1989

Subject sensitivity to changes in task and stimulus characteristics :: effects on judgement behavior.

R. Kevin Stone
University of Massachusetts Amherst

SUBJECT SENSITIVITY TO CHANGES IN TASK AND STIMULUS CHARACTERISTICS: EFFECTS ON JUDGMENT BEHAVIOR

A Thesis Presented

by

R. KEVIN STONE

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

MASTER OF SCIENCE

February, 1989

Department of Psychology

SUBJECT SENSITIVITY TO CHANGES
IN TASK AND STIMULUS CHARACTERISTICS:
EFFECTS ON JUDGMENT BEHAVIOR

A Thesis Presented

by

R. KEVIN STONE

Approved as to style and content by:

Arnold D. Well

Arnold D. Well, Chairperson of Committee

James I. Chumbley

James I. Chumbley, Member

Marvin Daehler

Marvin Daehler, Member

S. M. Berger

Seymour Berger, Department Head
Department of Psychology

ABSTRACT

SUBJECT SENSITIVITY TO CHANGES
IN TASK AND STIMULUS CHARACTERISTICS:
EFFECTS ON JUDGMENT BEHAVIOR

FEBRUARY, 1989

R. KEVIN STONE, B.A., KEENE STATE COLLEGE

M.S., UNIVERSITY OF MASSACHUSETTS

Directed by: Professor Arnold D. Well

The present study was designed to investigate two questions concerning how subjects judge the strength of relation between two continuous variables. The first question was whether subjects' judgments would be affected by changes in slope, variance of X, and/or variance of estimate of Y on X (error variance) in a way similar to how these changes affect the Pearson product-moment correlation. The second question concerned whether the sensitivity exhibited to the variables in question would change as a result of the manipulation of the instruction sets. Three instruction sets were used: a neutral set where subjects were simply asked to assess the strength of relation (Group J), a set which discussed predictability as a means of

understanding strength of relation (Group P), and a set which discussed the concept of error variability as a means of understanding strength of relation (Group F). It was found that subjects judgments were influenced by error in Group F more than in Groups P and J, with correspondingly less influence exerted by changes in the level of variance of X (and to a lesser degree, slope). Future avenues of research are also discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In order to understand the world, we must detect whether or not variables we encounter are related, assess the nature and strength of relations, and make predictions about the state of one variable given information about the state of another. How people are able to perceive and evaluate the degree to which two variables are related is a topic which has received a great deal of attention in the past few decades. Researchers working in the areas of probability learning, social psychology, clinical psychology, and even animal learning have contributed to our understanding of how covariation is detected and evaluated.

To date, most of the research has investigated the extent to which people are able to detect and evaluate covariation, and discussion has centered on what constitutes accuracy and what conditions elicit more or less accurate judgments. Perhaps a better focus might

have been on what kinds of information people use in making judgments of strength of relation, and how they use this information differently depending upon the exact nature of their task.

One concern about research dealing with covariation detection and estimation is that we do not have a clear idea about what subjects are sensitive to when making judgments of strength of relation. Another concern is that we do not know what subjects are doing when they generate judgments in such a task, or even what it is they understand their task to be when instructed to make such judgments. The present research was designed to address these two issues.

The first issue addressed by the present research is what information subjects use when making judgments about imperfect linear relations. Are they directly sensitive to relations between variables in the same way as the Pearson correlation coefficient, or are they more sensitive to other descriptors of linear relations, such as the slope of or variability around the regression line. This issue was investigated by how changes in slope, variance of X, and error variance were reflected in subjects' judgments of strength of relation, compared to how changes in these variables affect the Pearson correlation coefficient.

The second issue addressed by this research was whether the manner in which subjects use information about slope, variance of X, and error variance differs depending upon the specific tasks they are instructed to perform. What was of interest here was 1) whether subjects' judgment behavior was different depending upon the specific task (judgment vs. prediction + judgment), and 2) whether, by means of instructions, subjects could be focused on a particular component, namely error variability (i.e. variability around the least-squares regression line), as a means of understanding strength of relation.

There is a large body of literature which examines how (and how well) people make judgments of strength of relation. Most of the research concerned directly with covariation detection and evaluation has used dichotomous variables. Few researchers have focused on how people perceive and use information concerning continuous variables, though much of the information that we interact with is not simply comprised of pairs of mutually exclusive events, but rather with continuous gradations of information, such as degrees and measures of some type. Although we are primarily concerned here with relations between continuous variables, we shall briefly refer to some of the findings of studies that have used dichotomous variables as they impact on the present research. Specifically, we shall look at some of

the factors that have been found to affect subjects' judgments of strength of relation in both of these sets of studies.

There are two main types of tasks that have been used in the study of covariation detection to try to understand whether, and how well, people use information concerning continuous variables: judgment and prediction. Each approach has its drawbacks as well as its strengths. Judgment tasks ask the subject a seemingly very straightforward question: based on the given data, how strong is the relation? However, the results have varied widely both within and between studies. Factors such as the manner in which the stimuli are presented (as well as the stimuli themselves), the instructions used, the way in which the judgment of strength of relation is elicited, how the stimuli are defined by the cover stories used, and whether subjects are given any context within which to view the stimuli have all potentially contributed to these varied results. These factors all fall into one of two categories: task characteristics and stimulus characteristics. We shall therefore examine how some of these factors have been shown to affect subjects' judgments in studies in the literature.

In covariation studies that have employed binary variables, much of the discussion has been about what combinations of cells subjects use to make their

judgments, as well as how that information is combined (eg. Beyth-Marom, 1982). There has been little discussion, however, concerning exactly what information subjects use in a covariation task using continuous variables. Some authors have used the correlation coefficient (*r*) as a normative criterion, implicitly assuming that subjects are directly sensitive to r, or perhaps sensitive to the components that are combined to make up r (eg. Well, Boyce, Morris, Shinjo, & Chumbley, 1988; Jennings, Amabile, & Ross, 1982). This practice of using *r* as a normative criterion has been recently questioned by a number of experimenters, since it is possible that subjects are sensitive to other characteristics of a linear relation. For example, Jennings et al (1982) reported that the group average ratings of relation in a continuous variable covariation judgment task tended to be characterized by the function $100(1 - \sqrt{1 - r^2})$, not *r*.

Also, Wright and Murphy (1984) have suggested that an alternative measure of correlation may be more appropriate in evaluating subject performance than the standard Pearson *r*. Specifically, they suggest that a more 'robust' measure of correlation, (that is, one that is affected less by single outlying data points than is Pearson's *r*) may be more in keeping with subject response behavior. They showed that subject responses were

affected less by such outliers than was the correlation coefficient. This clearly suggests that subjects may be sensitive to certain direct measures of strength of relation, and in a way different than the correlation coefficient is. To quote Well et al (1988):

"It is possible that it is more adaptive to be sensitive to separate components of a relation such as rate of change and predictability than it is to be directly sensitive to a composite measure such as the correlation coefficient."(p. 18).

So what aspects of the relation might subjects be sensitive to? One possibility is that they could be sensitive to the degree of slope of the relation: how much does one variable (Y) change on the average as the other (X) changes. Or, if expected to predict one variable from the other, they could be sensitive to the amount (either in an absolute or relative sense) by which those predictions were incorrect.

Lane, Anderson, and Kellam (1985) focused on the question of just what it is about covarying variables that subjects are sensitive to when making judgments of strength of relationship. Their argument is that "in order for judgments of covariation to be a monotonic function of Pearson's correlation, these judgments must be the same for all data having the same value of Pearson's correlation, regardless of the values of the

individual components"(p. 641). It must be stressed here that while Lane et al. (1985) refer to these variables as 'components' of the correlation, they each describe aspects of the linear relation (at least, slope and error variance do so), and as such are useful indicators of the degree to which two things are related. What is of interest here is that these variables can be combined in a certain way so as to produce the correlation coefficient. The Pearson product-moment correlation combines these types of information (slope ($\underline{b}$), variance of X (S_x^2), and error variance (S_{yx}^2)) in the following way.

$$\underline{r}^2 = \underline{b}^2 S_x^2 / (\underline{b}^2 S_x^2 + S_{yx}^2)$$

If subjects are directly sensitive to the Pearson correlation coefficient $\underline{r}$, there should be no difference in subject judgments of strength of relationship if these components are varied, so long as $\underline{r}$ remains the same. By systematically varying these components, Lane et al. (1985) showed that, in a task where subjects were to judge the strength of relation exhibited in a scatterplot or in a table, subjects were more sensitive to changes in error variance than to slope or the variance of X, (relative to their effects as indicated in expressions for $\underline{r}$). While this was statistically significant only when stimulus information was presented in the form of a scatterplot, there was also a suggestion that the effect

was present for tabular presentation. It appeared that as error variance was made smaller (with concurrent changes in slope and/or the variance of X, in order to maintain the same correlation), judgments of strength of relationship were higher. This finding led Lane et al. (1985) to conclude that the Pearson product-moment correlation was not necessarily the best criterion by which to evaluate subject performance in a covariation task.

