March 1992

Effects of Expectancy on Assessing Covariation in Data: “Prior Belief” versus “Meaning”

Dorritt Billman
Georgia Institute of Technology

Brian H. Bornstein
University of Nebraska-Lincoln, bbornstein2@unl.edu

Jeffrey Richards
University of Pennsylvania

Published in *Organizational Behavior and Human Decision Processes* **53** (1992), pp. 74–88.

http://www.sciencedirect.com/science/journal/07495978

Effects of Expectancy on Assessing Covariation in Data: "Prior Belief" versus "Meaning"

Dorrit Billman
Georgia Institute of Technology

and

Brian Bornstein and Jeffrey Richards
University of Pennsylvania

A large number of researchers have addressed the question of how prior beliefs affect assessment of covariation in new data. Some have suggested that prior beliefs disrupt covariation assessment (Nisbett & Ross, 1980), while others have claimed they help (Wright & Murphy, 1984). Research in this tradition has not consistently distinguished meaningfulness of the data from expectations about the particular relationship between the variables to be assessed. We collected covariance judgments on meaningful variable pairs where subjects had a prior belief in a positive relation, had a prior belief in a negative relation, had a prior belief that the variables are unrelated, or were agnostic about the existence or nature of relation. Subjects rated data with negative, positive, and zero correlations. We evaluated performance in terms of subjects' ability to discriminate objectively different correlations, rather than simply comparing to a reference statistic, and also on the bias subjects showed. Subjects with no prior belief, with positive beliefs, and with negative beliefs were all reasonably well able to discriminate among different objective correlations. In addition, subjects with no prior belief showed appropriate use of the judgment scale, while those having a positive or negative expectation were biased in the direction of their prior belief. In contrast, subjects with the prior belief that the variables were unrelated showed essentially no discrimination. Our results disconfirm the hypothesis that prior beliefs generally facilitate correlation assessment of summarized data. Judgments of meaningful data were best when subjects were initially agnostic.

INTRODUCTION

Assessing contingency is a fundamental aspect of learning from experience and has been widely investigated for its importance in both informal and

The research was supported by NIMH Grant R23H220522 to Dorrit Billman and was conducted during Brian Bornstein's tenure as a NSF predoctoral fellow. We thank Jon Baron and Saul Sternberg for comments on the research. Address correspondence and reprint requests to Dorrit Billman, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332.

scientific reasoning (cf. Alloy & Tabachnik, 1984, and Crocker, 1981, for reviews). Some researchers have stressed how poorly people assess actual contingency and how much they are overwhelmed by prior belief; the identical objective correlation is judged very differently if it conflicts versus is consistent with existing beliefs. Effects of self-fulfilling prophecies (Snyder & Swann, 1978), confirmation bias in evidence selection (Trope & Hassock, 1982; Skov & Sherman, 1986), selective encoding of evidence to match prior beliefs (Cohen, 1981; Lord, Ross, & Lepper, 1979), and overweighing priors in belief revision all focus on the problems stemming from overuse of prior beliefs. Chapman and Chapman (1967, 1969) in their work on illusory correlation stressed how heavily people relied on prior beliefs which made them insensitive to actual patterns in the data. Jennings, Amabile, and Ross (1982) emphasized how poor people were at assessing contingency, whether the data were meaningful (pictured heights of men and walking sticks) or abstract (unlabeled pairs of numbers). Comparing performance across studies, Alloy and Tabachnik (1984) suggested that sensitivity to objective correlations is only good when relevant prior beliefs are absent or when the expected and actual relation are congruent.

The fact that people are affected by background knowledge, however , does not mean that performance is worse when background beliefs are available. Only a few researchers have compared performance on covariation assessment tasks with abstract versus meaningful variables while equating other aspects of the task. Miller (1971) and Muchinsky and Dudycha (1974) looked at how accurately subjects' predictions reflected objective correlation. Wright and Murphy (1984) looked at direct estimation of correlation from summarized data. These studies found that people do worse on the abstract variables, though this is where prior belief should least disrupt performance.

