while the results of this experiment indicate that predictions from Mental Concentration are regressed against predictions from Motivation, a less than perfect predictor. Perhaps this discrepancy is related to the fact that in the present experiment predictions were made on the same scale as the predictor. This might sensitize subjects to respond with values that are different from the value of the predictor. Under these circumstances both weak regressiveness and leniency might exert more influence on predictions.

The influence of weak regressiveness and leniency on predictions was also revealed in explanations subjects gave to the asymmetry in their predictions, when asked about it after the experiment. Some responders mentioned "giving the benefit of the doubt" to students with low scores on the predictor as the reason for the asymmetry; this is, in fact, a regression principle suggesting regrcssiveness in the low value of the predictor by referring to the principle of leniency. Another typical response was "The student's motivation (mental concentration) can be low, but he can have other important characteristics that will help him succeed in school." Here, "named error" rcasoning for regressive predictions (Ganzach & Kranz, in press) is applied only when it is in line with leniency.

EXPERIMENT 3

Explaining Regression by Leniency

In this experiment the asymmetry of intuitive regression was tested using a different experimental paradigm, one that does not require numerical predictions. Subjects were presented with two descriptions of a particular regression phenomena, one based on a high value of the predictor and one on a low value, and asked to decide which phenomenon is more likely to occur. In addition, subjects were presented with two competing explanations of these regression phenomena, one statistical and one causal, and were asked to pair the two explanations with the two phenomena in a way that seemed appropriate.

Method

Subjects. Fifty-three undergraduates at Tel Aviv University participated in the experiment during a class.

Procedure. The experiment consisted of two questions on a one-page questionnaire:

A. It is a common occurrence that students who received a high grade in their first exam decline in their achievements and receive average grades in subsequent exams. On the other hand, students who received low grades in their first exam increase their achievements and receive average grades in subsequent exams.

Where do you think this phenomenon is more frequent?

1. Among students who received a high grade in the first exam.
2. Among students who received a low grade in the first exam.

B. Two explanations were offered for this phenomenon:

1. The grade in the first exam is partly the result of chance, and does not reflect the true ability of the student.
2. The results of the first exam change students' motivation, and therefore change their achievement in subsequent exams.

Please associate the two explanations in Part B with the most appropriate corresponding cases above; that is, associate one of the explanations with the decline suffered by some students who received a high grade on the first exam and the other explanation with the improvement achieved by some who received a low grade on the first exam.[4]

Results and Discussion

A summary of the answers to the two questions is presented in Table 3. A total of 85% of the subjects felt that regression to the mean is more likely to appear for students at the low value of academic achievement than students at the high value. (The 95% confidence interval is 72–93%.) However, 68% of subjects associated the nonstatistical explanation with the regression phenomenon that they considered more likely (95% confidence interval is 54–80%). A majority of subjects choose the statistical explanation for the phenomenon they believe is less likely to occur and

[4] The items above are translated from the Hebrew; and the final instruction, to associate one explanation with each phenomenon, is a rough paraphrase of the written plus oral instructions actually given.

TABLE 3
JOINT FREQUENCIES FOR JUDGMENTS OF REGRESSION PHENOMENA

		Regression more likely to occur in			
		Low level	High level	Row total	Row percentage
Nonstatistical Explanation better associated with	Low level	33	5	38	72%
	High level	12	3	15	28%
	Column total	45	8	53	
	Column percentage	85%	15%		

the deterministic one for the phenomenon they believe is more likely to occur.

These results confirm the asymmetry of regression effects in intuitive prediction with respect to the low and high values of the predictor, using a quite different method from the preceding experiment. They are also consistent with the assertion that when regression phenomena are recognized they are represented in the form of intuitive prediction rules that are nonstatistical (Kahneman & Tversky, 1973). This experiment suggests that the complementary assertion is true as well. Regression phenomena are more likely to be recognized because they are in line with intuitive nonstatistical rules of prediction. Thus, regression phenomena in the prediction of academic achievement are recognized for students with low scores since a deterministic explanation which is in line with the intuitive principle of leniency is readily available.