It remains unclear just what subjects are responding to in a judgment task. The Lane et al. (1985) study asked the proper question when they looked at what people may be sensitive to in judging strength of relation, in terms of variables that are related to, but not identical to the commonly used Pearson correlation coefficient. It may be, however, that they did not ask the question properly. To return to the theme that task as well as stimulus characteristics affect subjects' judgment behavior, it is important to examine the effects of instruction and task on subject performance when stimuli are manipulated as in the Lane et al. (1985) study. It is our contention that this is an area where the Lane et al. (1985) study may have suffered, and which may have been responsible for the non-significant results in the tabular condition. They instructed subjects to assess the relation of the X and Y variables by using a number line (0-100), where "0

means no relation, and 100 means a perfect linear relation." It is unclear what subjects understand the phrase 'perfect linear relation' to mean. It is likely that subjects could interpret these instructions in the graphical format condition, because they could be readily translated into 'how well the stimulus approximates a line'. In the tabular format, where the results were not significant, these instructions may have been confusing, or simply unhelpful. The statistically fluent subjects who participated in their third experiment would have been more likely to understand what a linear relation was, and in fact, there seemed to be a greater, although still statistically non-significant, effect of error variance in the statistically fluent tabular format group.

While the effect of instructions has not been given much discussion in the realm of continuous variable covariation studies, there has been a great deal of discussion of this topic in covariation studies using dichotomous variables. These studies present stimuli which can be fit into the cells of a 2x2 table (such as the presence or absence of a symptom and the presence or absence of a disease) in the following way:

		disease	
		present	absent
symptom	present	cell a	cell b
	absent	cell c	cell d

The stimulus information is presented either serially, one pair of stimuli at a time, as in Smedslund (1963), and Jenkins and Ward (1965), etc., or in a summary table form, where all of the information is presented in a 2x2 table (such as in Ward and Jenkins, 1965). Subject judgment performance is then evaluated in terms of some normative model, such as the phi coefficient (which for dichotomous variables is equivalent to the correlation coefficient), or delta P (which is the difference between the two conditional probabilities).

Beyth-Marom (1982) challenged the interpretation of the results of the studies which stated that subjects cannot accurately evaluate the strength of relationship of binary variables. She suggests that subjects may in fact have been doing what they were instructed to do.

By analyzing the instructions used in several prior studies, Beyth-Marom showed that the instructions given to subjects in the Smedslund study focused subjects' attention on the frequency of Cell A relative to the

number of cases, and not on the correlation at all. Ward and Jenkins (1965) had subjects evaluate data that was constructed in terms of how much control cloud seeding had on rain. Beyth-Marom noted that subjects were actually instructed that "complete control means that whenever you seed, it rains, and whenever you don't seed, it doesn't rain", which she maintains, focused subjects on only the confirming cases in Cells 'a' and 'd' of the 2x2 table.

Instructing subjects to use information in a particular way, especially when those instructions are unintentional, raises a twofold problem. First, the accuracy of subject's judgments were then analyzed in terms of how well they approximated a normative model, which weighted information from each cell equally. So, studies in which this problem is not controlled for will mistakenly strengthen the concept that people are not very good at evaluating this sort of information. Secondly, it makes comparing the results of these experiments a risky endeavor, since subjects were not necessarily doing the same tasks, but merely similar ones.

Apart from the instructions themselves, other task characteristics have been shown to affect performance. For example, Einhorn and Hogarth (1986) suggest that the labelling of variables may affect understanding in

another way. Information about variables was couched in terms of either causal or coincidental contexts, and either forward or reverse inferential contexts. In their discussion of research on subjects understanding of causality, they conclude that the type of instructions given, and the type of questions asked of the subjects strongly affect the degree to which subjects weigh information in different cells. Causally focused subjects pay more attention to cells in which the cause (present or absent) brings about a positive effect (a and c cells) than a negative effect (b and d cells). According to Einhorn and Hogarth, this difference in attention does not occur in tasks where the context/instructions suggest no causal connection between the variables. For a more thorough discussion of this subject, see also Crocker (1981).

The context within which the subjects view the stimulus information is often defined in terms of a cover story. This gives them a basis for reasoning about the variables presented, and also gives a reason for the variables being presented in certain units. Wright and Murphy (1984) suggest that when subjects have a theory from which to consider the information presented, they perform better than when no context is present.

Lane et al. (1985) elected to use an abstract context setting (no cover story) in presenting stimulus

information. While this avoided the problem of whether the context of the variables suggested any causal framework which may have confounded the results, it also left the subjects with no theoretical base from which to evaluate the variables. Wright and Murphy (1984) showed that any theory which helps the subjects think reasonably about the stimuli is helpful. Wright and Murphy, in their brief review of the utility of theories about data suggest that 1)"People seldom collect data without a theory in mind", 2) "That a theory may help people make judgments--even when the theory is at odds with the data to be judged", and that 3) "having a theory may make people more resistant to noise and may thereby engender more accurate judgments." (pp. 303-304). The results of their experiments do indicate that having a theory about the data to be judged, whether it is a good theory or not, is better (in that it results in better and less variable subject judgments) than not having a theory.

Because of the importance of the questions at hand and some of the concern we had about the Lane et al. (1985) study, we designed the present study to investigate two questions: 1) What are subjects sensitive to in tabular format covariation situations when they are asked to judge the strength of relation? The basic paradigm contrasted effects of slope, variance of X, and error variance at different levels of $\underline{r}$. 2) How does this

sensitivity vary as a function of what task the subjects are instructed to perform, and how they are instructed to understand the task? By having subjects evaluate strength of relation while responding from different instructional frameworks, we expected to get a better idea about what subjects consider to be "important information" within a relation. The Lane et al. (1985) instructions may have forced subjects to pay close attention to error variance by their scale being in terms of "perfect linear relation." Subjects in the graphical format would be almost forced to react to the degree to which the scatterplot did not form a straight line, which is very dependent on error variance. Subjects may have a default strategy for approaching such a task, but this strategy may be superseded by instructions that suggest a different approach. If this is the case, then the results of Wright and Murphy (1984), who instructed their subjects to think of the strength of relation in terms of the predictability of one variable from another, are not necessarily comparable to the results of other studies where subjects were given instructions that gave no such suggestion.

We examined whether the effects of instruction on covariation estimation behavior using instructions which were in terms of 1) a neutral interpretation of the term strength of relation, 2) an instruction set which

suggested an approach to judging strength of relation based on predictability, and 3) an instruction set that focuses on the idea of error variability as a means of understanding strength of relation.

CHAPTER II

METHOD

Subjects

One hundred and forty-four undergraduates were recruited to participate in this experiment, seventy-two from the University of Massachusetts at Amherst, and seventy-two from Keene State College. All subjects received extra credit in psychology courses for participation.

Materials

Twelve sets of stimulus materials were formed by varying slope (b), variance of X (S_x^2), and error variance (S_{yx}^2), so as to have correlations of a given strength which had different constituent components. It should be noted that, except for conditions in which b=1, these are the values and combinations as were used in Lane et al. (1985). The b=1 conditions were added to increase the number of data points in the design, and the range of the stimuli. Table 1, below, contains the correlation coefficients, and how they were derived. For example, a correlation of r=.53 could have b=2, S_x^2=100, and S_{yx}^2=1000. It could also have b=1, S_x^2=400, and

s_{yx}^2=1000, or $\underline{b}$=4, s_x^2=100, and s_{yx}^2=4000, or $\underline{b}$=2, s_x^2=400, and s_{yx}^2=4000.

Four data sets were created, which contained all 12 combinations of $\underline{b}$, s_x^2, and s_{yx}^2, but included different values of X and Y. This was done to avoid basing conclusions on the idiosyncrasies of a particular data set.

Table 1

Correlation coefficients for stimulus sets

		SLOPE					
		1		2		4	
	s_{yx}^2=	1000	4000	1000	4000	1000	4000
s_x^2	100	.30	.16	.53	.30	.78	.53
	400	.53	.30	.78	.53	.91	.78

Subjects were randomly assigned to one of three instruction groups (See Appendix A for instruction sets). The instructions in Group J (judgment only, neutral instructions group) instructed subjects to look at each stimulus set, and to evaluate the strength of relation, but did not suggest a means of interpretation of the term

"strength of relation." In Group P (prediction group) subjects were instructed to think about strength of relation in terms of predictability, and were presented examples of perfectly correlated and completely uncorrelated stimulus sets. These subjects performed the same evaluation as Group J after predicting six missing Y-values from six extra X-values. Group F (focused prediction group) required subjects to perform the same task as Group P, but stressed the effects of error variance on predictability. The instructions for Group F were designed to approximate as closely as possible, in a tabular format situation, the way in which the Lane, et al (1985) graphical format condition subjects may have interpreted the instructions in their task.