It is possible that simply the meaningfulness of the data, aside from particular beliefs about them, is the primary factor affecting performance. In general, more meaningful material can be held in short-term memory and it is easier to encode and retrieve from long-term memory. Since remembering and comparing data across the data set is needed for contingency assessment (even when all the data is simultaneously present), this might be the primary or sufficient mechanism to explain performance differences with abstract versus meaningful data. If this were the case, the nature of the prior beliefs would not affect the process of assessment. On the other hand, something about the particular content of the prior beliefs does influence the resulting contingency judgment; where prior beliefs are for a positive correlation, judgments are biased positively, and where negative, negatively. While we know background knowledge affects contingency assessment, we know little about the manner in which different sorts of background beliefs affect assessment or the mechanisms for the effects. Given that abstract variables are more difficult to assess than meaningful ones, the current research investigates how different types of prior beliefs about meaningful variables affect the process and outcome of contingency assessment.

The research reported here addresses three goals. First, we assess how the particular type of prior belief affects contingency assessment. Second, we consider what is meant by "good" or "bad" assessment and what measures will be most informative. Finally, we layout some alternatives in how processing may change with different prior beliefs and map these onto expected patterns of performance.

Meaningfulness and Types of Prior Belief

We distinguish between meaningfulness of individual variables and prior belief about how the variables are related. While one cannot have a prior belief unless the variables have an established meaning, there are variable pairs which are individually meaningful but for which one has no opinion on whether or how they are related. These two factors have not been distinguished in previous work; subjects probably did have some prior beliefs about how meaningful variables were related but these were not assessed.

Our experiment used meaningful variables which differed in the type of prior belief. We compared four belief conditions: Positive (Correlation) Belief, Negative (Correlation) Belief, Belief in Zero Correlation, and Don't Know. The first three conditions use pairs of meaningful variables about which subjects have a specific expectation about the nature of the relationship (or its absence); in the latter condition the variables are independently meaningful but subjects were agnostic about whether or how they might be related.

Assessing Effects of Prior Belief

We assume that the primary value of detecting contingency is telling that some things are more related than others. This requires discriminating objectively different correlations and encoding this information to allow comparison and communication. Typically subjects' correlation judgments are correlated with or compared to a reference statistic, be it Pearson's correlation, percentage variance, or some other metric. This approach assumes that the particular scale is psychologically real and should be the standard for estimation and it combines underlying discrimination with how the rating scale is used.

In contrast, we assume only that the direction of the scale and the reference point of no relation are meaningful. Rather than comparing to a reference statistic, we look for rating differences for different objective contingencies and prior beliefs. By looking for changes in the pattern of subjects' ratings we

can still evaluate subjects' estimation. In particular we can assess how well subjects discriminate objectively different correlations and whether or how this is affected by prior belief.

We need to contrast *discrimination*, *sensitivity*, and *bias*. Good discrimination depends on having reasonably good sensitivity and bias. As in signal detection theory (Green & Swets, 1966) *sensitivity* refers to underlying ability to distinguish between objectively different stimuli and *bias* refers to boundaries the subjects sets in using the response categories or points on a rating scale. Prior work, comparing judgments to a reference statistic, focused on how strongly subjects are biased by prior belief. We are particularly interested in whether prior belief affects discrimination. Poor sensitivity always leads to poor discrimination, but poor bias can as well. For example, subjects might have such a strong positive bias that all judgments would be compressed to the "most positive" end of a rating scale and underlying sensitivity could not be expressed.

In sum, several changes follow from focusing on differences across belief conditions rather than comparison to a reference statistic. First, we are primarily interested in discrimination: Is eating vitamin C or getting lots of sleep more related to preventing colds? How does prior belief affect such comparisons? While we are interested in differences in bias, we are most interested in their possible effects on discrimination. Second, we are interested in identifying where poor discrimination might be due to lack of sensitivity and where to bad bias. In turn, this helps pin down where in the evaluation process prior belief has its effect(s): relatively late in the process, such as deciding how to use the rating scale, or earlier, such as making subjects oblivious to data in the first place.

