

1990

The influence of causality on strategies for making judgments of strength of relation.

Celia M. Klin
University of Massachusetts Amherst

Follow this and additional works at: <https://scholarworks.umass.edu/theses>

Klin, Celia M., "The influence of causality on strategies for making judgments of strength of relation." (1990). *Masters Theses 1911 - February 2014*. 2192.
<https://doi.org/10.7275/7675884>

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.



312066013813091

THE INFLUENCE OF CAUSALITY ON STRATEGIES
FOR MAKING JUDGMENTS OF STRENGTH OF RELATION

A Thesis Presented

by

CELIA M. KLIN

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

September 1990

Department of Psychology

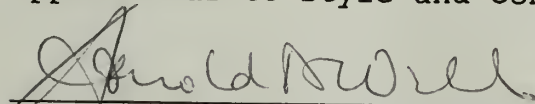
THE INFLUENCE OF CAUSALITY ON STRATEGIES
FOR MAKING JUDGMENTS OF STRENGTH OF RELATION

A Thesis Presented

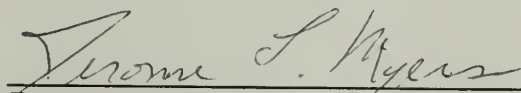
by

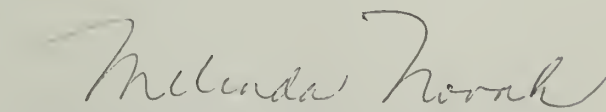
CELIA M. KLIN

Approved as to style and content by:


Arnold D. Well, Chairperson of Committee


Susan T. Fiske, Member


Jerome L. Myers, Member


Melina Novak, Department Head
Psychology

ACKNOWLEDGEMENT

I would like to thank my committee for their support and patience with this project. First I would like to thank Arnie Well for chairing my committee and for introducing me to the covariation literature. I would also like to thank Susan Fiske for bringing a different perspective to this work. Finally I would like to thank Jerry Myers for his seemingly endless supply of time and attention.

I would also like to thank all of the "kids" in the program for being a constant source of support and humor. I would especially like to thank Ruth for the million times she has made me dinner at the end of the day, and Rene' for dragging me out of the lab at those times when I was most resistant to being dragged out. I would like to thank Kevin for providing a great collection of office toys and for being cheerful in the morning when I'm not, Robin for sharing the wisdom she's gained from her years in the program, and John for sitting through a whole year of graduate statistics and only taking a few pages of notes. And of course no acknowledgment is complete without a thank you to that ultimate source of inspiration, the somewhat elusive Wanda Merrill.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	<u>Page</u> iii
LIST OF TABLES.....	v
Chapter	
1 INTRODUCTION.....	1
2 EXPERIMENT.....	18
Design.....	19
Subjects.....	21
Procedure.....	22
Materials.....	22
Results.....	25
3 DISCUSSION.....	35
APPENDIX: TABLES.....	46
BIBLIOGRAPHY.....	53

LIST OF TABLES

	<u>Page</u>
1. The standard lay out of a contingency table for two binary variables.....	46
2. Contingency tables for the six problems, and normative judgments for each problem, based on each of the five strategies.....	47
3. Matrix of Pearson Correlation Coefficients for judgments predicted by each strategy.....	48
4. Mean ratings on the six problems, and delta p_r and ϕ	49
5. Number of subjects using each strategy when cut-off for classification is placed at $r = .3$	50
6. Number of subjects using each strategy when cut-off for classification is placed at $r = .6$	51
7. Average Pearson correlation coefficients of subjects' judgments and normative responses for five strategies.....	52

CHAPTER 1

INTRODUCTION

In attempting to understand how people make judgments about the strength of relation between variables, few researchers have distinguished between events which are causally related and those which are noncausally related. Einhorn and Hogarth (1986) emphasize the theoretical distinction between judgments of covariation and judgments of causality. They consider covariation as just one of many "cues to causality," along with such things as temporal order and contiguity. Hilton (1988) also discusses the limitations of a covariational definition of causality. He notes that often there are conditions which covary with an effect, but are not considered causes of that effect. For example, there may be a high correlation between hair color and eye color, yet one would not claim that this relation is causal. A covariational definition of causality is particularly incomplete in situations in which there is interdependence among causes. In such cases, there may be a chain of causal events leading up to an outcome, and intuition will easily point to one of the events as the cause of the outcome. These judgments are dependent on more than the size of the correlation, or the degree of covariation, between the event and the outcome. According to Hilton, certain types of causal connections, such as abnormal conditions and goal states, are

most likely to be selected from the causal chain as the cause of the outcome.

There is empirical support from a number of domains that suggests that people are sensitive to causal relationships, and that causal information is processed differently than noncausal information. In a study of the perseverance of people's social theories, Anderson, Lepper and Ross (1980) found that subjects were more likely to hold on to their beliefs after being presented with discrediting information if they had first generated causal elaborations for their beliefs. In studies of text comprehension and memory, there is a great deal of empirical evidence which points to the importance of the causal structure of text on readers' understanding and memory of the text. For example, Myers, Shinjo, and Duffy (1987) found that subjects had better recall for sentence pairs if they formed causal elaborations to integrate the two sentences while they were reading.

Einhorn and Hogarth (1986) argue that people are likely to use information from all four cells of a traditional 2 X 2 table (See Table 1) when they make a judgment of covariation. On the other hand, they suggest that people may ignore information from some of the cells when making a judgment of causality. Einhorn and Hogarth also distinguish between two types of causal judgments. People may interpret the task of assessing causality as requiring a judgment about the sufficiency or the necessity of the cause. If suspected

cause X is sufficient to cause effect Y, then cell B should be empty, or in the least be considerably smaller than cell A. If, on the other hand, subjects are thinking about causality in terms of necessity, they are more likely to contrast cell A with cell C. If suspected cause X is necessary for effect Y to occur, then cell C should be empty, or at least be smaller than cell A.

Despite the variety of evidence which suggests that causality is a psychologically real and important construct, it is not clear how strategies for making judgments of strength of relation are affected by whether or not the relation is causal. Most of the covariation research has concentrated on the general question of competence for making a judgment of strength of relation. For example, Nisbett and Ross (1980) made the strong claim that people do not have an understanding of the concept of correlation, while Crocker (1981) argued that Nisbett and Ross were overly pessimistic, and that people are often able to use covariation information appropriately in order to make reasonable judgments of strength of relation.

In order to reach any consensus about people's ability to make judgments of relation, there must be an understanding of the processes that are involved. The development of a process model is complicated by the possibility that there may be many ways of making such judgments. Not only may there be significant individual differences, but people may be flexible

enough to tailor their strategies to deal with the particular characteristics of the judgment task. Empirically, subjects have been shown to be sensitive to a variety of task characteristics, demonstrating different strategies and levels of competence as these task characteristics change. Therefore, as Einhorn and Hogarth have stated, "...the question is not whether people sometimes ignore relevant information but rather, under what conditions is particular information ignored" (Einhorn & Hogarth, 1986, p. 8).