A single cover story was used to provide a context for the variables presented (Wright and Murphy, 1984). It was written in terms of the relation between a fictional Generalized Professional Aptitude Test and starting salary following graduation for a given fictional student. The cover story was constructed so as to not give subjects any predisposition to expect specific degrees of relation. (See Appendix B for cover story).

At the bottom of each stimulus set a number line was presented on which the subjects were instructed to indicate his or her judgment of strength of relation. The number line ran from 0-100, with 0 labelled "no

relation", and 100 labelled "perfect relation". There was also a space for subjects to write in their numerical judgment value.

Stimuli

Stimulus sets consisted of 14 X-Y pairs with a specified correlation, as well as a specific slope, variance of X, and error variance. Stimulus sets were constructed to have particular characteristics, and were developed as follows: First, using SYSTAT, a set of seven z scores (z_x's) was randomly selected from a normal distribution. Then, a second set of seven z scores was selected from a normal distribution. By performing the regression of the second set of z scores on the z_x's, seven residuals (z_e's, uncorrelated with the z_x's, were obtained. Both the z_x's and the z_e's were then standardized. The eighth through the fourteenth values of the z_x's were obtained by multiplying each z_x by -1. For example, if the first value was -1.033, then the fourteenth value became +1.033.

The eighth through the fourteenth z_e's were obtained in a similar manner. Both of these sets of scores were then re-standardized. The sets of X values were created with the appropriate means and standard deviations using the formula:

$$X = \underline{z}_x S_x + \bar{X}$$

where S_x was either 10 or 20, and the mean of X was 200. Then Y values that had the correct mean, standard deviation, slope, and error variance for the particular condition were obtained using:

$$Y = \underline{bz}_x S_x + \bar{Y} + \underline{z}_e S_{yx}$$

This resulted in two variables with the desired correlation, $\underline{b}$, S_x^2, and S_{yx}^2. These sets of 14 X-Y pairs were then ordered on X.

For the two prediction groups, 6 additional values of X were interspersed within each set of 14 pairs, and were paired with blank lines in place of the Y-values. Subjects were instructed to predict the value of the Y's that was paired with each X. These X-values represented standardized scores of +/- .5, 1.0, and 1.5, but did not duplicate any score appearing in the set of generated pairs. Subjects in Group J had just the 14 generated pairs.

Design

Each subject was randomly assigned to one of the three groups. Each subject was presented with all 12 stimulus sets, and made judgments (and predictions for the Group P and Group F subjects, prior to those judgments) for each of them. The order of the stimulus sets given to each subject was balanced using a Latin Square design, in which the initial order was randomized.

Procedure

Subjects were given the stimulus booklets in small groups (between 2 and 10 subjects at any session). A session took approximately one hour.

The instructions informed the subject that the task was self-paced, that they had as long as they needed to complete the task, and that calculating devices would not be allowed.

The experimenter also informed subjects that they could ask questions at any time. The few questions that were asked were primarily about procedure. The experimenter answered these questions in a general and non-prejudicing manner.

CHAPTER III

RESULTS

It was decided *a priori* to delete the data of any subject who judged the stimulus set with r=.16 to have a stronger relation than the stimulus set with r=.91. The data for nineteen such subjects was deleted (three from Group J, and eight each from Groups P and F). Instances in which there were ties (12 cases) were retained.

A 2 x 3 x 3 x 2 x 2 analysis of variance was performed on factors of school, (University of Massachusetts and Keene State College), instruction group, slope, variance of X, and error variance. As expected, there were no differences between the two subject pools used, and so school was dropped as a factor in all subsequent analyses.

We performed a repeated-measures analysis of variance on the judgment scores, to observe any Group by factor interactions. It should be noted that although the three stimulus variables (b, S_x^2, and S_{yx}^2) were varied factorially in the experiment, they are confounded with the objective level of the correlation coefficient (see Table 1). Thus, when slope was high (b=4), the mean value of r was .75, and when slope was low (b=1), the mean value of r was .32. Similarly, when variance of X was

high (S_x^2=400), the mean value of $\underline{r}$ was .64, but when variance of X was low (S_x^2=100), the mean value of $\underline{r}$ was .43. Also, when error variance was low (S_{yx}^2=1000), the mean value of $\underline{r}$ was .64, but when the error variance was high (S_{yx}^2=4000), the mean value of $\underline{r}$ was .43.

Lane et al. (1985) dealt with the confounding of level of slope, variance of X and error variance variables with the correlation by considering a set of specific contrasts within certain levels of $\underline{r}$. They made specific comparisons at certain levels of $\underline{r}$, as follows: "...judgments of relatedness in the low slope, low error variance, high variance of X condition were quite a bit higher than judgments in either the high slope, low error variance, low variance of X condition or the high slope, high error variance, high variance of X condition even though Pearson's correlation was .78 in all three cases." (p. 644). This particular comparison (the statistical significance of which was determined by Newman-Keuls test) would support a statement that slope was less important to subjects' judgments of relatedness than were error variance or variance of X.

Another way of determining whether there are effects of $\underline{b}$, S_x^2, and S_{yx}^2 over and above the way in which they contribute to the correlation coefficient is to regress subjects' judgments on $\underline{r}$, as well as $\underline{b}$, S_x^2, and S_{yx}^2. If the population regression coefficient for a given

predictor variable differs from zero, this means that the variable makes a contribution to predictability over and above the contributions made by the other predictor variables. Both of these methods of inquiry were used in our attempt to understand subject judgment behavior, and those results follow.

For the analysis of variance, significant main effects for $\underline{b}$, S_x^2, and S_{yx}^2 were obtained, $\underline{F}(2,244)=100.52$, $\underline{F}(1,122)=60.3$, $\underline{F}(1,122)= 83.86$, respectively, all $\underline{p}$'s=.0000. Subjects judged relations to be stronger when slope was high, when variance of X was high, and when error variance was low. These means are reported in Table 2, below. Also, the means for each factor, by Group and by level of each factor appear in Table 3 (also below).

There was a significant slope X error variance interaction, $\underline{F}(2,244)=4.61$, p=.0108. These within-subject interactions are difficult to interpret, due to the confounding of the factors ($\underline{b}$, S_x^2, and S_{yx}^2) with the correlation coefficient. However, due to the design of the stimuli in this experiment, some comparisons can be made that allow us to assess the effects of certain variables independent of the correlation coefficient

No main effect of Group on strength judgments was found, $\underline{F}(2,122) <1$. However, there was a significant Group X S_x^2 interaction, $\underline{F}(2,122)=11.73$, $\underline{p}$=.0000. As is

shown in Figure 1, there is a large difference between subjects' judgments at high and low levels of S_x^2, (as reflected in the mean judgments of strength of relation), in Group J. This difference is less pronounced in Group P, and in Group F there appears to be almost no effect of different levels of S_x^2 on subject judgments.

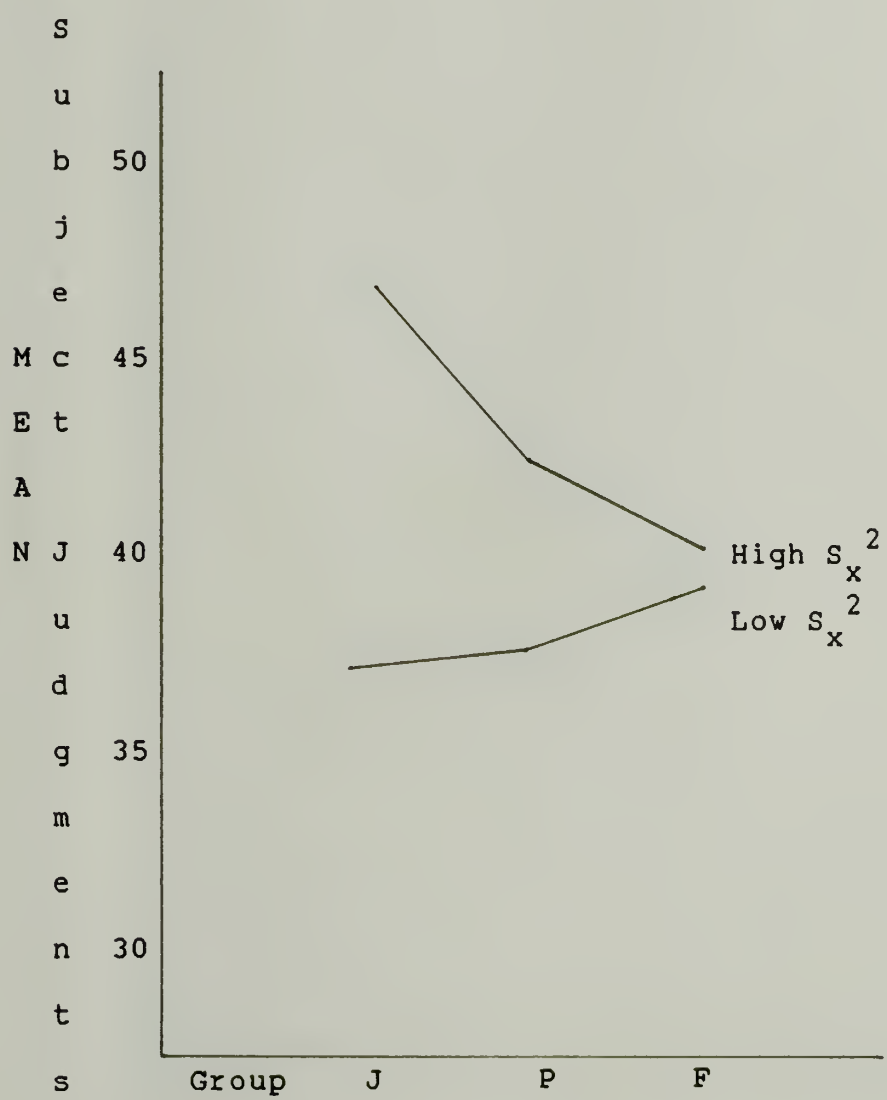

Figure 1. Group by S_x^2 interaction.