Effects of Prior Belief on Processing

Prior research strongly suggests that prior beliefs will have a biasing effect on assessment of new data. When asked to assess a data set, people seem to have a hard time separating what they believe generally is true in the world from assessing one particular data set, even when the contrast is made very explicit (Kuhn, Amsel, & O'Loughlin, 1988). We consider three, alternative effects prior belief might have beyond a simple bias.

One possibility is that while meaningful data can be assessed more effectively than abstract data, the particular sort of prior belief about the variables does not change the assessment process. Given meaningful data, a subject will not differentially tune out, compare, or evaluate data, from one type of belief to another. Subjects would show equal ability to distinguish between objectively different correlations in all conditions, once subjects were assessing meaningful data. The "Don't Know" condition would differ in the absence of bias, but would be otherwise the same. Indeed, discrimination might be uniformly *good* across different prior beliefs, but prior work, focusing on the match to some particular measure of correlation, might not have identified this.

Another possibility is that availability of any hypothesis helps in the evaluation of data (Wright & Murphy, 1984), above and beyond any effects of meaningfulness of the individual variables. Just as the additional structure from meaningful variables leads to improved performance, the additional structure from a prior belief, be it in positive, zero, or no relation, might further help in the assessment process. So, while a prior belief might bias the final estimate, it might also help the subject encode or compare data. The value of having a hypothesis has been suggested by others for organizing what data to look for (Karmiloff-Smith & Inhelder; 1974). While this differs from covariation assessment tasks, both require integrating evidence across cases. Having a hypothesis might benefit tallying consistent versus inconsistent cases, looking for exceptions, or guiding encoding. If this is correct, the Positive, Negative, and Belief in Zero Conditions would show similar, good, discrimination, while the "Don't Know" condition would show least discrimination, though also least bias.

Finally, the particular prior belief might affect assessment in a very specific way, such that the important contrast might be between believing variables are totally irrelevant versus believing they are or might be related. This contrast may be the most fundamental of all. Related variables usually come from the same domain; variables we strongly believe could not be related usually come from different domains. With variables in different domains, we typically lack theoretical or causal links; hence, forming explanations of correlations across domains may require appeals to causal principles we do not trust or believe in, such as ESP, divine intervention, or conspiracy theories. Belief in no relation seems to entail the greatest change should we be proved wrong and usually we only consider whether variables are negatively or positively related if we have some reason to think them mutually relevant. Hence subjects might be least affected by the data and least able to discriminate differences in the belief in Zero Condition and do roughly the same in the three conditions where subjects believe the variables are or might be related.

THE EXPERIMENT

A pilot study using a mixed within and between-subjects design found that prior belief interacted with data type: when subjects were in the Positive, Negative, or Don't Know Conditions, they discriminated among the data sets, but when subjects believed the variables were unrelated they could not. The present study used a completely within-subject design. In addition to a rating task, it included a forced choice task to provide another tool for separating the effects of bias versus sensitivity. Subjects' sensitivity, not just discrimination, was assessed by having them pick which of two data sets had the stronger correlation. Since no rating scale was used, bias could not affect judgment.

Method

Materials

Variable pairs. We compared performance in four belief conditions: Positive (Correlation) Belief, Negative (Correlation) Belief, Belief in Zero Correlation, and Don't Know. We piloted to find variables where subjects shared the same expectation. Variable pairs were selected so that while the intended relation was generally believed, it was far from logically required. We wanted our subjects to be able to imagine how alternative data (with which they were presented) might be produced. Variables in the Don't Know condition were individually meaningful but subjects did not have a specific correlational belief connecting them. Subjects were instructed to respond "don't know" if they "suspect{ed} that there might be some sort of relationship, but {were} sure neither about the existence of such a relationship nor about its possible direction." For Belief in Zero variable pairs, subjects reported a prior belief that the variables were specifically unrelated and independent.