EXPERIMENT 4

Leniency and Multiple Determination

In Ganzach and Krantz (in press) we demonstrated that predictions from a single variable become less extreme following experience in which that variable constitutes one component of a multivariate predictor; and we showed (Experiment 3 in that paper) that this result cannot be obtained by merely introducing a second relevant predictor into the "causal field" (Einhorn & Hogarth, 1986). The theory introduced in that paper asserts that such moderation of univariate predictions occurs when, following experience with multivariate-prediction subjects generate and use an additional predictor and assign it a moderate value. Thus, the same result should be attainable even if subjects have no previous experience with multivariate prediction, provided that they generate the additional predictor, perceive it as relevant to the prediction, and assign it a modcrate value. Now we have just shown that subjects assign moderate values to their predictions on the basis of leniency and weak regressiveness. Thus, in the final experiment of this series, we asked some of the subjects to predict two correlated outcomes on the basis of one predictor variable. Predicting the first outcome (the intermediate predictor) was labeled as an "intermediate step" toward predicting the second outcome. The idea is that subjects would generate a moderate value for the first outcome and then, because of the perceived correlation, they would use this moderate value as one component of a multivariate predictor, which would further moderate the prediction of the second outcome.

Method

Overview. Subjects predicted GPA on a 1–4 scale. The design was 2 × 2 factorial, with the two factors being predictor (either Motivation or SAT) and presence or absence of the intermediate prediction task. The intermediate variable was either Intelligence (when the original predictor was Motivation) or Achievement Test score (when the original predictor was SAT). Since the perceived correlation between these two pairs of predictors was found to be different in pretest subjects, the predictor manipulation can be regarded as a correlation manipulation.

Subjects. Two hundred and twenty-four undergraduates in various courses in the Psychology departments of Barnard College and Columbia College participated in the experiment. The experiment was run during class. Subjects were asked to participate but participation was voluntary. Subjects were assigned randomly to conditions.

Procedure. Subjects in the Motivation [SAT] condition with intermediate prediction received the following written instructions (in which information about mean GPA was supplied in order to reduce error variance):

> "The purpose of this prediction experiment is to see how people make predictions when they have only partial information. You receive Motivation [SAT] scores of Columbia students and you are asked to predict their GPAs. (The GPA scale at Columbia is between 1.0 and 4.0 and its mean is 2.86.) The other major indicator of GPA—the student's Intelligence [Achievement Test score]—is not given. However, as an intermediate step, make a prediction of the Intelligence [student's Achievement Test score]. Finally, based on both Motivation and Intelligence [SAT and Achievement Test score] make a prediction of the student's GPA."

For subjects in the control conditions, the final two sentences of this paragraph were omitted.

In a second paragraph, subjects in all conditions were given a brief explanation of percentile. Control subjects then wrote predictions of GPA based on the 15, 50, and 85 percentile of the predictor. The remaining subjects were required initially to write their intermediate prediction of the other predictor (in percentile) and only then to write their predictions of the GPA.

Pretest. Twenty additional subjects answered the following short questionnaire:

> "The following question is about your perception of variables that lead to academic success. Please circle the answer you believe to be most valid:
>
> a. If a student has high Motivation (towards academic success) he is also likely to have high Intelligence.
>
> b. If a student scored high on the SAT he was also likely to score high on the Achievement test."

Results and Discussion

Pretest. Seventeen out of 20 subjects chose SAT to be more highly correlated with Achievement Test than Motivation is with Intelligence.

Intermediate prediction. The means of the intermediate predictions in the two exposure conditions are plotted in Fig. 3. Standard errors are given in the caption.[5]

For SAT, the results of Experiment 2 are confirmed, though the effect is not large: the average prediction from the 15 pereentile is reliably higher than 15 (19.0 ± 2.8), that from the 85 percentile is not reliably below 85 (83.7 ± 2.6). The small effect accords with the fact that SAT score is regarded as an excellent predictor of Achievement test score.

For Motivation, however, the results are dramatic. The asymmetry in moderation, attributed to leniency, persists, but the effects of weak regressiveness are large indeed, so that predictions are moderated even from the high value of the predictor. (The 95% confidence interval is 71.6 ± 3.0.) In fact, 4 subjects felt that Motivation was uncorrelated with Intelligence, as evidenced by the same predictions of Intelligence for all three Motivation percentiles, 4 others indicated a negative correlation, and 11 gave U or inverted U responses. Even when the analysis is restricted to the 34 subjects who showed a monotonic increasing prediction of Intelligence with Motivation, the large regressiveness persists. The graph for this subgroup is similar to that for the full group, but somewhat steeper; it shows highly reliable moderation of predictions even from the 85 percentile. Thus, the extreme subjects can be regarded as lying on a continuum with the other subjects in regard to their theories about the relation between Motivation and Intelligence: people are unwilling to make extreme predictions about Intelligence on the basis of information about Motivation.