Table 2

Mean subject judgment for each stimulus and instruction condition (Standard deviations in parentheses)

		Slope 1		Slope 2		Slope 4	
		s_{yx}^2 1000	s_{yx}^2 4000	s_{yx}^2 1000	s_{yx}^2 4000	s_{yx}^2 1000	s_{yx}^2 4000
Group	s_x^2						
J	100	32.5	28.0	38.4	31.7	45.8	42.7
		(20.1)	(18.9)	(21.2)	(18.5)	(24.2)	(19.5)
	400	43.2	31.3	53.6	37.6	67.4	48.9
		(22.9)	(21.6)	(21.8)	(22.6)	(17.7)	(24.7)
P	100	37.9	25.9	34.6	33.6	50.4	38.4
		(26.6)	(19.3)	(24.0)	(25.9)	(25.9)	(24.6)
	400	39.3	32.0	40.6	37.6	58.3	43.8
		(24.0)	(22.4)	(25.2)	(23.9)	(23.7)	(26.4)
F	100	37.8	28.4	40.0	34.9	52.1	37.1
		(22.1)	(18.9)	(23.7)	(21.0)	(23.1)	(19.4)
	400	38.7	28.0	40.9	38.2	53.6	41.8
		(22.9)	(25.9)	(23.6)	(25.4)	(22.6)	(25.3)

Table 3

Means for each level of each factor, by Group.

Group	b			s_x^2		s_{yx}^2		Overall
	1	2	4	100	400	1000	4000	
J	33.7	40.3	51.2	36.5	47.0	46.8	36.7	41.8
P	33.8	36.6	47.8	36.8	42.0	43.5	35.2	39.4
F	33.3	38.5	46.2	38.4	40.3	43.9	34.8	39.3
Means	33.6	38.5	48.4	36.6	43.1	44.7	35.6	40.1

Most interesting was the Group X s_{yx}^2 X s_x^2 interaction, $F(2,122)=5.19$, $p=.0068$. Figure 2 shows that, for cells where s_x^2 and s_{yx}^2 were both low versus when they were both high, an interesting pattern emerges. (Each of these plotted points is equated for correlation. The objective correlations combined to produce each point are .3, .53, and .78). In Group J there was no difference in judgments when both s_x^2 and s_{yx}^2 were high and when both were low (t=.120, df=39, p=.905). In Group P (in

which subjects were asked to predict certain Y's from X's prior to making judgments), the effects of error variance significantly

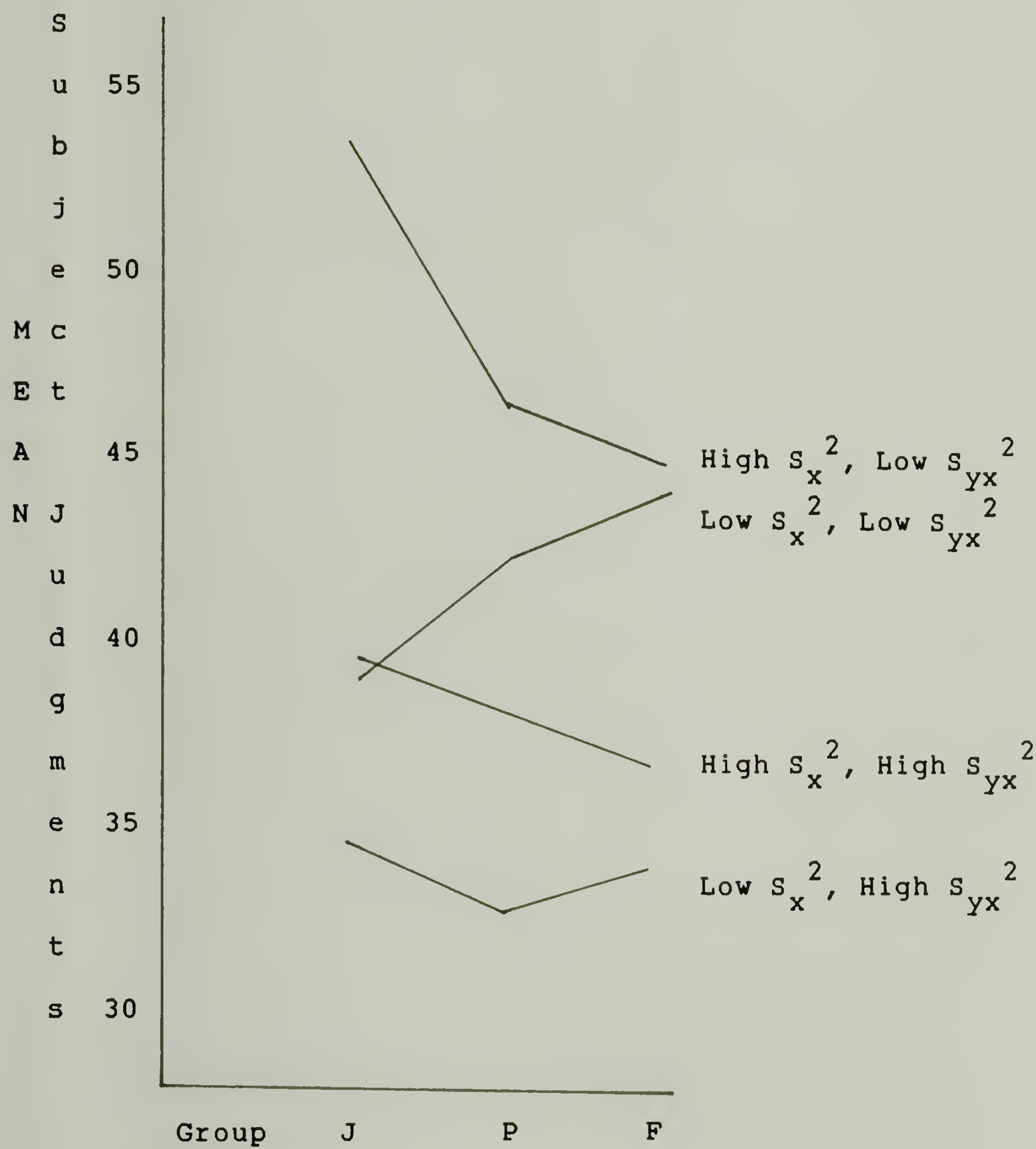

Figure 2. Group by S_x^2 by S_{yx}^2 interaction.

outweighed the effects of variance of X, (t=2.208, df=39, p=.033). That is, relations where s_x^2 and s_{yx}^2 were low were judged to be stronger than relations where s_x^2 and s_{yx}^2 were high. In Group F, the effects of s_{yx}^2 again significantly outweighed that of s_x^2, (t=3.464, df=39, p=.001).

We then looked to see whether the difference between the size of this effect was significantly different between groups. The effect in Group F was greater than in Group J (t=3.21, df=39, p<.01). The effect in Group F was also greater than in Group P (t=1.88, df=39, p<.05). However, there was no difference between Group P and Group J (t=1.33, df=39, p>.05), all one-tailed tests.

There is an interpretation of subject behavior which is supported by this Group by s_x^2 by s_{yx}^2 interaction. When subjects are put into a situation in which they are asked to predict Y's from given X's, this active participation causes them to be more sensitive, directly, to the amount of error in the relation represented. Further, when subjects are focused on the effects of error (by means of instructions in this experiment, and by means of the graphical format and instructions in the Lane et al. (1985) study) this sensitivity becomes more acute, to the extent that subjects are less aware of other factors which are normally evaluated; specifically, variance of X. (Alternatively, it could be argued that

subjects may be sensitive to the ratio of S_y^2 to S_x^2. Since this ratio changes as S_{yx}^2 changes, it is not possible to differentiate the effects of S_{yx}^2 from S_y^2/S_x^2 in this experiment. (Chumbley, personal communication)).