Piloting identified 12 variable pairs for which there was good agreement about the relationship between the variables: two each in the Positive and Negative Belief conditions, and four each in the Don't Know (which showed the least consensus) and Belief in Zero conditions (see Table 1). More pairs were included for the latter two conditions because we were especially interested in a comparison between these conditions.

Data sets. Three types of data sets—positive, negative, and zero correlation—were constructed for subjects to judge. Each data set consisted of 15 observations for each of two variables, with the observations paired to generate the desired correlation. Each set was generated independently to ensure that results were not an artifact of one particular data sample. For each pair of variables, a data set of each type was generated: Zero: Pearson's $r = 0.0$ (±.07); Positive: $r = +.5$ (±.03); and Negative: $r = -.5$ (±.03). Standard normal distributions for each set were produced with a random number generator, and the values linearly transformed to the mean and standard deviation appropriate for the variables.[1]

[1] Due to our relatively small sample size, any two sets of variable values might not be exactly independent and have the target correlation, The solution was to generate many sample distributions and pick those pairs between which the required correlation and independence held. These distributions were then linearly transformed to meet the sample mean and standard deviation requirements of a given variable, and occasionally slightly modified to repair round-off error.

TABLE 1
PERCENTAGE AGREEMENT WITH THE INTENDED RELATIONSHIP FOR VARIABLE PAIRS USED

Variable Pair	Relationship	% Agreement	Confidence
Body weight Daily caloric intake	Positive	100.0	6.00
No. traffic violations Car insurance premium	Positive	90.9	6.25
Amount of T.V. watched No. of books read	Negative	95.7	5.41
Amount of rainfall No. of sunny days	Negative	82.6	5.11
Height College GPA	No Relation	86.4	6.42
No. movies attended Amount red meat eaten	No Relation	87.0	5.50
White blood cell count IQ	No Relation	87.0	5.50
No. concerts attended Body weight	No Relation	82.6	5.79
% Body fat Red blood cell count	Don't Know	77.3	*b*
Daily calcium intake[a] Avg. resting pulse	Don't Know	72.7	*b*
Silicon dioxide[c] Content rock density	Don't Know	58.3	*b*
Winter rainfall Spring caterpillar length	Don't Know	77.7	*b*

Note. Pairs were selected for highest percentage agreement. Agreement percentages come from one of two pilot studies ($n = 23$ or $n = 22$), except as noted.
[a] Based on results of a separate pilot ($N = 11$).
[b] Confidence ratings were not given for ''Don't Know'' responses.
[c] Percentage based on pretask rating, not piloting; pretask ratings were consistently lower than pilot ratings for pairs where both ratings were available.

In all, 72 sets of observations were independently generated: (12 Variable Pairs) * (2 Variables in the Pair) * (3 Correlation Sets per Variable Pair).

Cover stories. Cover stories for each of the variable pairs described the data as generated by a scientific study which gathered information about both variables. "Subjects" in the invented studies were given names to increase the data's meaningfulness. Subjects who were questioned informally about the data's credibility after the experiment believed the stories to be authentic.

Stimulus booklets. Each story booklet consisted of instructions, a rating sheet for prior belief about each variable pair, and a page with data for

each variable pair to be assessed. A cover story appeared at the top of each page, followed by two data sets labeled Group A and Group B, each with 15 observations on both variables.

For each variable pair, one data set had a zero correlation and the other either a +.5 or a –.5 correlation. Across subjects, eight different orderings of the variable pairs were used. On half their judgments (Rating), subjects rated each of the two data sets separately; on the other half (Forced Choice), subjects made forced choice judgments, picking which of the two data sets showed a stronger correlation. Each subject saw (1) half the pages with –.5/0 and half with +.5/0 pairs of data sets, (2) half with the zero data set on top and half on the bottom, and (3) half forced choice and half rating judgments. Each variable pair was presented in all eight combinations, creating eight Randomized Sets with 12 items in each set. Order of variable pairs was randomized differently in each Randomized Set. The answer book had six pages for rating judgments and six for forced choice responses. There were two Task Orders of each Randomized Set, one with the block of Forced Choice judgments presented first and the other with the Rating judgments presented first.