[5] One subject in the Motivation group and 6 in the SAT group failed to write down their intermediate predictions.

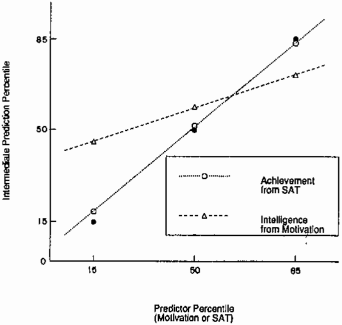

FIG. 3. Intermediate predictions of Achievement from SAT ($N = 56$) or of Intelligence from Motivation ($N = 54$) in Experiment 8. Open circles/triangles are the group means of predictions from the 15, 50, and 85 percentiles. The standard errors of these means are respectively 1.4, 1.5, and 1.3 for prediction of Achievement from SAT and 3.5, 2.2, and 3.0 for prediction of Intelligence from Motivation. Lines are fit by least squares to the three points for each group. Darkened circles are predictions by matching.

Predictions of GPA. Figure 4a shows the group means of predictions of GPA from Motivation and Fig. 4b shows the group means of predictions from SAT.[6] At the 15 percentile of Motivation, the subjects who made intermediate predictions predicted somewhat higher GPA than Control subjects (.28 ± .22, 95% confidence.) The other differences in Fig. 4a, and those in Fig. 4b, are not reliably different from 0, but the latter figure shows the same trend toward elevation of predictions from the 15 percentile of SAT, for the subjects in the Intermediate Prediction group.[7]

The effect of the intermediate predictions on the predictions of GPA was small, especially in view of the large regressiveness shown by the intermediate predictions of Intelligence from Motivation. For predictions

[6] The analyses in this section include all subjects, including those whose intermediate predictions were missing or not strictly increasing with the predictor.

[7] Again, the standard deviation of predictions from the low value was higher than that from the high value. The ratio of these two standard deviations was about 2 for all groups.

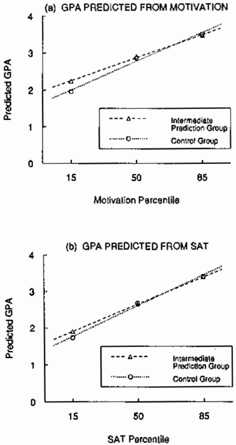

FIG. 4. Predictions of GPA from Motivation and from SAT for two conditions, Control and Intermediate Prediction. Points are group means and lines are fit by least squares. At the 15 percentile the standard error of the difference between condition means is about 0.11 in both graphs; the observed difference is reliably different from 0 in (a) but not in (b). Other standard errors are all smaller but no other group differences are reliably different from 0.

from the 15 percentile, a linear fit shows a reliably positive but small effect of the intermediate prediction on the final prediction: the coefficient is estimated as .0073 ± .0058, or probably less than 0.1 GPA scale units per 10 percentiles of the intermediate prediction. A similar magnitude is found at the 85 percentile. The most likely possibility for these small effects is that the intermediate predictor has limited effect because some subjects make use of it in their final prediction of GPA and others do not.

An alternative measure of moderation in prediction is the individual regression slope relating GPA to Motivation or to SAT. Table 4 gives the

TABLE 4

REGRESSION SLOPES FOR GPA IN EXPERIMENT 8

	Predictor	
Condition	Motivation	SAT
Control	.0220	.0240
	(.0024)	(.0022)
Intermediate Prediction	.0177	.0214
	(.0020)	(.0023)

means and standard errors of these regression slopes for each of the four groups in the present experiment. Since this measure was used by us (Ganzach & Krantz, in press, Table 1) to measure moderation in prediction after experience with multiple determination, it is possible to compare the moderation effect of intermediate predictions to the moderation effect after experience with multiple determination. For prediction from Motivation, the effect of intermediate prediction is $.0044 \pm .0031$, which is somewhat smaller (though not reliably so) than the effect of experience with multiple determination. For prediction from SAT, the effect is $.0026 \pm .0032$, which is smaller than the moderation effect of experience with multiple determination.

In summary, this experiment demonstrates that intuitive predictions can be strikingly regressive, probably when the predictor and the outcome are represented as weakly related. The hypothesis that making an intermediate prediction will moderate final predictions was supported as well. Although the size of this effect is small we believe that it is meaningful. First, it offers a possible explanation for the processes by which weak regressiveness operates: people generate spontaneously values of intermediate predictors which in turn moderate their predictions of the outcome. Second, it is potentially a debiasing technique: forecasters can avoid too extreme predictions by generating values of other relevant but unknown predictors based on the value of the given predictor.