As mentioned earlier, regression analysis provides another way to deal with the confounding between correlation and the other variables. Regression analysis allows us to focus on one particular predictor variable by partialing out the effect of all of the other variables, so that we may ask whether, and to what extent, it can add to the predictability of subjects' responses. Partialing out the effects of variables refers to holding the variables at certain levels, so that one can observe the behavior of the variable in question. Judgments for each subject were regressed on the four predictor variables of interest. This regression provided us with four partial regression coefficients corresponding to $\underline{r}$, $\underline{b}$, S_x^2, and S_{yx}^2 which best predict the twelve judgments made by each subject.

Table 4 presents the correlation matrix for the predictors, along with means and standard deviations for each of the predictor variables. Each measure is based on the twelve values of the stimulus sets used in the experiment. It should be noted that the means and standard deviations are based on only two or three

different values. For example, S_{yx}^2 can be only 1000 or 4000 in this experiment.

TABLE 4

Intercorrelations, Means, Standard deviations, and Tolerances for the predictor variables

Predictor variable	r	b	S_x^2	S_{yx}^2
r	1.000			
b	.750	1.000		
S_x^2	.448	.000	1.000	
S_{yx}^2	-.448	.000	.000	1.000
Means	.536	2.333	250.	2500.
SD	.239	1.303	156.670	1566.699
Tolerance	.048	.081	.204	.204

Note. r=correlation; b=slope; S_x^2=variance of X; S_{yx}^2=error variance. Tolerance = 1 - the square of the multiple correlation of one predictor variable with the remaining predictor variables. (See footnote 1)

These predictor variables have widely different ranges, (i.e. r can change only .75 units, from .16 to .91, while S_{yx}^2 can change by 3000 units). The actual value of the regression coefficient, therefore, is not very helpful in evaluating the size of the effect produced by a predictor variable, in comparison to the

other predictor variables. For this reason, we have chosen to present what is referred to as a semistandardized regression coefficient. These coefficients are the product of Beta'(the standardized regression coefficient) and the standard deviation of the criterion variable. The semistandardized regression coefficient indicates the change in subject's strength of relation judgments with one standard deviation unit change in the predictor variable. Using this will allow us to compare the size of the effects of the predictor variables. (From Balota and Chumbley, 1984). Table 5 presents the mean semistandardized regression coefficients and the standard deviations for each predictor variable.

The first analyses performed on these semistandardized regression coefficients were tests of whether each of the predictor variables had an effect on subjects' judgments of strength of relation over and above the effects of the other predictors. This was done by testing whether the mean of each of the predictor variable coefficients was different from zero for each group. As can be seen in Table 5, in Group J and Group F there was no unique contribution of $\underline{r}$ to subject judgments after the effects of $\underline{b}$, S_x^2, and S_{yx}^2 were partialed out. This was not the case in Group P, which due to a few large negative values, had a mean value of

Table 5

Mean semistandardized regression coefficients for each Group.

	r	b	s_x^2	s_{yx}^2
Gp. J	-.000478	7.522**	5.483**	-4.7001***
	(22.20)	(17.51)	(11.44)	(9.40)
Gp. P	-8.508*	12.621***	6.423***	-7.834***
	(22.17)	(15.49)	(11.28)	(12.53)
Gp. F	.5466	5.095	.6267	-4.7001***
	(21.03)	(16.89)	(9.24)	(10.97)

Note. Standard deviations are in parentheses. *=.05, **=.01, ***=.001 p-values.

-8.508, (which was significantly different from 0, $p<.05$). This caused us some initial concern, since it seemed most unlikely that the level of correlation would have such a large negative effect on subject judgments. For that matter, it seemed odd that as r increased, subject judgments would decrease at all. The same analysis was performed after deleting coefficients that were more than one standard deviation from the mean value of the coefficients for each predictor variable, and it

was found that the pattern of effects was identical to the original pattern, (that is, each of the mean values that were significant in the original analysis continued to be significant) except that the mean value for *r* in Group P was no longer significantly different from zero. That *r* was no longer significantly different from zero suggests that this apparent effect was due to a few large negative outlying scores, and is not a systematic effect of the predictor variables on subject judgments.

The results in Table 5 indicate that subjects can be focused, by means of instructions, on particular aspects of a set of covarying stimuli. For example, in Groups J and P, *b*, S_x^2, and S_{yx}^2 all contributed significantly to subject judgments. This suggests that subjects were, in both of these groups, paying attention to all of these variables. However, when subjects are focused on error variance, (Group F), S_{yx}^2 is the only predictor variable that contributes uniquely to subject judgments.

A two-factor mixed design ANOVA was performed on the regression coefficients of each group. This allowed us to examine the effect of any predictor variable on the judgment behavior by Group interaction. The results of this analysis suggest that the subjects were differentially sensitive to characteristics of the stimulus sets in different instruction groups, $F(6,366)=2.154$, $p=.047$. We then performed 4 one-way

ANOVA's, one on the coefficients of each predictor variable. The only variable that differed significantly by Group was s_x^2, $F(2,122)=3.355$, $p=.038$. (The other variables, r, b, and s_{yx}^2, had F-values of 2.205, 2.126, and 1.207, respectively, all $p>.05$). We then performed t-tests to examine this difference of the effect of s_x^2. There was a difference between Group J and Group F, ($t=2.140$, $df=39$, $p=.039$), as well as between Group P and Group F, ($t=2.139$, $df=39$, $p=.039$). Group J and Group P were not statistically different ($t=.314$, $df=39$, $p=.755$). What this tells us is that subjects used information about s_x^2 differently due to instructions. Specifically, this shows that subjects tend to stop using information about s_x^2 and slope when instructed to focus on error in making their evaluations about strength of relation.

CHAPTER IV

CONCLUSIONS

One objective in this project was to further our understanding about what people are sensitive to when evaluating relations. Specifically, we examined four different descriptors of linear relation; slope, variance of the X-values, and error variance, as well as a particular combination of these elements, the Pearson product-moment correlation coefficient. This last descriptor was included in our study due to its popularity as a normative criterion by which performance in covariation estimation tasks is evaluated. The preference for r as a normative model has recently been challenged in the covariation literature. Jennings et al. (1982) showed that mean performance in their study could be better characterized by $1-\sqrt{1-r^2}$. Wright and Murphy (1984) suggested that a measure not as sensitive (as Pearson's r) to outlying data points may be more descriptive of subjects' estimation performance. Lane et al. (1985) suggest that subjects are more sensitive to differences in the amount of error variability in a data set than is the Pearson correlation coefficient. Taken together, these studies suggest that the use of Pearson's

r as a criterion by which to evaluate the quality of a subjects performance on a covariation estimation and prediction task may not be the best approach.

In the present study, it was found that subject judgments were more affected by the different levels of error variance than was the objective correlation. Specifically, we found that subjects judged the strength of a relation to be higher when error variance was low than when error variance was high, (with corresponding changes in $\underline{b}$ and S_x^2), for stimulus sets with the same correlation.

Another equally important focus of this study was to observe whether, and how, subjects used different indicators of linear relation when the task instructions were changed. As expected, we found that subjects were more sensitive to differences in error variance in a task that required them to predict Y-values from X-values and instructed them to think about strength of relation in terms of how inaccurate their predictions may have been, (which is akin to their degree of confidence in their predictions) prior to estimating strength of relation than they would be in a task where they were only required to estimate strength of relation. As can be seen in Figure 2, in the results section, judgments of strength of relation were significantly higher for low S_{yx}^2 than for high S_{yx}^2 conditions, regardless of level

of S_x^2, across level of slope, in Group F than in Group P. This difference was also apparent between Group F and Group J.

Cumulatively, what these results suggest is that, (a) estimation behaviors which occur in an prediction+estimation (Groups P and F) and estimation-only (Group J) tasks are not the same, (b) these differences result in systematic differences in subject judgment behavior. This systematic difference can be examined by observing to what extent any particular element of linear relation is able to describe these differences, and (c) subject judgment behavior can be manipulated **in a predictable manner** merely by suggesting a means by which the covariation estimation task can be understood, without going into any mathematical explanations.

Another interesting development was that subjects had a tendency (non-significant) to be more sensitive to changes in level of slope in Group P than in Group J. This tendency makes sense, since one way to approach prediction is to see how much one variable changes on the average as the other variable changes, which is slope. Alternatively, this effect may simply be due to the amount of attention invested in the task, since a subject in this task could select her own subset of information to pay attention to. This idea that differences in

attention-investment resulting in differences in subject judgment behavior would mesh well with theories put forth by both Yates and Curley (1986), and Brehmer (1979), who suggest that overt differences in subject response behavior may be due to attentional constraints.

It should be emphasized that these findings are of particular importance to the study of estimation behavior, since they underscore a problem in the literature: studies of estimation behavior have used different kinds of instructions and different kinds of stimuli, which limits what they can say collectively about estimation behavior.