Crocker (1981) has begun to examine the process more systematically by dividing the task into six subtasks including: (1) deciding what data are relevant, (2) sampling cases, (3) classifying instances, (4) recalling the evidence and estimating the frequencies of confirming and disconfirming cases, (5) integrating the evidence, and (6) using the covariation estimate. Although these may all be involved in real-world judgments, laboratory studies vary greatly as to which of these tasks subjects are actually asked to perform. For example, in some studies, subjects are presented with frequency information in summary form (e.g. Shaklee & Mims, 1982; Ward & Jenkins, 1965). In these studies, the task is simplified because the first four steps have already been completed. On the other hand, there is serious question as to subjects' ability to understand the 2 X 2 table in which the summary information is presented. (e.g. Peterson & Beach, 1967). In other studies (e.g. Shaklee & Hall, 1983), data are

presented case by case so that subjects are responsible for steps three and four as well, and an additional source of potential errors is introduced.

When task demands are made more complex and the information processing demands increased, subjects often respond by ignoring much of the relevant data and basing their judgments on a degraded set of stimuli (Crocker, 1983). Shaklee and Mims (1982) found that subjects who had to remember frequency information were less accurate in their judgments than subjects who were presented with frequency information in already tabulated form. Jenkins and Ward (1965) also found that subjects were more likely to base their judgments on all the available information when they were given data in summary form than when the information was presented serially.

Altering the wording of the question that subjects are asked has also been shown to bias the interpretation of the task (Crocker, 1982; Seggie, 1987). Crocker compared three different questions, two of which clearly emphasized a certain type of information: (1) "You want to find out if there is a relationship or connection between working out the day before a (tennis) match and winning that match." (2) "You want to find out if there is a relationship or connection between working out the day before a match and losing that match." In the third condition, the question was unbiased and did not emphasize any one type of instance: "You want to find

out if there is a relationship or connection between whether or not you work out the day before a match and whether you win or lose the match." Subjects were asked to indicate what information they thought they needed in order to make a judgment of the relation or connection between the two variables. Subjects in the unbiased condition requested more information than subjects in either of the biasing conditions. Also, as was expected, subjects in the two biasing conditions most often requested information of the type mentioned in their questions: Subjects in the first condition were more likely to ask how often the match was won after working out the day before the match, while subjects in the second condition more frequently asked how often people lost their match after working out the day before the match. Thus Crocker concluded that the wording of the question can significantly influence subjects' understanding of the task and their strategy for making that judgment.

Other task demands which have been shown to influence subjects' strategies are characteristics of the stimuli themselves. Jennings et al. (1982) compared subjects' judgments of covariation in a data-based condition and a theory-based condition. In the data-based condition subjects were required to make judgments about stimuli for which they had no real-world theories. For example, they examined drawings of men of varying heights carrying walking sticks of varying heights, and were asked about the relation between the

two. In the theory-based condition subjects' judgments were based only on their prior expectations or theories and no data were provided. For example, they were asked to judge the strength of relation between such variables as "students' self-ratings of ambitiousness" and "students' heights," or "students' ratings of Congress's performance in the past decade" and "students' ratings of labor leaders' performance in the past decade."

Jennings et al. found that subjects were much more conservative when their judgments were based only on the data. They were unlikely to venture into the upper half of the 100 point scale even when the objective correlations were as high as .85. They concluded that unless our prior theories lead us to predict strong relations, we are unlikely to detect them even when they occur. Wright and Murphy (1984) also found that it was beneficial for subjects to have a theory about the data, even if their theory was eventually disconfirmed by the data. Subjects with prior theories were less sensitive to outlying values and made judgments which were more congruent with normative measures.

Given that subjects have been shown to be sensitive to some characteristics of the stimuli, it is reasonable to ask how the causal nature of the relation between the variables affects subjects' performance. Most of the empirical and theoretical work in the covariation literature has tended to gloss over the distinction between causal and noncausal

relations even though questions relating to causality have long had a place of importance in the field of psychology, as well as in fields as diverse as philosophy, statistics, psychology, law, and economics (Einhorn & Hogarth, 1986). In fact, the philosopher Hume (1739/1960) believed that our understanding of causation was the "cement of the universe" (White, 1988).

Some theories suggest that our ability to detect causal relations is a very basic, and possibly even innate ability. According to Michotte, an ability to detect cause and effect is necessary in order to function successfully in the world. Michotte believed that even infants' perceptions of physical objects include more than physical features such as color and shape. Our initial perceptions are more abstract and include a knowledge of an object's causal qualities.

White (1988) also suggests that the ability to detect cause is a very basic ability which is either innate or, in the least, present within a few months of birth. He suggests that causal reasoning has its beginning in iconic processing. From a very young age we are able to temporally integrate events which are stored in a single icon, thus perceiving one as causing the other. White, like others, suggests a set of invariant features which define causal relations: temporal contiguity, spatial contiguity, temporal order, and similarity between the antecedent and the consequent. White believes that whereas iconic processing is automatic, the infant soon

learns to generalize these four primitives and then uses them to evaluate more abstract causal relations.

If there are a set of primitives which are unique to causal reasoning, people may treat causal relations in a qualitatively distinctive manner which differs from their treatment of relations which are not causal. As mentioned, Einhorn and Hogarth suggest that judgments which are strictly judgments of covariation are more likely to involve equal attention to all four cells of the contingency table.

Like Einhorn and Hogarth, Green et al. (1979) believe that the attention given to the four types of instances may depend upon the causal nature of the relations. When subjects are asked to make a "choice," they are likely to consider a causal, or one-way relationship, in which one event must precede the other. When instructed to make a "judgment," they may view the relationship as a two-way, or noncausal relationship, in which neither event necessarily precedes the other. For example, if instructed to "choose" which of two medications caused patients to recover most quickly, subjects are likely to view the medication as causing the recovery. On the other hand, if instructed to make a "judgment" about the relation between hair color and eye color, subjects are not likely to perceive one variable as causing the other, because there is no obvious way to order the two variables. Subjects should be as likely to consider the number of instances of blue eyes in the presence of blond hair as they

are to consider the number of instances of blond hair in the presence of blue eyes.

There may be confounds in the covariation literature due to the tendency to ignore this distinction, and to assume the same process for judgments of causality and judgments of covariation. For example, little attention has been paid to the effect of the wording of the question on subjects' understanding of the causal nature of the task, and as Crocker's work suggests (Crocker 1982), subjects' performance may be influenced by the specific wording of the question. In some studies subjects are asked to judge the degree to which an outcome is controlled or influenced by some action performed by the subject (e.g. Alloy & Abramson, 1979; Allan & Jenkins, 1980), while other studies ask subjects to judge the "degree or strength of relationship." It is reasonable to hypothesize that certain questions may lead subjects to assume that the variables are causally related, while other questions may leave the relations unspecified or imply that the relations are noncausal.

Not only may some questions bias subjects to interpret the task causally or noncausally, but as noted earlier, there may be more than one type of causally biasing question. Einhorn and Hogarth (1986) discuss the distinction between forward and backward causal inference. A question such as "You will be asked to judge the degree of influence that medication has on recovery." is a forward causal inference, from event

to outcome. According to Einhorn and Hogarth this asks the question "Will medication lead to recovery?", and suggests to subjects that they are to look for cases in which recovery did not occur in the presence of the medication. This focuses attention on cell B, or on the difference between cell A and cell B.