GENERAL DISCUSSION

Two issues are discussed in this section. First, the leniency heuristic in prediction is compared to the positivity bias in social judgment—the tendency of ratings to be too positive (Guilford, 1954; Bruner & Taigiuri, 1954).[8] Second, the relationship between lenient predictions and regres-

[8] The adequacy of comparing the two could be questioned since, normatively, rating is different from prediction. Rating involves evaluation of an input. Prediction involves estimating an outcome. There is more uncertainty in the latter task than in the former, and, therefore, prediction should be less extreme than rating. This normative rule is ignored in

sive predictions is examined in the context of a general model for the representation of inferential reasoning.

An illustrative example of the positivity bias is found in Sears (1983). He showed that between 75 to 91% of the students of UCLA rated most of their professors as better than "average." As this example shows, the positivity bias in social judgment is demonstrated by comparing ratings against some objective criteria. What about comparison between intuitive predictions and an objective criteria? Strictly speaking, the design of the studies reported here does not allow such a comparison. However, when the predictor's value is the 50th percentile, the "objective" value of the prediction is the 50th percentile, since the prediction is based on neutral information. Compared to this objective criteria, subjects' predictions showed leniency, since they were higher than 50. (In Experiment 2 predictions of percentile GPA from mental concentration and from motivation were 57.2 ± 4.5 and 56.1 ± 4.2, respectively; In Experiment 4, predictions of intelligence from motivation and ACH from SAT were 59.3 ± 5.0 and 52.0 ± 3.0, respectively).

While prediction leniency is consistent with the general notion of the positivity bias, there is, to the best of our knowledge, no direct evidence in social judgment research showing that positivity increases with uncertainty. Nevertheless, there are two indications that the relationship between leniency and uncertainty is not limited only to predictions, but is relevant generally to social judgment. First, it is possible that uncertainty is a major factor in determining when a positivity bias will occur and when a negativity bias—the assignment of relatively greater importance to negative features (Kanouse & Hanson, 1971)—will occur. For example, after reviewing the literature on the positivity and on the negativity biases, Markus and Zajonc (1985) conclude: "Positivity bias is not contradictory to negativity bias in that the former refers to making guesses when virtually no information is available. The latter occurs under less uncertainty . . ." (p. 186)

Second, the relation between leniency and uncertainty is consistent with most explanations of the positivity bias, which are motivational in nature [e.g., people feel better when they think that their environment is pleasant (Maltin & Stang, 1978); people are concerned about the consequences of their responses (Decker & Cornelius, 1981)]. Although in the experiments reported here no motivational manipulation was involved, it is quite possible that whatever the motivational reasons that lead to le-

intuitive prediction, rendering them similar to rating. For example, responses to the task of rating students' academic potential based on verbal descriptions of the students do not differ significantly from responses to the task of predicting their achievement based on the same descriptions (Kahneman & Tversky, 1973).

niency, they operate more freely to influence predictions when uncertainty is high. People are motivated to give "benefit" in their judgments about their fellow human beings, and they feel that they are allowed to do that more when doubt is higher.

All the prediction tasks used in this experiment were predictions of human performance. Therefore, one should be cautious in generalizing the results of these experiments to predictions in other domains (e.g., economic forecasts). However, since the positivity bias was shown to exist for ratings of non-human objects (Maltin & Stang, 1978),[9] it is reasonable to ask if the leniency heuristic operates also in prediction tasks other than prediction of human performance. Our discussion suggests that this question is threefold. First, whether people tend to be unjustifiably optimistic in their forecasts; second, whether they project more of this optimism when the uncertainty about the forecasts increases; and third, whether this optimism interacts with weak regressiveness to create an asymmetry in predictions from low and high values of the predictor.

It is possible to ask whether only two heuristics—representativeness and leniency—can account for the asymmetry in predictions from low and high values of the predictor, as well as the differences in variability, without referring to weak regressiveness. One can argue that if motivational reasons are the cause of leniency, they might influence predictions from low values of the predictor more than predictions from high values. For example, if people feel that they are accountable for their judgments, they might be more lenient in predictions from low values of the predictor, since lower evaluations are difficult to defend if one does not have solid reasons.