Two other paths of inquiry are opened at this point. First, the question of the effects of instructions on subject covariation estimation behavior has only begun to be explored systematically. Since it has been shown that subjects can be focused on error variance as a means of understanding covariation, it would be interesting to see whether one could also elicit increased sensitivity to other elements of linear relations by means of instruction. If so, this would suggest that subjects have no hard and fast rule by which to make estimates of strength of relation, but that given a certain context or situation (experimentally, instructions), they have different foci, or goals which are altering how they approach the task. This in itself suggests that it may

not be meaningful to discuss the idea of accuracy of subject judgments, except in the context of particular tasks, given particular constraints. And if this is the case, the construction of instructions and stimuli for subjects in such a task needs to be carefully considered.

A second path is to examine subject sensitivity to s_{yx}^2 further. Just as the correlation coefficient can be thought of as a combination of $\underline{b}$, s_x^2, and s_{yx}^2, s_{yx}^2 itself can be reduced to its component parts. A certain amount of error can be distributed in a number of different ways. All of the 'pieces' of error could be of a moderate size, or some could be quite small and others correspondingly quite large. So, from a graphical standpoint, what would be manipulated would be the shape of the envelope that includes the points comprising the correlation. Or, from a statistical standpoint, it is the degree of homoscedasticity that the relation contains that would be manipulated. If, as Wright and Murphy(1984) suggest, a measure that discounts (or underweighs) outlying data points is more descriptive of subject judgment behavior, then it seems reasonable that the limits of acceptable data could be better understood by performing a manipulation of this sort, and these limits could then be tested in conjunction with other characteristics of the stimulus environment.

Footnote 1

The tolerance values given in Table 4 are measures of the extent to which a given predictor variable is not a linear combination of the other predictor variables in the regression. The maximum tolerance value possible is 1.0(totally orthogonal predictor), and the minimum is 0.0(totally predicted from the other variables)(From Chumbley and Balota,(1984))

As can be seen, the tolerances here are extremely low, indicating that each of these predictor variables can be expressed to a certain extent as a linear combination of the others. However, this low tolerance raises a problem of interpretation of the regression results, since a small value of tolerance indicates a high level of multicollinearity, which can provide unstable and distorted estimators of the population (Pedhazur (1982)). This can seriously impair the experimenters' ability to interpret the results of the analysis with confidence.

One way to attempt to understand data suffering from difficulties with high multicollinearity is to abandon the standard ordinary least squares analysis, in favor of a method which yields somewhat biased estimates of the population regression coefficients, but with much smaller standard errors. (For a fuller discussion of difficulties associated with high multicollinearity, see

Pedhazur(1982)). One such method is ridge regression. (for a discussion of the ridge regression method, see Draper and Smith (1981)).

When a ridge regression analysis was performed on the present data, the same story emerged as with the original regression analysis; the regression coefficients became less variable, but did not seriously change in their other characteristics (i.e. by changing sign). Therefore, we present only results from the ordinary least squares analyses, since the problems associated with multicollinearity seemed to be less drastic in this data than is often the case with data that has a high multicollinearity.

APPENDICES

Appendix A:
Instruction Sets

{Sample Group J instructions}

Instructions

In our experience, we know that some things are strongly related to others, for instance, a person's height and the length of their arm. We also know that some things are not strongly related to each other, like a person's height and the amount of iron in their blood. The first is considered a strong relationship because if a person is tall, it is likely that all of the other bones in their body are long as well.

Some relationships have an obvious cause-effect explanation, like the age of an elm tree and its height. This may not be a perfect relationship, (as you can probably think of reasons for an old elm tree to be short), but it is a pretty strong relationship, as age is obviously a cause of height in elm trees.

Some relationships do not have any cause-effect explanations, or it may be that both of the components are caused by something else. However, they may still be strongly related. For example, the health of the stock market and the length of the new fashions in women's dresses are strongly related (REALLY!) though no-one knows why. It's just something that someone noticed and measured over a period of time.

What we would like you to do is to evaluate the information in each of the following cases, and make a judgment about how strongly or weakly related the components are. To do this, put a mark on the number line at the bottom of the page of information. The number line runs from 0 -> no relation to 100 -> perfect relation. Then write in the number you marked in the space provided.

There are 12 sets, and they should take you several minutes to accurately make the judgment about how strongly related the information is. Please take your time and be as accurate as you can. You may take as long as you wish; there is no hurry. Also, feel free to reread these instructions at any time. And feel free to ask the experimenter questions at any time.

You may of course withdraw from this experiment without beong penalized in any way. Any information we obtain from you will be kept strictly confidential.

Please begin the experiment when you have finished reading these instructions. Thank You.

{Sample Group P instructions}

Instructions

In our experience, we know that some things are strongly related to others, for instance, a person's height and the length of their arms. Tall people are most likely to have long arms, and short people are most likely to have short arms. We also know that some things are not strongly related to each other, like a person's height and the concentration of iron in their blood.

One way to think about strength of relationship is in terms of predictability: given some information about the relationship between things and the corresponding value of the one event or thing, we can make a good prediction about the other event or thing. For example, try to predict the missing Y scores in the table below, by using the X and the the relationship of X to Y. Fill in the blanks and then continue reading:

	X	Y	W
Case 1	50	150	250
Case 2	60	___	?
Case 3	75	200	200
Case 4	100	250	350
Case 5	120	___	?
Case 6	125	300	300
Case 7	150	350	150

Looking at this example, you can see that as the X values get larger by a certain amount, the Y values also get larger by a certain amount. In this example, there is perfect predictability. As X gets larger by 10 points, Y gets larger by 20 points, and this happens at every point. This is an example of a perfect relationship.

Most relationships are far from perfect, however. For example, if you were given the X's above and asked to predict the W's, it is unlikely that you could accurately do this. When X gets larger by a certain amount, W doesn't seem to get larger in any corresponding way. This is an example of no relationship, where knowing the X value tells you nothing about what the W value might be.

All of the sets that you work with today will be more predictable than W, but less predictable than Y.

Your task involves two parts. First we would like you to look at the information on a given page and then

make the best predictions you can using that information. Second, we would like you to make a judgment about how strongly related the two columns are. To do this, place a mark on the number line at the bottom of the page of information. The number line runs from 0 -> no relation to 100 -> perfect relation. Finally, write in the number you marked in the space provided.

There are 12 sets, and each will probably take you several minutes. You should carefully decide on which values to write in the blanks. This is a difficult task, so please take your time and be as accurate as you can be. You may take as long as you wish; there is no hurry. Also, feel free to reread these instructions at any time, and to ask the experimenter questions at any time.

Please begin the experiment when you have finished reading these instructions. Thank You.

{Sample Group F instructions}

Instructions

In our experience, we know that some things are strongly related to others, for instance, a person's height and the length of their arms. Tall people are most likely to have long arms, and short people are most likely to have short arms. We also know that some things are not strongly related to each other, like a person's height and the concentration of iron in their blood.

One way to think about strength of relationship is in terms of predictability: given some information about the relationship between things, and the corresponding value of the one event or thing, we can make a good prediction about the other event or thing. For example, try to predict the missing Y scores in the table below, by using the X and the relationship of X to Y. Fill in the blanks and then continue reading:

	X	Y	W
Case 1	50	150	250
Case 2	60	___	?
Case 3	75	200	200
Case 4	100	250	350
Case 5	120	___	?
Case 6	125	300	300
Case 7	150	350	150

Looking at this example, you can see that as the X values get larger by a certain amount, the Y values also get larger by a certain amount. In this example, there is perfect predictability. As X gets larger by 10 points, Y gets larger by 20 points, and this happens at every point. You also see that large values of X are paired with large values of Y, and small values of Y are paired with small values of X. This is an example of a perfect relationship.

Most relationships are less than perfect, however. You can think about the strength of relationship getting smaller if there are fewer small values paired with small values, and fewer large values paired with large values. The less the large values of one set are paired with large values on the other set, the lower the strength of relationship.

Also, one set of values is supposed to get bigger by a certain amount as the other gets bigger by a certain amount. The degree to which the amount of change is different from what you would expect is an indicator of strength of relationship. The more the amount of change differs from your expectations, the lower the strength of relationship. For example, if you were given the X's above and asked to predict the W's, it is unlikely that you could accurately do this, because there seems to be little correspondence between the sets. When X gets larger by a certain amount, W doesn't seem to get larger in any corresponding way. This is an example of no relationship, where knowing the X value tells you nothing about what the W value might be.

All of the sets that you work with today will be more predictable than W, but less predictable than Y.

Your task involves two parts. First we would like you to look at the information on a given page and then make the best predictions you can using that information. Second, we would like you to make a judgment about how strongly related the two columns are. To do this, place a mark on the number line at the bottom of the page of information. The number line runs from 0 -> no relation to 100 -> perfect relation. Finally, write in the number you marked in the space provided.

There are 12 sets, and each will probably take you several minutes. You should carefully decide on which values to write in the blanks. This is a difficult task, so please take your time and be as accurate as you can be. You may take as long as you wish; there is no hurry. Also, feel free to reread these instructions at any time, and to ask the experimenter questions at any time.

Please begin the experiment when you have finished reading these instructions. Thank You.