On the other hand, the question: "You will be asked to judge the degree to which recovery is influenced by the medication." requires a backward causal inference, from the outcome to the event. This question implies the counterfactual question, "Would recovery have occurred if the medication was not taken?" This may encourage subjects to consider cases in which taking the medication did not lead to recovery, and focus attention on cell C, or on the difference between cell A and cell C.

A second factor influencing judgment is the degree of causality suggested by the variables. It has been hypothesized that subjects are more likely to reason causally when presented with variables which have a one-way relation, such as medication and recovery, than with variables which have a two-way relation, such as hair color and eye color. And as Jennings et al. (1982) have shown, subjects' judgments may be influenced by the specific characteristics of the variables.

Although there have not been any thorough and conclusive investigations of the effects of the causal versus noncausal distinction on subjects' strategies, several studies have

investigated certain aspects of the question. Both Adi et al. (1978) and Green et al. (1979) have concluded that subjects display a higher level of thought, in Piagetian terms, when making judgments about noncausal, or two-way relations, than with judgments of causal, or one-way relations.

Adi et al. compared ninth and twelfth-grade high school students' judgments of a causal relation (pill taking and body size of rats) with a noncausal, or "coincidental" relation (tail color and body size of rats). Subjects were asked if a "relation existed" between the two variables, and were subsequently interviewed as to how they had arrived at their conclusions. Subjects' judgment rules were divided into nine categories, which corresponded to four Piagetian stages of development.

A significant difference in response type was found as a function of type of relation, causal or coincidental. In the coincidental task, 34 of 40 subjects were classified in the top two of the four Piagetian stages, while only 17 of the 40 subjects in the causal condition were classified at these levels.

Unfortunately, this study contained several serious methodological flaws. Distinctions between the nine categories were somewhat unclear. Furthermore, classifications were based entirely on subjects' self-reports rather than on their actual judgments. There is good reason to question subjects' ability to analyze and communicate their

own reasoning, especially for a task as complex as a judgment of covariation. Shaklee and Hall (1983) found limited congruence between subjects' self-reports of their covariation judgment rules and their actual judgments.

Green et al. (1979) also compared causal, or one-way, and noncausal, or two-way, relationships. They found significant differences between the two conditions, although again, the methodology was questionable. Like the Adi et al. study, responses were classified into Piagetian levels of thought. The classifications were inconsistent. For example, the comparison of the two row conditional probabilities, $(A/(A+B) - C/(C+D))$, is included in both level 3 and level 4. In addition, a comparison of conditional probabilities is rated as an inferior strategy than a comparison of diagonals, or of confirming versus disconfirming instances: $((A+D) \text{ versus } (B+C))$. In reality, in situations in which the marginal frequencies are unequal, a comparison of conditional probabilities provides a more accurate judgment than a comparison of confirming and disconfirming cases. For example, assume that the likelihood of recovering when given medication A is equal to the likelihood of recovering when given medication B, but medication A is prescribed twice as often as medication B, so that the marginal frequencies are unequal. In this case, a comparison of conditional probabilities will yield a correlation of zero, accurately indicating the lack of relationship between the type of medication taken and

recovery. On the other hand, a comparison of confirming and disconfirming cases will suggest that the two medications are not equally related to recovery, as the unequal number of instances of medication A and medication B will be taken into consideration.

Green et al. found that subjects who showed a high level of thought in the causal task (transitional formal thought) advanced to a higher level of thought (fully formal) with the noncausal task. On the other hand, subjects who showed a lower level of thought (concrete operations), regressed in the noncausal task. Green et al. concluded that judgments of noncausal relations require a higher level of thought than judgments of causal relations. This is consistent with Einhorn and Hogarth's prediction that judgments of noncausal relations encourage the use of more of the available information. According to Green et al., this can result in either an increase or a decrease in an individual's performance, depending on the sophistication of their reasoning.

Most studies in the covariation literature base their conclusions solely on subjects' final judgments of the strength of relation between the variables. This is usually a number on a scale from 0 to 100, or -100 to +100 if negative correlations are included. In order to develop a process model of how judgments of covariation are made, an understanding is needed of the effects of task demands on

subjects' actual judgment-making strategies, rather than only on their final judgments. A design developed by Shaklee and Tucker (1980) and Shaklee and Hall (1983), the "rule diagnostic approach," differentiates between four possible data combining strategies. Each of the strategies represents a different combination of the four cells of the contingency table. The following four judgment rules were included in their work:

1. Cell A information only.
2. Cell A versus Cell C.
(Shaklee et al refer to this strategy as A versus B, but in a traditional contingency table it would be labeled A versus C.)
3. Comparison of the sum of the diagonals, or delta d.
($(A+D)$ versus $(B+C)$)
4. Comparison of the conditional probabilities, or delta p.
($(A/A+B)$ versus $(C/C+D)$)

The cell A strategy is considered the simplest, and the least adequate, because it ignores three out the four types of information. The comparison of the conditional probabilities is considered the most sophisticated because it includes all four types of instances and combines them in such a way as to be sensitive to marginal frequencies. In the Shaklee et al. studies, the problems are structured so that the cell A problems can be solved correctly by all four strategies, A versus C problems by strategies 2, 3, and 4, sum of diagonals problems by strategies 3 and 4, and conditional probability problems by strategy 4 alone.

The Shaklee et al. task required subjects to view a set of observations which described each of two events. For example, a plant was pictured as healthy or sick, as a function of amount of water, large or small. After receiving 24 observations, subjects made a judgment about the relation between the two events on a three point scale. The three points represented a negative correlation between the events, a correlation of zero, or a positive correlation. Each subject was then classified as using one of the four strategies depending on the number of problems they judged correctly.

This design was successful in detecting systematic differences in subjects' strategies. For example, Shaklee and Tucker (1980) found that while 17% of high school students used the cell A strategy, the simplest of the four, only 1% of the college students did so. Furthermore, in one study, Shaklee and Hall found that only 14 of 80 subjects failed to match any of the four judgment patterns they included, and subjects appeared to be fairly consistent in their use of a particular strategy.

Clearly the four judgment rules which were included are but a subset of all of the possible strategies that subjects could be using. Although the use of each of the four strategies provides a unique pattern of results, alternate strategies could also account for the same pattern of results. In particular, a comparison of cell A and cell B would have

produced judgments which were indistinguishable from the A versus C strategy, and in some of their experiments, the results would have been indistinguishable from the sum of the diagonals strategy and the comparison of conditional probabilities as well.

If Einhorn and Hogarth's (1986) distinction between analyses of causality based on sufficiency and necessity is valid, it would be particularly beneficial to construct problems to distinguish between the cell A versus B and the cell A versus C strategies. The use of the cell A versus B strategy represents viewing causality in terms of sufficiency whereas the use of the cell A versus C strategy represents viewing causality in terms of necessity.

In summary, most of the previous work has ignored the distinction between causal and noncausal relations. The work which has focused on this question has relied solely on subjects' self-reports. The present research was designed to examine the effects of the causal nature of the variables and of the instructions on subjects' process for making a judgment of covariation.