However, this two-heuristic model does not explain adequately all the results presented in this paper. In particular, it does not explain why the relative values of the predictions from high values of the predictor are not higher than the relative values of the predictors, and why in the high values of the predictors the relative values of the predictions from low validity predictors are not higher than the relative values of predictions from high validity predictors. Furthermore, the results of the intermediate predictions from high values of the predictor in Experiment 4 (especially the predictions of Intelligence from Motivation) are even contrary to the two-heuristic model, since the values of predictions from high values of the predictors are regressed against the values of the predictors, and since in the high values of the predictors, predictions from the low validity predictor are more regressive than predictions from the high validity pre-

[9] This effect appears, however, to be weaker for non-human objects than for human objects (see Sears, 1983).

dictor [the latter finding is also evident in Kahneman & Tversky (1973, Fig. 3)].

The model presented in Fig. 1 can account for this finding. It shows that the relative value of the predictions depends on the comparable influence of leniency and weak regressiveness. If their influence is equal, then the relative value of the predictions will be equal to the relative value of the predictor and changes in validity will not lead to changes in the relative value of the predictions. On the other hand, if weak regressiveness is "stronger" than leniency, predictions from high value of the predictor will be regressive (although less regressive than predictions from the low value).

Weak regressiveness can be viewed as the heuristic that determines the manner by which the benefit of the doubt is given. One can give the benefit of the doubt in the low value of the predictor, but not in the high value of the predictor, since only in the low value is leniency compatible with one's experience with regression phenomena. Thus, involved in intuitive numerical predictions are nonstatistical heuristics (the representativeness heuristic and the leniency heuristic), as well as a statistical heuristic (weak regressiveness). In the following paragraphs we discuss the relevance of this model of intuitive numerical predictions to the representation of inferential reasoning.

In general, there are two extremes by which inferential reasoning can be represented. First, it can be represented by nonstatistical heuristics such as the representativeness heuristic. Second, inferential reasoning can be represented by statistical heuristics, or heuristics that resemble abstract statistical rules (Nisbett *et al.*, 1983). An example for such heuristic, that was extensively investigated, is the intuitive equivalent of the law of large numbers. This heuristic was shown to be used by laymen in everyday problems (Nisbett *et al.*, 1983), and its abstract nature—its resemblance to the statistical law of large numbers—was demonstrated by the influence of statistical training on reasoning about problems requiring the use of this law, whose context is remote from the context of the training (Fong, Krantz, & Nisbett, 1986).

Quite often, however, inferential processes involve both statistical and nonstatistical heuristics. The representativeness heuristic, for example, is only a first approximation for the processes of intuitive numerical predictions. Our experiments and other findings (Ganzach & Krantz, in press; Nisbett *et al.*, 1983; Fischoff *et al.*, 1979) indicate that some recognition of regression effects is involved in the processes of numerical predictions. Thus, the process that governs numerical predictions is predominantly nonstatistical, but still contains some recognition of a statistical principle. The following question given to several classes at Columbia University indicates that a nonstatistical element is involved in the representation of an inferential rule that is predominantly statistical. "A person is standing in the middle of a road after drinking all evening. He is so drunk that his steps are completely random. He is as likely to make a step to the right as he is to make a step to the left." Subjects were asked to estimate when this inebriated person is more likely to be closer to the point of origin, after 100 steps or after 1000 steps. This question was given to subjects who were statistically naive (Introduction to Psychology students, first year students in a Philosophy class) and to subjects who had at least one course in Statistics (graduate students in Psychology, MBA students, and undergraduate Psychology students at the end of their Statistics course). While 70% ($N = 58$) of the statistically naive subjects chose the correct answer (100 steps), only 46% ($N = 57$) of the subjects who were exposed to Statistics chose the correct answer. The reason for the higher proportion of incorrect answers among the subjects who had exposure to Statistics is most likely misapplication of the law of large numbers. Thus, the heuristic that leads subjects who have statistical training to outperform subjects who do not have such experience on many other problems whose solutions require the use of the law of large numbers (Fong *et al.*, 1986, present numerous examples for this) is discrepant from the statistical law of large numbers. This discrepancy leads them to inferior performance in the problem presented here.

Taken together, the studies reported here indicate that the representation of statistical reasoning is best described on a continuum between nonstatistical heuristics and statistical heuristics. Some statistical laws (e.g., the law of large numbers) have rather accurate, although not perfect, intuitive counterparts. Other statistical laws (e.g., the principle of regression to the mean) do not have strong or accurate intuitive counterparts, although some recognition of these laws does influence human inductive reasoning. We suggest that this recognition is learned from experience under conditions in which statistical rules (such as regression to the mean) accord with other intuitive inductive rules (such as the leniency heuristic).