Appendix B:
Cover Story

Cover Story

A new study is being run, looking at the strength of relationship between the test scores of students on graduate school entrance exams, and their starting salary in their first job following graduation. We wanted to look at several schools and several professions to see how the strength of this relationship is different due to school and profession. You can imagine how this relation could be different. Some schools accept only the top-scoring students, and a degree from this school may command a big salary for its students. On the other hand, some schools are not as concerned with exam scores, (so their students may have a wider range of scores), and they may not have the prestige to command a high starting salary for their graduates. In the first case, the score doesn't tell you much about what the student is earning, because everyone from that school earns about the same, and their scores are about the same. In the second case, the scores can tell you quite a bit, because a top scoring student at this school will probably get a better paying job than a lower scoring student, based on her ability, rather than the name of her school.

However, it is difficult to compare different professions due to their different entrance exams. The MCAT, GMAT, GRE, ASAT, LSAT, ETC., are all different in what they measure and how they scale their scores. So, a few years ago we developed the Generalized Professional School Test (GPST). A large number of students took this test along with the tests required by the schools they were applying to. After several years, many of them have graduated, and have communicated with us regarding their starting salaries.

What we want you to do is to make a judgment about how strongly related the test scores are (for a given school and profession) to the starting weekly salary of the graduates. We have coded the profession and school by number in each case so as not to confuse your evaluation of the strength of relationship.

Each page represents a different school and/or profession. Using the information on a given page, decide how strongly related test score and salary are to each other, and mark the number line at the bottom of the page appropriately. This task requires a good deal of thought, so take your time and be as accurate as you can be.

Feel free to look back at this page and at the instructions if something seems unclear. Please turn the page and begin. Thank You.

Appendix C:
Sample Stimulus Sets

{Sample Page of stimulus set (neutral condition)}
{$r = .30$, $b = 1$, $Sx^2 = 400$, $Syx^2 = 4000$.}

(please be as accurate as possible, and remember that each page represents different information.)

	Generalized Professional School Test score	Weekly Income immediately following graduation (first job)
Student A	167	990
Student B	176	1015
Student C	177	1004
Student D	186	904
Student E	187	1074
Student F	192	911
Student G	194	980
Student H	205	991
Student I	207	927
Student J	213	1101
Student K	214	931
Student L	222	1049
Student M	223	1061
Student N	232	1056

Now please indicate your judgment of how strongly these two tests are related on the number line below, and then write the number in the space provided.

0-------------25-----------50------------75-----------100

^ no relationship

^ perfect relationship

write judgment of strength of relationship (corresponding to your marked judgment) here:______

{Sample Page of stimulus set (prediction condition)}:
{$r = .30$, $b = 1$, $Sx^2 = 400$, $Syx^2 = 4000$.}
(please be as accurate as possible, and remember that each page represents different information)

	Generalized Professional School Test score	Weekly Income immediately following graduation (first job)
Student A	167	990
Student B	172	____
Student C	176	1015
Student D	177	1004
Student E	181	____
Student F	186	904
Student G	187	1074
Student H	190	____
Student I	192	911
Student J	194	980
Student K	205	991
Student L	207	927
Student M	209	____
Student N	213	1101
Student O	214	931
Student P	220	____
Student Q	222	1049
Student R	223	1061
Student S	230	____
Student T	232	1056

Now please indicate your judgment of how strongly these two tests are related on the number line below, and then write the number in the space provided.

0-------------25------------50-------------75------------100
^ ^
no relationship perfect relationship

write judgment of strength of relationship (corresponding to your marked judgment) here:______

Appendix D:
Raw Data

Raw Data

Loc	Ins	Sub#	Judgments 1-12											
1	1	001	25	40	60	50	30	65	80	10	50	30	90	15
1	1	002	30	65	50	25	50	50	60	60	50	45	90	60
1	1	003	15	00	35	25	45	15	75	35	25	45	85	75
1	1	004	00	75	50	25	00	50	25	25	25	25	75	75
1	1	005	25	25	50	50	50	25	50	75	50	75	75	75
1	1	006	55	40	55	40	60	55	70	35	50	50	80	50
1	1	007	65	60	40	71	75	60	80	50	75	75	75	75
1	1	008	40	22	37	07	50	20	71	09	18	28	42	47
1	1	009	20	10	50	30	30	30	30	30	40	30	40	40
1	1	010	23	20	60	55	25	65	52	18	48	75	45	40
1	1	011	30	20	30	30	30	40	40	20	40	35	60	50
1	1	012	10	60	25	25	20	30	75	20	40	20	80	20
1	1	013	01	01	05	03	03	00	20	02	09	05	25	12
1	1	014	50	25	75	50	75	25	75	25	75	25	75	75
1	1	015	25	25	00	25	50	00	50	25	50	25	50	25
1	1	016	50	50	50	50	25	50	75	75	75	75	50	75
1	1	017	52	86	68	75	65	50	27	57	26	26	48	78
1	1	018	25	20	30	30	30	25	40	25	10	30	70	50
1	1	019	25	15	27	30	20	12	40	25	20	30	70	40
1	1	020	80	21	70	29	52	15	98	60	52	72	85	15
1	1	021	25	20	20	20	25	23	30	30	20	15	20	26
1	1	022	30	10	45	30	20	20	30	35	25	50	55	30
1	1	023	25	20	40	15	85	30	75	20	80	50	80	60
1	1	024	20	15	40	40	45	30	55	30	85	40	90	70
2	1	025	25	25	75	75	50	50	75	50	50	75	75	75
2	1	026	00	15	85	35	15	25	50	65	50	50	75	75
2	1	027	25	25	25	25	70	25	40	50	20	10	50	20
2	1	028	10	05	05	05	10	10	05	10	05	70	75	05
2	1	029	21	50	35	30	15	30	40	63	48	35	80	82
2	1	030	25	25	60	25	50	25	05	40	25	05	30	25
2	1	031	75	25	00	50	00	00	50	75	50	75	75	75
2	1	032	45	30	15	05	50	08	60	28	40	20	80	75
2	1	033	00	25	60	00	30	35	05	60	75	50	65	50
2	1	034	45	00	90	20	95	25	70	10	65	30	75	85
2	1	035	25	00	50	00	00	99	25	00	50	00	75	00
2	1	036	50	60	25	50	10	25	50	25	40	60	60	40
2	1	037	25	20	50	14	65	25	25	25	75	50	50	30
2	1	038	55	50	25	07	45	10	10	40	60	30	40	25
2	1	039	80	65	60	55	70	55	70	65	85	75	80	75
2	1	040	25	10	30	10	25	25	50	20	25	23	50	20
2	1	041	40	45	48	45	48	20	60	55	42	75	80	75
2	1	042	25	20	45	25	25	25	75	50	50	45	75	70
2	1	043	60	15	70	50	50	20	70	45	30	40	85	30

(cont'd)

Raw Data (cont'd)

Loc	Ins	Sub#	Judgments 1-12											
2	1	044	65	40	35	25	55	50	65	75	75	60	65	60
2	1	045	05	25	10	36	25	20	20	15	10	10	20	15
2	1	046	30	28	28	12	65	28	83	20	34	15	87	17
2	1	047	35	35	60	60	35	40	75	50	80	40	70	45
2	1	048	37	37	47	37	00	47	70	70	37	70	70	70
1	2	049	25	25	25	25	25	25	50	25	75	25	75	50
1	2	050	00	25	25	25	25	00	75	25	75	25	75	25
1	2	051	60	30	50	60	75	46	40	40	75	55	60	70
1	2	052	25	25	10	10	10	20	10	25	25	10	35	25
1	2	053	52	55	70	25	45	76	49	70	90	70	55	75
1	2	054	55	15	60	30	50	60	60	60	40	60	80	75
1	2	055	45	45	75	45	55	60	40	55	50	50	75	70
1	2	056	49	48	39	47	35	47	47	27	36	15	37	40
1	2	057	40	50	34	55	30	55	70	60	70	75	65	30
1	2	058	70	55	55	75	90	59	51	55	95	49	95	87
1	2	059	76	50	40	23	60	00	85	53	48	60	72	70
1	2	060	95	78	08	21	83	40	85	62	78	73	26	29
1	2	061	25	25	75	25	25	50	25	50	50	50	50	50
1	2	062	40	10	35	45	30	65	65	40	55	65	15	30
1	2	063	00	15	00	00	00	05	10	25	10	25	25	10
1	2	064	50	25	60	10	25	40	10	25	50	25	45	40
1	2	065	30	25	45	25	30	40	25	60	70	50	55	40
1	2	066	25	45	25	45	65	25	50	25	15	25	25	50
1	2	067	20	25	50	55	65	30	30	60	80	70	60	70
1	2	068	50	00	50	00	25	50	50	50	50	50	75	50
1	2	069	00	00	05	00	25	25	25	00	25	25	50	75
1	2	070	30	45	60	35	30	45	30	20	45	15	80	20
1	2	071	00	62	00	10	10	25	47	05	40	15	45	15
1	2	072	75	25	80	75	40	75	60	60	90	50	80	75
2	2	073	06	05	07	15	10	06	04	02	27	04	73	02
2	2	074	30	46	20	40	50	15	50	35	50	40	45	45
2	2	075	75	40	42	99	99	70	50	62	90	50	80	83
2	2	076	15	10	15	10	15	25	25	30	25	25	50	25
2	2	077	75	20	45	25	70	30	70	45	60	65	80	55
2	2	078	00	25	25	00	25	00	25	00	25	25	25	00
2	2	079	60	00	50	25	71	10	42	60	33	72	68	81
2	2	080	70	45	65	65	80	45	70	85	85	65	80	75
2	2	081	30	10	20	10	19	45	40	10	20	20	19	10
2	2	082	10	00	10	00	25	10	00	05	00	05	00	00
2	2	083	15	00	00	25	00	10	10	00	50	00	75	00
2	2	084	45	30	30	35	30	25	50	49	30	35	55	45
2	2	085	20	40	70	50	20	65	65	55	75	65	80	60
2	2	086	18	10	20	10	12	10	25	25	20	20	37	35