Design

The balanced design used two experimental factors, Belief (Positive, Negative, Belief in Zero, Don't Know) and Data (Positive, Negative, Zero), and three counterbalancing factors, Randomized Set, Task Order, and Item.

Subjects

Subjects were 48 undergraduate students from a University of Pennsylvania paid subject pool. The majority of subjects in this pool have had no statistics; some have had an introductory course. They received $3 each for their half-hour participation.

Procedure

Subjects participated individually. The experimenter explained the general nature of the task and gave out a story booklet with instructions and an answer booklet. The instructions gave an overview of the task, explained and gave examples of the different types of relationship that might exist between a pair of variables (strong or weak; positive, negative, unrelated, or "don't know") and told subjects they would be judging sets of new data.

First subjects indicated their current belief about the relation between the variables, then they judged the data sets. For each of the 12 variable pairs subjects read the cover story, studied the two data sets, and indicated their judgment in the answer booklet. For half the variable pairs subjects rated each data set individually; for half they decided which had a stronger relation.

Forced Choice and Rating judgments were blocked for each subject and analyzed separately. In the rating judgments, the subject indicated the strength of relationship in the data by marking his or her estimate on a scale of 0 to 100, where 0 meant "not related" and 100 meant "strongest possible relation." Subjects then indicated for each data set whether the relationship was positive or negative (unless they chose 0 on the scale). We used this measure because we thought subjects might find strength of relation easy to judge but might not operate well in distinguishing positive from negative relations; were this so, judgments of size of relation might be a more sensitive index. In the forced choice task, the subject picked the data set (Group A or Group B) that showed the stronger relationship and then indicated whether the relationship in the Group was positive or negative. The two tasks are difficult in different ways: the rating task requires assigning a number, which the forced choice does not, but the forced choice requires considering and comparing two data sets together. Finally, subjects were debriefed, and the fabricated nature of the data and the purpose of the experiment were discussed.

Results

Response Classification

Overall, subjects' initial beliefs about the relation between two variables differed from the intended belief on 31% of the judgments. Subjects were especially reluctant to report lack of an opinion (59% disagreed with our "Don't Know variable pairs). To avoid dropping such a large part of the critical data, we used subjects' reported beliefs to redefine the Belief condition for these judgments.

Main Effects of Belief and Data on Relatedness Ratings

The primary analysis was a repeated-measure ANOVA of rated relatedness with Belief, Data, and Subjects as factors. We combined the magnitude of the relatedness estimate with whether it was negative or positive to get the rated relationship. (We did comparable analyses of absolute value of the rated relationship, but subjects were always less sensitive. Thus, we used the real rather than absolute value.) Randomized Set, Task Order, and Item within belief were assessed separately; only Item was significant [Item $F(41, 506) = 2.43$, $p < .0001$]; including Item in the main ANOVA did not change any effects.

Mean ratings by Belief and Data are shown in Table 2 and Fig. 1. The main effects of Belief [$F(3, 497) = 16.90$; $p < .0001$] and Data [$F(2, 497) = 45.42$; $p < .0001$] were both significant. (Reduced *df* come from reclassification to use subjects' own beliefs.) Subjects judged data more positively with positive prior

TABLE 2

CELL MEANS AND STANDARD DEVIATIONS OF RATING SCORES FOR EXPERIMENT 2

Data correlation	Belief (correlation expectancy)			
	Positive	Negative	Zero	Don't Know
Positive ($r = .5$)				
Mean	26.7	7.5	9.6	16.5
SD	(29.6)	(38.5)	(25.6)	(24.5)
Zero ($r = 0$)				
Mean	18.1	−12.2	−0.4	−5.7
SD	(27.5)	(28.5)	(16.5)	(25.1)
Negative ($r = -.5$)				
Mean	−5.5	−39.0	−4.5	−17.8
SD	(36.0)	(27.2)	(22.5)	(30.3)

beliefs and more negatively with negative prior beliefs, relative to judgments in the Don't Know and Belief in Zero conditions. Post hoc analyses showed that the Positive, Negative, and Belief in Zero belief conditions all differed from each other but that the Don't Know condition differed only from the Positive condition (Tukey's HSD, $p < .05$). Thus, there was a general biasing effect of Prior Belief, such that the specific nature of the belief—whether positive, negative, or zero—shifted subjects' evaluations of the data in the corresponding direction. Subjects' judgments were also appropriately influenced by the relation