APPENDIX

Since the sample size of prediction trials and of subjects was too small to examine a full model, we used the following approach: we added to the main effect model each of the 10 possible two-way interaction terms, each at a time. Out of these interactions, 2 were significant ($p < .05$): the interaction between DEV and X_1 and the interaction between DEV and SAT. Consequently a model that include the main effects and the two interactions was explored. To facilitate interpretation of the results, we centered the variables at their means. Because of the interaction between

TABLE 5
SUMMARY OF PREDICTION EQUATIONS WITH INTERACTIONS

		$X_1 = 1$		$X_1 = 0$	
Variable	Coefficient	Mean	Standard error	Mean	Standard error
SAT	b_1	.00756	.00064	.00648	.00060
ACH	b_2	.00597	.00075	.00666	.00062
DEV	b_3	.00274	.00047	.00119	.00028
DEV * SAT	b_4	.000008	.000022	.0000479	.000016
X_2	b_5			2.09	.22

X_1 and DEV we estimated the model for vignettes with low PS ($X_1 = 1$) and vignettes with average and high PS ($X_1 = 0$) separately.

For the vignettes with low PS the model that was estimated is

$$\text{GPA} = b_0 + b_1 * \text{SAT} + b_2 * \text{ACH} + b_3 * \text{DEV} + b_4 * \text{SAT} * \text{DEV}.$$

For vignettes with high PS the model that was estimated is

$$\text{GPA} = b_0 + b_1 * \text{SAT} + b_2 * \text{ACH} + b_3 * \text{DEV} + b_4 * \text{SAT} * \text{DEV} + b_5 * X_2.$$

Table 5 shows the means and the standard errors (across the 55 subjects) of the raw regression coefficients of these models. The table shows that for the vignettes with low PS the relationship between DEV and GPA is positive. For the vignettes with average and high PS this relationship depends on the value of SAT. However, since the variables are centered at their means, b_3 represents the "typical" or average relationship between DEV and GPA (Cohen & Cohen, 1983, p. 325), and it is clear from the table that this relationship is positive.

REFERENCES

Agnoli, F., & Krantz, D. H. (1988). *Suppressing natural heuristics by formal instructions: The case of the conjunction fallacy.* Unpublished manuscript.

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, **8**, 1–8.

Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *The handbook of social psychology*. Cambridge, MA: Addison–Wesley.

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

Decker, P. J., & Cornelius, E. T. (1981). The effect on leniency of justifying performance ratings to a supervisor. *Journal of Psychology*, **108**, 211–218.

Einhorn, H. J., & Hogarth, M. H. (1986). Judging probable cause. *Psychological Bulletin*, **99**, 3–19.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, **23**, 339–359.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, 18, 253–292.

Ganzach, Y., & Krantz, D. H. The psychology of moderate prediction: I. Experience with multiple determination. *Organizational Behavior and Human Decision Processes*.

Guion, R. M. (1965). *Personnel testing*. New York, NY: McGraw–Hill.

Guilford, J. P. (1954). *Psychometric methods*. New York, NY: McGraw–Hill.

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

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

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

Kanouse, D. E., & Hanson, L. R., Jr. (1971). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Hillsdale, NJ: Erlbaum.

Markus, H., & Zajonc, R. B. (1985). The cognitive perspective in social psychology. In E. Aronson & G. Lindzey (Eds.), *The handbook of social psychology*. New York, NY: Random House.

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory and thought*. Cambridge, MA: Schenkman.

Nisbett, R. E., Fong, G. T., Lehmann, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, 238, 625–631.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339–363.

Sears, D. O. (1983). The person positivity bias. *Journal of Personality and Social Psychology*, 44, 233–250.

Sears, D. O., & Whitney, R. E. (1973). Political persuasion. In I. de S. Pool, W. Schramm, F. W. Frey, N. Maccoby, & E. B. Parker (Eds.), *Handbook of communication*. Chicago: Rand McNally.

Sharon, A. T., & Bartlett, C. J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology*, 22, 251–263.

Slovic, P. (1966). Cue-consistency and cue-utilization in judgment. *American Journal of Psychology*, 79, 427–434.

Zukier, H., & Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology*, 47, 349–360.

RECEIVED: October 3, 1988