CHAPTER 2

EXPERIMENT

The primary manipulation in the present study involved a comparison of subjects' strategies in a noncausal and in several causal conditions. In each causal condition, subjects were presented with information about variables with a one-way relation. They were read a cover story about patients who either took or did not take a medication, who subsequently recovered either slowly or quickly from a disease. In the noncausal condition, subjects were given information about a two-way relation. They heard a cover story about patients who had or did not have a particular physical symptom and their speed of their recovery from a disease.

The second manipulation involved a test of Einhorn and Hogarth's (1986) hypothesis about forward biasing and backward biasing questions. Three different questions were used in conjunction with the causal variables: The first of these was an unbiased question. The second question focused on how the event, taking or not taking the medication, influenced the speed of recovery from the disease. The third question was biasing in the backward direction and focused on how the speed of recovery was influenced by taking or not taking the medication.

In sum, there were four between-subject conditions: The first group of subjects received the noncausal cover story

and the unbiased question. The second, third, and fourth groups of subjects received the causal cover story. The second group of subjects were asked the forward biasing question, the third group of subjects were asked the backward biasing question, and the fourth group of subjects were asked the nonbiasing question.

Design

In the following study, there were several modifications of the rule diagnostic approach used by Shaklee et al. (1980, 1983). Problems were included to distinguish between the (A-B) and the (A-C) strategy, although no problems were included to identify the cell A strategy because such a low percentage of college age subjects have been shown to use this strategy (Shaklee & Tucker, 1980). In addition, problems with negative correlations were not used, as it has been demonstrated that people have particular difficulty understanding the concept of negative correlations and in making accurate judgments. (Malmi, 1986; Seggie, 1975; Jurd, 1975)

Delta p and phi were included as two normative measures of strength of relation. Delta p, the difference between the two row or column conditional probabilities, is considered to be the appropriate statistical measure for judgments of strength of relation between two binary variables when there is a one-way, or causal relationship. For this study, delta

p represents the difference between the row probabilities: Δp_r . For these particular problems, Δp does not differ much for the columns and rows. Φ , the correlation coefficient, is included as well as it is generally considered to be the normative measure when there is a two-way, or noncausal relationship between row and column variables (Allan, 1980).

As in the rule diagnostic approach, problems were structured so that the different strategies produced distinctly different patterns of results. Whereas subjects in the Shaklee et al. studies made categorical judgments on a three-point scale, in this study subjects made judgments of strength of relation on a 100 point scale.

Subjects were given six problems, each with 24 observations. The objective strengths of relation, as measured by Δp_r , ranged from .56 to -.2.¹ Table 2 presents the 2 X 2 contingency tables for each of the six problems. For both the causal and the noncausal cover stories, the column variable (B) was rate of recovery. The two levels of this variable were recovering quickly and recovering slowly. For the causal cover story, the two levels of the row variable (A) were taking a medication and not

¹ As discussed, the decision was made not to include problems with negative correlations. The intended value of Δp_r and Φ for problem 6 was zero, but in designing the stimuli, the value was unintentionally set to -.2.

taking a medication. In the cover story for the noncausal group, the two levels of the row variable were having a symptom and not having a symptom.

The problems were structured to differentiate between (A-B), (A-C), and the three strategies which make use of all four cells of the contingency table. The problems were not designed to distinguish between the sum of the diagonals strategy (Δd), Δp_r , and ϕ . These three strategies involve attention to all four cells of the table and represent a high level of sophistication of reasoning. In order to differentiate between them, a substantial increase in the number of problems would be required.

Table 3 presents the Pearson correlation coefficients for the normative judgments based on each of the five strategies. Other than Δd , Δp_r , and ϕ , which are highly correlated, the correlations of the other strategies are all negative and close to zero.

The six problems were presented in one of two orders. The observations within each problem were randomly ordered. Each problem was presented in one of two orders, the second of which was the reverse of the first.

Subjects

Eighty-eight University of Massachusetts students participated for course credit or for five dollars. The data from five subjects were dropped because their responses

included a judgment of zero for four or more of the six problems. Subjects were run in groups of five or fewer; some subjects were run individually. Groups of subjects were randomly assigned to instructional groups, with the restriction that there be an equal number of subjects in each instructional group, and in each of the four orderings within an instructional group. An experimental session lasted approximately 25 minutes.

Procedure

Each subject was read the causal or the noncausal cover story as well as the general instructions. They were told to write down their judgment of strength of relation, a number from zero to 100, at the end of each of the six problems. Subjects were given the opportunity to ask questions about their task before beginning.

The 24 observations for each problem were played on a portable cassette recorder. At the end of each problem, the tape recorder was turned off and the question subjects were to respond to was repeated. When everybody had written down their answer, the next problem was started.

Materials

Causal task.

The causal cover story was introduced as follows:

"A new strain of the flu has just been identified. Not much

is known about it yet, but some people who come down with it recover in about 24 hours while others take much longer.

As little is yet known about the virus, different doctors are prescribing different medications to their patients. A group of medical researchers is now attempting to determine if there is a relationship between the type of medication a patient takes and the rate at which they recover.

You will be asked to make judgments about six different medications. For each medication you will listen to a series of statements. Each statement will provide information about one patient. You will be told if they took that type of medication or did not take that type of medication, and if they recovered slowly or quickly."

This cover story was then followed by one of three questions:

- 1.) The non biasing question: "You will be asked to judge how strongly the medication and the recovery rate are related."
- 2.) The forward inference question: "You will be asked to judge the degree of influence that the medication has on the rate of recovery."
- 3.) The backward inference question: "You will be asked to judge the degree to which the rate of recovery is influenced by the medication."

Subjects were then read the following instructions regarding the judgment scale:

"Your judgment will range from 0-100. A judgment of 0 means that there is no relationship between the type of medication a person takes and their rate of recovery while a

judgment of 100 means that there is a perfect relationship between the type of medication a person takes and their rate of recovery."

Subjects then listened to 24 observations for each of the six medications:

Cell A information: "One person took the medication and recovered quickly."

Cell B information: "One person took the medication and did not recover quickly."

Cell C information: "One person did not take the medication and recovered quickly."

Cell D information: "One person did not take the medication and did not recover quickly."

Noncausal task.

The noncausal cover story was introduced as follows:

"A new strain of the flu has just been identified. Not much is known about it yet, but some people who come down with it recover in about 24 hours while others take much longer.

Different people with this virus have different symptoms, including coughing, sore throat, fever, headache, congestion, and stomach pains. A group of medical researchers is now attempting to determine if there is a relationship between the symptom that the patients have and the rate at which they recover.

You will be asked to make judgments about six different symptoms. For each judgment, you will listen to a series of statements. Each statement will provide information about one patient. You will be told if they have that symptom or do not have that symptom, and if they recovered slowly or

quickly. You will be asked to judge how strongly the symptom and the recovery rate are related.

Your judgment will range from 0-100. A judgment of 0 means that there is no relationship between the type of symptom a person has and their rate of recovery, while a judgment of 100 means that there is a perfect relationship between the type of symptom a person has and their rate of recovery."

Subjects then listened to 24 observations for each of the six symptoms:

Cell A information: "One person had the symptom and recovered quickly."

Cell B information: "One person had the symptom and did not recover quickly."

Cell C information: "One person did not have the symptom and recovered quickly."