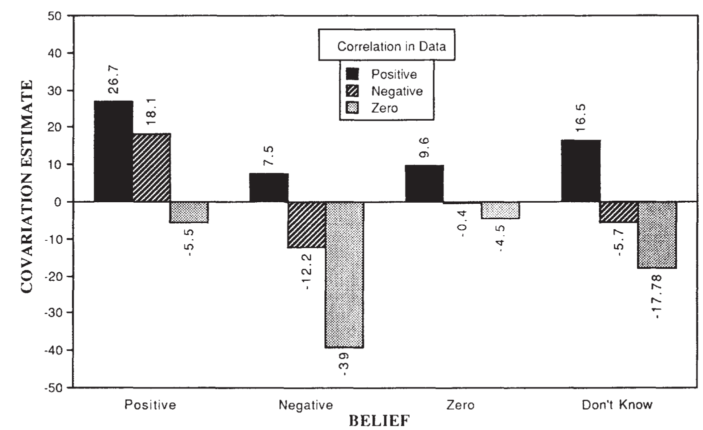

FIG. 1. Judgments of relatedness from Experiment 1. Judged relation for positive (+.5), zero, and negative (−.5) correlations in the presented data sets, grouped by prior belief about the variable pairs presented.

Fig. 1. Judgments of relatedness from Experiment 1. Judged relation for positive (+.5), zero, and negative (–.5) correlations in the presented data sets, grouped by prior belief about the variable pairs presented.

objectively present in the data set. All levels of Data differed significantly from one another (Tukey's HSD $p < .05$).

Interaction Effects of Belief and Data on Relatedness Ratings

The Belief × Data interaction is our primary measure of differential use of the data in different conditions. The interaction was significant , [$F(6, 497) = 3.44$; $p < .003$] and shows clearly in Fig. 1. The Positive, Negative, and Don't Know belief conditions all showed substantial differences in ratings of the positive and negative data, differences of 34 to 47 points. In contrast, ratings in the Belief in Zero condition differed only 14 points for positive and negative data. As in the pilot experiment, these means suggest that discrimination varies across belief conditions, with subjects in the Belief in Zero condition least able to discriminate between different objective correlations.

We explored the nature of this interaction by one-way ANOVAs run separately within each belief condition. The ANOVAs for all four belief-conditions were significant (all F's > 5.0; p's < .005). Thus, subjects in the Belief in Zero condition did show significant discrimination of the data, although the magnitude of the effect is less than in the other conditions (14 vs 34–47 points.). Further, subjects in the Belief in Zero condition showed the least variation in their ratings. Standard deviations across data types ranged from 16.5 to 25.6 for the Belief in Zero condition, as opposed to 24.5 to 36.0 in the other conditions. The existence of an interaction, as well as the difference in variances, suggests that a Belief in Zero correlation exerts a stronger bias than other types (or an absence) or prior belief. It forcefully compresses all ratings toward zero, thereby impairing discrimination between different data sets.

Effects of Belief on Forced Choice Judgments

The forced choice procedure was included to avoid the effects of bias in using a rating scale. Differences among conditions in performance on this measure could not be due to different biases in using the response scale and must be truly due to differential sensitivity. Performance overall was modest, with 60% correct and chance performance equal to 50%. However, performance is best in the Belief in Zero condition with 73% correct, compared to 53% for Positive and Negative belief combined, and 57% for Don't Know. These differences in percentage correct are significant [$\chi^2 = 8.09$; $p < .02$], and only the 73% score is above chance [binomial probability <.0001]. These results suggest that if the task is to compare data sets directly, the Belief in Zero condition is not at all disadvantaged. The fact that subjects in this condition perform at least as well as subjects in the other Belief conditions suggests that a major part of their poor discrimination on the rating task is the result of bias and not a lack of underlying sensitivity.

This finding was confirmed by rescoring the rating data as forced choice to compare performance in the forced choice and rating tasks. Equal ratings, of which there were many in the Belief in Zero condition, were treated as guesses and scored half correct and half as errors. By this scoring, the Positive and Negative Belief conditions together had 65% correct, the Don't Know condition 60%, and the Belief in Zero condition 59%. Thus, when effects of bias were removed by this rescoring or by the true Forced Choice task, the Belief in Zero Condition did not show a disadvantage. Any differential performance appears to be caused by an especially strong bias when making quantitative judgments, and not by diminished sensitivity.

DISCUSSION

Effects of Belief on Performance

Our subjects' primary task was rating strength of relation between two meaningful variables. From a normative perspective, performance was best in the Don't Know condition. Here subjects showed good discrimination among objectively different correlations. They also showed no evidence of bias: the zero point of the rating scale was used for variable pairs with no relation while positive and negative judgments were symmetrical around this zero point. In comparison, judgments with positive or negative prior beliefs showed comparable discrimination but biased use of the scale. Finally, when subjects believed that two variables were unrelated they showed the poorest discrimination and judgments were compressed in toward the zero rating on the scale.

These findings reject the possibility that all meaningful data are created equal and it is just the contrast between meaningful versus abstract data that affects discrimination; different types of prior belief about meaningful data did have differential effects. Further, these findings reject the hypothesis that having any hypothesis, whatever it may be, helps assess contingency in new data. The Don't Know condition, where subjects lacked such a hypothesis, fared best considering both sensitivity and bias, and the Belief in Zero condition, where prior beliefs are probably most firmly held, showed the worst overall discrimination. Finally, the findings are consistent with the idea that working under the belief that variables might but need not be related provides all the benefits but none of the costs of assessing meaningful data.

Clues about Processing: Bias, Discrimination, and Sensitivity

As in prior research, subjects were biased by their prior beliefs. In addition, we showed that discrimination, not just bias was affected; the Belief in Zero condition showed very poor discrimination. We would like to be able to determine whether this poor discrimination is due to poor underlying sensitivity or simply the product of a strong bias which com presses all judgments toward the zero point. While it would be possible to have good discrimination over a very narrow interval of the scale, this would require a corresponding very strong reduction in variability. The forced choice judgments were designed to distinguish sensitivity from bias. Overall performance on this task was rather poor, but here there was no disadvantage (and in fact better performance) for the Belief in Zero condition. Thus our findings are consistent with the view that differences in discrimination across belief conditions stem from differences in bias. This suggests that the prior beliefs have most of their effect relatively late in the assessment process at the point when a rating is assigned to a perceived degree of relatedness. We have no evidence that there are effects on initial encoding or evaluating evidence in our tasks.

From a performance perspective, it does not matter whether poor sensitivity contributes to the poor discrimination; a prior belief in unrelatedness produces a bias so strong that it would ordinarily preclude the expression of underlying sensitivity. From a theoretical perspective, it would be valuable to distinguish between these more sharply.

CONCLUSION

One of the most pervasive findings in contemporary cognitive psychology has been the effect of task content. Different content produces different performance in logically equivalent tasks. In problem solving (Kotovsky, Hayes & Simon, 1985; Klahr & Robinson, 1981), logical reasoning (Cheng & Holyoak, 1985; Griggs & Cox, 1982; Johnson-Laird & Wason, 1977), probabilistic decision making (Tversky & Kahneman, 1980, 1983), and statistical inference (Chapman & Chapman, 1967, 1969; Jennings et al., 1982), formally isomorphic tasks are treated very differently depending on their content. Our experiment traced out similar effects in covariation assessment.