Cell D information: "One person did not take have the symptom and did not recover quickly."

Results

Table 4 presents a summary of subjects' judgments of the strength of the relation between the row and column variables for each of the six problems. The first four columns of Table 4 contain the average judgments for each of the four instructional groups. The fifth column contains the overall average judgment for each problem, and the sixth and seventh columns contain the objective measures, delta p , and phi respectively, of the strength of the relation in each of the six problems.

As can be seen in Table 4, there was a considerable

amount of variability in subjects' ratings, with standard deviations ranging from about 18 to 34. This seems to be comparable to the variability reported in other studies (e.g. Arkes & Harkness, 1983; Wright & Murphy, 1984), although in much of the literature there is no report of variability, so the comparison is limited. Despite this variability, on the average, subjects were sensitive to the variations across problems. Overall, mean ratings varied from 37.84 for problem two, in which $\Delta p_r = \phi = 0$, to 61.83 for problem six, in which $\Delta p_r = .56$ and $\phi = .49$. In an analysis of variance on subjects' judgments, with instructional group, order of problems, and sequence of trials within a problem as the between-subjects variables, there was a highly significant effect of problems, $F(5,335) = 11.02, p < .001$.

With the exception of problem six, the rank order of subjects' judgments followed that of the rank order of Δp_r and ϕ . If problem six is not included (a discussion of problem six is to follow), the Pearson correlation coefficient for the mean subject ratings for problems one through five, with the normative responses according to Δp_r and ϕ , were $r = .991$ and $r = .993$ respectively.

Although problem six had a lower Δp_r and ϕ than problem 5, the average judgment in problem six was larger than the average judgment in problem five. Δp_r and ϕ for problem six were $-.2$ and the average judgment was 47.72, while

delta p_r and phi for problem five were zero and the average judgment was 37.84. These ratings differed significantly from each other: $F(1,67) = 11.36$, $p = .001$. Problem six had a higher rating than problem five in all four instructional groups as well as in the overall mean.

One explanation for this reversal was that when the materials were prepared, delta p_r and phi for problem 6 were set to $-.2$ instead of zero, as was intended. This may have created some confusion for the subjects since they were instructed to express their judgments for each problem as a number between zero and 100. If subjects were aware that the direction of the correlation was negative, they may have been confused as to how to translate their judgment into a positive rating. Not only did the judgment scale include only positive numbers, but, as was mentioned, it has been demonstrated that people have particular difficulty understanding the concept of negative correlations and in making accurate judgments (Jurd, 1975; Malmi, 1986; Seggie, 1975).

Another possibility is that subjects were aware that the **absolute** relation between the variables was greater than zero. They may have then based their judgment on the magnitude of this relation and not the direction. This would explain why the mean judgment for problem six was higher than for problem five. In fact, if delta p_r and phi for problem six had actually been $+.2$ instead of $-.2$, the correlation of subjects' ratings for all six of the problems and delta p_r or

phi, would have been $r = .987$ and $r = .992$, respectively.

Alternatively, it is possible that subjects were unaware of the negative correlation, and the reversal for problems five and six was based on the representation of the 24 observations in each of the problems. Although problems five and six did not differ greatly in Δp_r and phi, the distribution of frequencies across the four cells of the contingency table was very different for the two problems. In problem five there were an equal number of occurrences in each of the four cells of the contingency table. Of the 24 patients, 12 were given the medication or had the symptom, and 12 patients recovered quickly. This is probably the most intuitive way of representing a correlation of zero. In problem six on the other hand, 17 out of the 24 patients were given the medication or had the symptom, and 16 out of the 24 patients recovered quickly.

Looking at the strategies which are included in Table 2, for problem five the judgments are the same regardless of which of the strategies is used. In problem six the situation is very different. Unless subjects used the information from all four cells of the contingency table and combined this information in a way which took the unequal marginal frequencies into account, their judgments would have been considerably larger than zero, and considerably larger than their judgment in problem five.

Looking at the average judgments separately for each of

the four instructional groups (see Table 4), one finds a somewhat different pattern than for the average over groups. In the ANOVA of subjects' ratings there was not a significant main effect of instructional group, although there was a significant interaction of problems and instructional group, $F(15, 335) = 2.112, p = .009$. This interaction reflects a different pattern of judgments over problems, depending on the set of instructions and the cover story that subjects received. Also significant in this ANOVA were two uninterpretable third order interactions: Problems X Instructional group X Order, $F(15, 335) = 1.940, p = .019$, and Problems X Order X Sequence, $F(5, 335) = 3.70, p = .003$.

One way of attempting to understand the interaction between problems and instructional group is by examining subjects' ratings in terms of the five strategies that have been discussed. Subjects' strategy use was analyzed in two ways: The judgments of each subject were correlated with the judgments predicted by each of the five strategies. In the first analysis an attempt was made to classify each subject according to one strategy. In the second analysis, an average correlation across subjects was determined for each strategy.

A somewhat arbitrary rule was used in order to decide which strategy each subject was using. Table 5 lists the classifications when subjects were classified as using a strategy if the correlation of their responses and the responses according to that strategy was $r > .3$. Strategies

delta d, delta p, and phi are not evaluated separately because they are so highly correlated with each other. With $r = .3$ as the cut-off, the number of subjects using each strategy is about the same for the three causal groups. On the other hand, the subjects in the noncausal group were less likely to be classified as using any of the strategies. In each of the four instructional groups, causal as well as noncausal, more subjects were classified as using one of the strategies which involves the use of all four cells of the contingency table than either the (A - B) or the (A - C) strategy.

When the cut-off is set higher, at $r = .6$, the pattern changes. Table 6 presents this classification. By raising the criterion, the number of subjects in the noncausal group who cannot be classified as using any of the strategies is now about the same as that of the causal forward and the causal unbiased group. In this analysis, the subjects in the causal backward group show a different pattern of strategy use than subjects in the other three groups. Consistent with the data in Table 5, subjects were classified as using delta d, delta p, or phi more often than either (A - B) or (A - C).

Although there is a simplicity in interpreting the data when subjects are classified categorically, this type of analysis does involve an arbitrarily chosen categorization rule. And as has been demonstrated, the choice of where to place this cut-off can strongly affect the final interpretation. Unfortunately, with this set of data, this

analysis does not allow for any conclusive theoretical claims about the effects of the causal nature of the variables on subjects' strategies for making judgments of strength of relation.

In the second analysis, the correlations of each subject's judgments and the responses predicted by each of the five strategies were averaged across subjects. Table 7 presents a summary of the average Pearson correlation coefficients. In the first four columns the average correlations are presented separately for each of the four instructional groups. In the fifth column the overall averages are presented for each problem.

As can be seen by looking at the fifth column of the table, there is no evidence that subjects are using either the (A - B) or the (A - C) strategy. Both of these correlations are very close to zero: for (A - B), $r = .068$, and for (A - C), $r = -.009$. The correlations with delta d, delta p, and phi are much higher, again suggesting that subjects may be using all of the information provided, or in the least, they are not systematically dismissing information.

In an ANOVA on these data, there was a main effect of strategies, $F(4, 268) = 7.357$, $p < .001$. The data set included the same between-subject variables as in the original ANOVA: instructional group, the order of the problems, and the sequence of trials within each problem, but the within-subjects variable was now strategy instead of problem. In

this analysis, the dependent variable was the correlation coefficient for subjects' judgments and strategies, instead of the actual judgments on the problems.

This main effect of strategies suggests that classifying subjects' ratings into strategies may provide a useful framework for attempting to understand subjects' behavior, and for examining the nature of the interaction between subjects' responses and instructional group.

Looking at the first four columns of Table 7, the correlations are somewhat different for each of the four instructional groups than for the overall average. As was found in Table 5, subjects in the noncausal group appear less likely to have used any of the strategies. Looking at the column averages in Table 7, the average correlation for the noncausal group is $r = .051$, which is lower than any of the causal groups.

This difference in the average correlation for each instructional group is reflected in a marginally significant between-subjects main effect of instructional group:

$F(3, 67) = 2.72, p = .051$. The within-subjects interaction of instructional group and strategies was not significant, $p = .201$. It therefore remains unclear how the causal nature of the variables influenced subjects' strategy use.

Although the interaction between instructional group and strategies is not significant, there is a trend which is worth noting. With the exception of the causal backward group and

the (A - B) strategy, none of the groups show any evidence of using the (A - B) or the (A - C) strategy. But the groups behave somewhat differently from each other with regard to the strategies Δd , Δp_r , and ϕ . Although these correlations are quite low in the noncausal group, they are somewhat higher for the three causal groups. This trend does not support the prediction (Einhorn & Hogarth, 1986; Green et al., 1979; Seggie et al., 1975) that subjects use more of the information when they are making judgments about variables which are not causally related.

In order to assess the magnitude of the influence that problem six had on the results, additional analyses were performed on problems one through five only. In an ANOVA on the judgments, the pattern of results did not differ from the ANOVA which included all six problems. There was a significant main effect of problems as well as a significant interaction between problems and instructional group. In an ANOVA on the correlations of subjects' responses and the judgments predicted by the five strategies, there was also no change in the pattern of results when problem six was excluded. There was a significant main effect of strategies and a significant between-subjects effect of instructional group. But, as in the original ANOVA, the interaction between strategies and instructional group was not significant.

In sum, in the ANOVA of the judgments on the six problems, as well as on problems one through five only, there

was a significant effect of problems, suggesting that subjects were sensitive to the variations across problems. There was also a significant interaction between problems and instructional groups. While this suggests that subjects were influenced by the set of instructions and the cover story that they received, further analyses did not provide much insight into the nature of this interaction. In the analysis of subjects' use of the strategies, there was not a significant interaction between strategies and instructional group. Due to the variability within each subject's responses as well as between subjects, it was difficult to assign subjects to strategies and to identify differences between instructional groups. There was somewhat of a trend for judgments in the noncausal group to correlate less strongly with any of the five strategies than the judgments in the causal groups. There was also a trend across all four instructional groups for subjects to use one of the strategies which involved all four cells of the contingency table rather than the (A - B) or the (A - C) strategy.

CHAPTER 3

DISCUSSION

The motivation for this study was to determine if subjects' strategies for making judgments of strength of relation are different for variables which are causally and noncausally related. Previous research has suggested that when subjects are presented with variables which are causally related they tend to ignore information from some of the cells of the contingency table, while subjects making the same judgment of strength of relation with variables that are not causally related use more of the available information. Furthermore, research in other areas of psychology, as well as fields outside of psychology, have provided empirical and theoretical evidence that suggest that causality is a psychologically real construct. The purpose of this study, therefore, was to investigate the possibility that subjects differ in their process for making a judgment of strength of relation when the variables are causally and noncausally related.

Unfortunately, based on the data and the analyses presented, it is difficult to draw definitive conclusions. Although there was a significant interaction between problems and instructional group, in subsequent analyses no clear pattern of results emerged to explain this interaction. Most notably, in examining subjects' use of the strategies there

was no clear pattern of strategy use within any one instructional group nor were there clear differences between instructional groups. Several of the trends which did emerge were not consistent across the different analyses.

One interesting finding was that the majority of subjects seemed to have used information from all four cells of the contingency table. As can be seen in Tables 5 and 6, more subjects were classified as using delta d, delta p_r , and phi than either (A - B) or (A - C). As well, as can be seen in Table 7, the average correlations are highest for these three strategies which involve all four cells of the table. In light of the debate over people's competence for making judgments of covariation, these data suggest that people do not systematically dismiss entire sources of information when making a judgment. These data also do not support the claim (Einhorn & Hogarth, 1986; Green et al., 1979; Seggie et al., 1975) that subjects use less of the available information when the variables are causally related. Looking at Table 7, the trend is actually in the opposite direction; the correlations with delta d, delta p_r , and phi are higher in the causal groups than in the noncausal group.

Some of the difficulties in drawing conclusions about these data are due to methodological problems which are specific to this study, while others are due to problems which are common to much of the covariation literature. One of the factors which makes it difficult to determine which, if any,

of the strategies subjects were using is the variability in subjects' responses. One of the reasons for this variability may have been the small number of trials in each problem. In the probability learning literature, subjects were faced with the simpler task of predicting which of two events was going to occur based on one or more probabilistic cues. Studies used up to 500 trials and judgments did not reach an asymptote for at least 50 trials (Castellan, 1977). If it is assumed that subjects start with a base judgment, for example of 0 or 50, and then increment or decrement this initial judgment after every new observation, after only 24 observations their judgments should still have been fluctuating significantly. If this is the case, the final judgment may have been strongly influenced by irrelevant dimensions, such as the order of the trials.

In addition, with only six problems per experimental session, subjects may not have settled into any one strategy until several problems into the experiment. When the average correlations between subjects' judgments and the normative responses according to the strategies were subjected to an ANOVA, the pattern of results was somewhat different when all six problems were included than when only the 4th, 5th, and 6th problems were analyzed. In the full ANOVA there was a main effect of instructional group: $F(3,67) = 2.72, p = .051$. When only the final three problems were included, this effect disappeared. ($p = .695$) This suggests that the pattern of

responses across problems may have not have been consistent. With a greater number of problems, subjects' behavior may have been more stable.

Another source of variability may have been the oral presentation of the information. This is in contrast with most of the other studies of covariation in which the information is presented visually. In this study, the subjects' task was fairly passive and involved little more than sitting and listening to the same four sentences over and over. After completing the experiment, several subjects reported that they had had a difficult time maintaining their attention throughout the experiment.

Beyond these specific criticisms of the task, there are limitations in using the framework of strategies as the basis for investigating the process of making a covariation judgment. One problem with this approach is that it is difficult to determine empirically which strategies subjects are using. Any attempt to develop a set of problems for which the use of the different strategies will produce a pattern of responses which are distinct from each other places heavy constraints on the problems. In this study, it was not possible to develop problems with very high objective correlations which produced different predictions for each of the strategies. And many other problems acted much like problem 5, in which all five strategies generated the same judgment.

Another difficulty arises in classifying subjects' responses into strategies, as was discussed in the previous section. Crocker (1981) discusses the difficulties of analyzing judgment data and suggests that this is one of the serious methodological problems in the field of covariation. Although there are several methods of interpreting judgment data, each of them has its limitations.