Meaningfulness of individual variables is distinct from beliefs about their relation and all beliefs are not created equal. There appears to be a three-way distinction: (1) believing that meaningful variables are related in a certain way, (2) believing that they are not related, and (3) not holding a belief at all. Being agnostic about the relation produces the best performance while a belief in no relation produces particular difficulties. This might result from difficulty in forming a causal link between the variables; just as causal explanation can lead to belief preservation (Anderson & Sechler, 1986), so might the implausibility of possible causal scenarios produce especially strong belief effects on data assessment.

Our ratings task does have a few close analogs in naturally occurring tasks with sports and business statistics, but it is very different from many informal

learning tasks. Hence, our findings are only suggestive of what the effects of prior belief may be when judgments are based on sequential or nonnumerical observations. Nevertheless, we believe that our findings may reflect some very basic properties of learning about how attributes are related and the use of such knowledge once learned. Specifically, establishing which variables are mutually relevant or belong to the same domain may be a very fundamental part of correlational learning. Some of this partitioning may be innately given; much of it is surely learned. Researchers such as Klayman (1984, 1988) have argued that discovering which features are predictive is the most important component of complex learning. His work shows that discovering what is mutually relevant and predictive goes a long way toward predictive success. Our work shows that success assessing correlations depends, in turn, on prior beliefs about mutual relevance. Indeed, it is striking how deeply the effects of prior belief pervaded our task: presumably there was no ambiguity about what information was relevant, no problem in selecting what information to attend to, and no domain-specific assessment procedures where prior beliefs might have exerted their influence. Nevertheless, prior beliefs about relevance had profound and distinctive effects on performance.

REFERENCES

Alloy, L., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review,* 91(1), 112–149.

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology,* 50(1), 24–34.

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

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology,* 74, 271–280.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology,* 17, 391–416.

Cohen, C. (1981). Person categories and social perception: Testing some boundaries of the processing effects of prior knowledge. *Journal of Personality and Social Psychology,* 40, 441–452.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin,* 90(2), 272–292.

Green, D. M., & Swets, J. A. (1966) *Signal Detection Theory and Psychophysics*. New York: Wiley.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology,* 73, 407–420.

Jennings, D., Amabile, T. M., & Ross, L. (1982). Informal covariation assessment: Data-based vs. theory-based judgements. In A. Tversky, D. Kahneman, & P. Slovic (Eds.), *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge Univ. Press.

Johnson-Laird, P. N., & Wason, P.C. (1977). A theoretical analysis of insight into a reasoning task. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science.* Cambridge: Cambridge Univ. Press.

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory *Cognition*, 3, 195–212.

Klahr, D., & Robinson, M. (\981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology*, 13, 113–148.

Klayman, J. (1984). Learning from feedback in probabilistic environments. *Acta Psychologica*, 56, 81–92.

Klayman, J. (1988). Cue discovery in probabilistic environments: Uncertainty and experimentation. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 14, 317–330.

Kotovsky, K., Hayes, J., & Simon, H. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology,* 17, 248–294.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego: Academic Press.

Lord, C. G., Ross, L., &Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology,* 47, 1231–1243.

Miller, P. M: ,(1971). Do labels mislead? A multiple case study, within the framework of Brunswik's probabilistic functionalism. *Organizational Behavior and Human Performance,* 12, 429–444.

Muchinsky, P. M., & Dudycha, A. L. (1974). The influence of a suppressor variable and labeled stimuli on multiple cue probability learning. *Organizational Behavior and Human Performance,* 12, 429–444.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Skov, R. B., & Sherman, S. I. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies and perceived hypothesis confirmation. *Journal of Experimental Social Psychology,* 22, 93–121.

Snyder, M., & Swann, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology,* 14; 148–162.

Trope, Y., & Bassock, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology,* 43, 22–34.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology.* Hillsdale, NI: Erlbaum.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjuction fallacy in probability judgment. *Psychological Review,* 90, 293–315.

Wright, J. C., & Murphy, G. L. (1984). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General,* 113, 301–324.

Submitted May 1990