One method that is commonly used (e.g. Allan & Jenkins, 1980, 1983), is to correlate subjects' judgments with the normative responses according to each of the strategies. This allows for some speculation as to the extent that subjects are using the information in the four cells of the 2 X 2 table and how they are combining this information, particularly if the results are clear cut and the correlations are very high or very low. But since the correlation coefficient is sensitive to such aspects of the data as the spread of the scores, the interpretation is often not clear. For example, if subjects were actually using one of the strategies but were only using a limited range of the 100 point scale, their responses might not correlate very highly with that strategy.

Another difficulty arises if these correlations are then averaged across subjects for each of the strategies, as was done in Table 7. If it is assumed that subjects are using only one of the strategies, then the correlation of their responses with the strategies they are not using is at best difficult to interpret, and at worst, meaningless. If, for

example, a subject is not using the (A - B) strategy, it is not necessarily meaningful to draw any theoretical conclusions based on the size of the correlation of their judgments with the normative responses according to the (A -B) strategy. The interpretation of a negative correlation between a subject's judgment and the normative responses of one of the strategies is particularly unclear.

It may therefore be more enlightening to disregard the correlations of subjects' responses with all but the one strategy that they are using, or that strategy with which their judgments correlate most highly. This involves classifying subjects categorically as using one (or none) of the strategies, as was done in Tables 5 and 6. Although this analysis is somewhat simpler to interpret it involves using a somewhat arbitrary categorization rule. Table 5 presents classifications of subjects based on the rule that the correlation of their responses and the judgments predicted by that strategy was .3 or greater; with this rule subjects in the noncausal group were less likely to use any of the a priori specified strategies than subjects in the three causal groups. On the other hand, it could reasonably be argued that a correlation of .3 is too low of a cut-off and that a .3 correlation between subjects' judgment data and the responses predicted by one of the strategies does not necessarily mean that the subjects were actually using that strategy. When the cut-off was set higher, at $r > .6$ (see Table 6), very

different theoretical conclusions could have been drawn from these data.

In addition to deciding on the appropriate statistical analysis, there are a number of difficult design choices to be made. The most obvious example of this involves the choice of the subset of strategies to include. In making this decision, not only does one have to consider all of the theoretically driven combinations of cells (A vs. B, A vs. C, etc.), but also the many ways the same cells can be combined. For example, in including a strategy that reflects Einhorn and Hogarth's hypothesis that subjects making a forward causal inference will focus on cells A and B, it is not clear if the strategy should be $(A - B)$ or $(A / (A + B))$.

Clearly it is not possible to develop a set of problems which makes it possible to distinguish between all of the possible strategies. Since subjects can only be said to be using one of the strategies which was included in the study, the design of the study puts serious constraints on the theoretical conclusions which can be drawn.

Another limitation of using strategies as a framework for covariation research is the assumption that subjects use the information from each of the cells of the 2 X 2 table in an all or none fashion. A more plausible assumption is that subjects differentially weight the four kinds of information. For example, in Einhorn and Hogarth's forward causal inference, even if subjects do concentrate on cells A and B,

it is unlikely that they completely ignore the information from cells C and D. A more plausible strategy than either $(A + B)$ or $(A/A + B)$ might be something like:

$(A/(A + B)) - k (C/(C + D))$, where $0 < k < 1$.

Up to this point, the discussion of the limitations involved in using the concept of strategies as a way of characterizing subjects' behavior has included concrete design problems: Because most subjects' responses do not correlate perfectly, or even strongly, with one strategy it is not clear how to best analyze the data and characterize any individual subject's process. Equally unclear is which strategies should be included. Both of those difficulties point to the fact that strategies are normative models rather than process models. Even if the methodological problems could be addressed, all of the plausible strategies specified, a set of problems developed to distinguish between them, and each subject classified definitively as using one and only one strategy, the result would be a description of performance rather than a process model. It would still not be clear as to how subjects had formed their judgments or how changing various parameters would affect this process.

As discussed earlier, Crocker (1981) breaks down the process of making a covariation judgment into six separate subcomponents: (1) deciding what data are relevant, (2) sampling cases, (3) classifying instances, (4) recalling the evidence and estimating the frequencies of confirming and

disconfirming cases, (5) integrating the evidence, and (6) using the covariation estimate. Even if subjects had been successfully classified as using one particular strategy, little could be said about their process in each of these stages. If, for example, there had been clear evidence to support Einhorn and Hogarth's position, and subjects in the forward causal inference group had tended to generate judgments that correlated highly with the (A - B) strategy, there would be several very different possible explanations for this effect. Utilizing Crocker's stages, this result can be explained in terms of one of several of the stages.

It could be claimed that the causal nature of the variables affects stage 2, sampling cases. When the variables are causally related, subjects may find the information from cells A and B more relevant than the information from cells C and D and therefore be more likely to encode this information. Another equally plausible interpretation would be that this effect was due to differences in stage 4, recalling the evidence and estimating the frequencies of confirming and disconfirming cases. While subjects were initially collecting the information, they may not have been biased in their memory for the four types of instances, but at the point in the task when a judgment was required, they may have had better recall for the information in Cells A and B. A third possible explanation for this effect involves stage 5, integrating the evidence. Subjects could have

accurately sampled cases, accurately recalled and classified this information, but their decision rules could have given greater weight to cells A and B than to cells C or D.

In most of the covariation literature, inferential leaps are made in an attempt to explain how a variety of parameters have affected subjects' process of making a judgment. Most studies only examine subjects' final judgments. However, the question of the influence of a parameter such as the causal nature of the variables on the process of making a judgment is actually several distinct questions. For example, assuming that Crocker's stages are valid, the causal nature of the relationship between variables may have a very different effect on the stage of classifying instances than it does on integrating evidence.

Although no one experiment can provide information about each of the stages involved in making a judgment of covariation, a more systematic exploration, or a bottom-up approach, would allow more concrete conclusions and less of a need to draw largely speculative inferences from the data. There has not been sufficient work which directly tests subjects' behavior with the simpler stages of the process. In attempting to answer the question about the effects of the causal nature of the variables, what is needed is a thorough investigation of its effects on each stage of the process. For example, in studying the effects of this parameter on the stage of the process when subjects sample instances, subjects

could be asked only to estimate the frequencies of the four types of instances rather than to make a judgment of strength of relation. The effects of a variety of parameters could then be studied in relation to this stage of the process: the number of observations, the order of the trials, the wording of the question, and so on. With this type of understanding of each stage of the process, a model could then be developed which could begin to successfully predict subjects' responses under a number of conditions, and adequately explain what subjects were doing at each of the stages.

Although the top-down approach which has primarily been used in the field has been valuable in identifying the parameters of interest and generating relevant questions, a more systematic exploration of the process is needed before these questions can be satisfactorily answered.

APPENDIX

TABLES

TABLE 1

The standard lay out of a contingency table for two binary variables.

		OUTCOME			
		present		absent	
POTENTIAL CAUSE	present	-	A	-	B
	absent	-	C	-	D

TABLE 2

Contingency tables for the six problems, and normative judgments for each problem, based on each of the five strategies.

Problem			A-B	A-C	Delta d	Delta p_r	Phi
1.							
	B1	B2					
A1	5	1	4	0	12	.56	.49
A2	5	13					
2.							
	B1	B2					
A1	3	2	1	-1	12	.39	.35
A2	4	15					
3.							
	B1	B2					
A1	9	8	1	7	4	.24	.22
A2	2	5					
4.							
	B1	B2					
A1	8	2	6	-1	2	.16	.17
A2	9	5					
5.							
	B1	B2					
A1	6	6	0	0	0	0	0
A2	6	6					
6.							
	B1	B2					
A1	11	6	5	5	0	-.2	-.21
A2	6	1					

TABLE 3

Matrix of Pearson Correlation Coefficients for judgments predicted by each strategy (See Table 2).

Strategy	Strategy				
	A-B	A-C	Delta d	Delta p_r	Phi
A-B	1.000				
A-C	-0.078	1.000			
Delta d	-0.129	-0.372	1.000		
Delta p_r	-0.127	-0.382	0.908	1.000	
Phi	-0.126	-0.403	0.885	0.997	1.000

TABLE 4

Mean ratings on the six problems¹, and delta p_r and phi.

Problem	Instructional Group				Mean	Delta	Phi
	Non-causal	Causal Forward	Causal Backward	Causal Unbiased			
1	52.27 (27.36)	65.19 (28.32)	73.55 (21.01)	57.10 (25.07)	61.83 (26.49)	.63	.49
2	63.32 (29.14)	50.62 (29.60)	43.60 (34.01)	51.70 (28.08)	52.55 (30.56)	.39	.35
3	50.09 (24.44)	51.29 (25.60)	43.50 (23.29)	48.68 (20.61)	48.68 (23.64)	.24	.22
4	45.86 (28.14)	48.62 (28.50)	53.60 (23.41)	36.60 (22.31)	46.19 (26.09)	.16	.17
5	45.27 (23.64)	43.14 (21.01)	29.20 (18.17)	32.75 (22.65)	37.84 (22.20)	0	0
6	50.05 (25.17)	52.57 (22.96)	51.00 (25.27)	36.80 (22.34)	47.72 (24.36)	-.21	-.20
Mean	51.15	51.91	49.10	43.94	49.14	.19	.17

¹ Standard deviations are given in parentheses.

TABLE 5

Number of subjects using each strategy when cut-off for classification is placed at $r = .3$

Instructional Group	Strategy			
	(A-B)	(A-C)	Delta d Delta p_r Phi	None
Noncausal	5	2	10	5
Causal Forward	7	5	9	0
Causal Backward	7	2	11	0
Causal Unbiased	4	2	13	1

TABLE 6

Number of subjects using each strategy when cut-off for classification is placed at $r = .6$

Instructional Group	Strategy			
	(A-B)	(A-C)	Delta d Delta p_r Phi	None
Noncausal	1	2	8	11
Causal Forward	2	2	5	12
Causal Backward	5	2	10	2
Causal Unbiased	0	0	9	11

TABLE 7

Average Pearson correlation coefficients of subjects' judgments and normative responses for five strategies.

Strategy	Instructional groups				Mean - all subjects
	Non Causal	Causal Forward	Causal Backward	Causal Unbiased	
(A-B)	-.090	.091	.344	-.057	.068
(A-C)	-.008	.012	-.036	-.006	-.009
Delta d	.159	.201	.295	.411	.263
Delta p _r	.099	.197	.303	.403	.246
Phi	.093	.188	.290	.390	.236
Mean	.051	.138	.239	.228	.161

BIBLIOGRAPHY

- Adi, H., Karplus, R., Lawson, S., & Pulos, R. (1978). Intellectual development beyond elementary school VI: Correlational reasoning. School Science and Mathematics, 78, 675-683.
- Allan, L. C. (1980). A note on measurement of contingency between two binary variables in judgment tasks. Bulletin of the Psychonomic Society, 415, 147-149.
- Allan, L. C. & Jenkins, H. M. (1980). The judgment of contingency and the nature of the response alternatives. Canadian Journal of Psychology, 34, 1-11.
- Allan, L. C. & Jenkins, H. M. (1983). The effect of representation of binary variables on judgment of influence. Learning & Motivation, 14, 381-405.
- Alloy, L. B. & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed college students: Sadder but wiser? Journal of Experimental Psychology: General, 108, 441-485.
- Anderson, C. A., Lepper, M. R., Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. Journal of Personality and Social Psychology, 39, 1037-1049.
- Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. Journal of Experimental Psychology: General, 112, 117-135.
- Castellan, N. J. Decision making with multiple probabilistic cues. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), Cognitive Theory, vol.2, New Jersey: Lawrence Erlbaum Associates, 1977.
- Crocker, J. (1981). Judgment of covariation by social perceivers. Psychological Bulletin, 90, 272-292.

- Crocker, J. (1982). Biased questions in judgment of covariation studies. Personality & Social Psychology Bulletin, 8, 214-220.
- Einhorn, H. J. & Hogarth, R. M. (1986). Judging probable cause. Psychological Bulletin, 99, 3-19.
- Green, S., Jurd, M., & Seggie, J. L. (1979). Formal thinking about correlation. Scandinavian Journal of Psychology, 20, 119-125.
- Hilton, D. J., Logic and causal attribution. In D. J. Hilton (Ed.), Contemporary science and natural explanation. Commonsense conceptions of causality. Harvester Press, 1988.
- Jenkins, H. M. & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. Psychological Monographs, 79, (1, Whole No. 594).
- Jennings, D. L., Amabile, T. M., & Ross, L. Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. London/New York: Cambridge Univ. Press, 1982.
- Jurd, M. F. (1975). Some aspects of the understanding of social science concepts in adolescents. Report to the National Committee for Social Science Teaching.
- Malmi, R. A. (1986) Intuitive covariation estimation. Memory & Cognition, 14, 501-508.
- Michotte, A. The perception of causality. New York: Basic Books, 1963.
- Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. Journal of Memory and Language, 26, 453-465.
- Nisbett, R. E., & Ross, L. Human inferences: Strategies and shortcomings of social judgment. Englewood Cliffs, N.J.: Prentice-Hall, 1980.

- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. Psychological Bulletin, 68, 29-46.
- Seggie, J. L. (1975). The empirical observation of the Piagetian concept of correlation. Canadian Journal of Psychology, 29, 32-42.
- Seggie, J. L. (1987). The judgment of covariation between binary variables: Some conditions that influence the process. Memory & Cognition. 15 (4), 341-348.
- Shaklee, H. & Hall, L. (1983). Methods of assessing strategies for judging covariation between events. Journal of Educational Psychology, 75, 583-594.
- Shaklee, H. & Mims, M. (1982). Sources of error in judging event covariations: Effect of memory demands. Journal of Experimental Psychology: Learning, Memory & Cognition, 8, 208-224.
- Shaklee, H. & Tucker, D. (1980). A rule analysis of judgments of covariation between events. Memory & Cognition, 8, 459-467.
- Ward, W. C. & Jenkins, H. M. (1965). The display of information and the judgment of contingency. The Canadian Journal of Psychology, 19, 231-241.
- White, P. A. (1988). Causal processing: Origins and Development. Psychological Bulletin, 104, 36-52.
- 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, 301-322.