(cont'd)

Raw Data (cont'd)

Loc	Ins	Sub#	Judgments 1-12											
2	2	087	25	05	25	23	03	03	02	25	25	02	12	12
2	2	088	25	10	25	20	20	20	50	25	25	40	55	37
2	2	089	45	50	50	50	50	45	50	55	63	45	40	45
2	2	090	55	84	78	25	31	75	45	79	70	89	74	75
2	2	091	10	10	10	10	15	10	10	10	20	10	20	10
2	2	092	25	05	00	10	05	00	25	15	45	10	00	15
2	2	093	65	65	45	90	70	75	80	65	70	80	80	45
2	2	094	75	50	80	25	00	50	80	50	50	25	80	50
2	2	095	99	70	70	75	25	00	30	25	60	00	75	50
2	2	096	10	20	15	25	15	10	35	20	60	25	65	40
1	3	097	15	30	50	25	60	15	45	20	50	25	75	50
1	3	098	02	00	00	00	00	00	03	03	22	04	25	15
1	3	099	70	55	60	50	45	40	50	45	60	65	75	55
1	3	100	25	25	50	25	25	50	00	00	75	00	25	00
1	3	101	35	20	37	10	30	17	45	60	70	54	65	35
1	3	102	25	00	00	25	25	25	50	00	50	00	75	00
1	3	103	55	70	60	40	80	40	70	65	75	25	45	50
1	3	104	65	50	10	48	25	50	60	50	50	25	60	75
1	3	105	25	50	65	50	20	75	25	60	75	60	75	75
1	3	106	75	70	75	75	75	60	99	90	99	94	99	85
1	3	107	20	25	55	30	45	40	10	45	65	15	75	45
1	3	108	80	25	75	00	30	40	50	10	75	80	25	75
1	3	109	35	20	30	30	30	20	20	20	55	20	64	20
1	3	110	25	25	00	50	50	25	50	25	25	25	25	25
1	3	111	30	10	55	10	25	60	45	40	10	30	15	25
1	3	112	50	50	55	50	60	50	50	50	55	50	55	55
1	3	113	80	60	07	50	49	60	60	10	25	35	50	25
1	3	114	50	25	75	25	85	40	75	25	50	25	50	75
1	3	115	50	10	60	05	30	10	70	60	75	50	85	40
1	3	116	53	30	30	23	40	45	47	67	77	29	62	78
1	3	117	30	27	50	25	50	30	50	60	80	55	75	70
1	3	118	75	50	50	50	95	50	50	75	75	50	75	50
1	3	119	30	20	60	15	50	45	30	45	50	45	45	60
1	3	120	25	35	30	10	53	50	60	54	24	52	60	27
2	3	121	20	15	11	12	20	15	20	15	32	05	30	20
2	3	122	00	15	25	30	50	10	00	15	00	25	10	25
2	3	123	50	00	10	00	00	10	20	00	10	00	10	00
2	3	124	75	30	20	60	80	50	80	90	85	85	75	50
2	3	125	44	36	35	47	32	41	31	47	47	32	47	47
2	3	126	25	02	05	05	26	30	45	02	25	10	00	24
2	3	127	50	60	60	40	60	50	30	80	50	60	40	50
2	3	128	62	67	34	44	79	48	70	60	80	50	63	53
2	3	129	34	20	50	22	47	10	53	30	45	36	50	60

(cont'd)

Raw Data (cont'd)

Loc
Ins
Sub#
Judgments 1-12

Loc	Ins	Sub#												
2	3	130	25	48	28	30	25	10	25	35	30	45	30	35
2	3	131	40	51	50	52	55	58	52	65	78	46	70	65
2	3	132	68	27	65	55	48	61	47	10	49	52	70	43
2	3	133	35	52	52	45	55	52	40	77	60	55	65	55
2	3	134	39	32	48	27	40	28	30	47	32	31	45	26
2	3	135	10	25	20	25	20	20	25	25	25	10	40	30
2	3	136	50	60	25	70	50	45	60	65	83	70	80	75
2	3	137	40	15	20	20	15	20	20	15	35	15	40	05
2	3	138	60	40	69	29	62	70	48	43	38	35	30	58
2	3	139	30	35	50	25	65	30	35	33	35	30	45	35
2	3	140	00	00	00	00	00	00	00	25	50	49	50	50
2	3	141	00	00	00	00	00	00	25	00	50	00	40	00
2	3	142	00	00	20	00	30	15	00	00	00	00	00	00
2	3	143	25	25	25	25	30	50	75	50	25	75	35	25
2	3	144	64	63	75	05	65	53	40	30	75	40	65	51

Appendix E:
Regression Coefficients

Group J Regression Coefficients

r	b	Sx^2	Syx^2
-44.8751	6.8857	0.0669	-0.0100
76.7923	-4.7661	-0.0220	0.0039
80.5072	1.0592	0.0478	0.0008
-71.6787	15.2211	0.1045	0.0007
-11.3116	11.8256	0.0772	0.0006
-9.0383	4.2779	0.0173	-0.0062
31.3450	0.8826	-0.0320	0.0011
137.7358	-17.1498	-0.0746	0.0025
-24.2491	6.7267	0.0499	-0.0039
-88.0982	16.5437	0.0679	-0.0049
-44.4690	12.4585	0.0498	-0.0050
95.1952	-9.9709	-0.0289	0.0023
10.1448	1.9962	0.0197	-0.0015
35.1242	-0.3714	0.0315	-0.0087
103.9489	-8.5022	-0.0710	0.0015
94.3834	-6.7358	-0.0368	0.0092
-74.0845	14.8372	0.1090	-0.0065
-98.9065	19.2281	0.1287	-0.0095
-22.4303	4.7716	0.0514	-0.0140
65.3801	-9.7624	-0.0347	0.0041
-140.8476	23.5750	0.1352	-0.0113
170.9308	-10.2974	-0.1168	0.0011
-7.9543	15.4589	0.0555	-0.0067
-31.8305	10.6222	0.1051	-0.0022
-10.9833	11.4222	0.1352	-0.0013
127.5397	-19.0573	-0.0678	0.0046
-260.6767	47.2866	0.1754	-0.0181
2.1153	9.2207	0.0713	0.0030
-32.9498	0.3338	0.0392	-0.0050
-171.0649	35.1314	0.1725	-0.0089
88.0594	-2.3108	-0.0212	-0.0009
-14.7038	15.0511	0.0239	-0.0018
66.5770	-1.0252	0.0045	-0.0104
-123.3295	20.5443	0.0709	-0.0154
-148.2357	23.0660	0.1041	-0.0087
2.1126	7.7181	-0.0381	-0.0069
-17.9980	7.3874	-0.0016	-0.0043
38.8851	-2.3162	-0.0005	-0.0027
-44.0882	14.3677	0.0818	-0.0032
94.6172	-2.9359	0.0186	0.0031
-125.7681	16.5929	0.1609	-0.0178
115.9426	-8.9027	-0.0903	0.0051
148.4456	-17.2553	-0.0753	-0.0014
30.6006	-0.3741	0.0319	-0.0026
3.0919	7.0088	0.0734	0.0041

Group P Regression Coefficients

r	b	Sx^2	Syx^2
11.5783	9.1210	0.0198	-0.0048
128.8233	-7.4481	-0.0326	0.0005
-73.9366	15.5091	0.0389	-0.0083
-40.0541	7.8328	0.0274	-0.0027
73.9130	-2.9849	-0.0749	0.0056
-16.7920	9.5358	0.0587	-0.0036
-75.6168	13.5254	0.0822	-0.0060
18.4381	2.2648	-0.0159	0.0021
-81.5836	17.5612	0.0558	-0.0098
77.4862	-5.3701	-0.0258	-0.0017
-64.8716	13.3855	0.0721	-0.0044
108.2749	-13.0271	-0.0935	0.0082
7.6184	3.4185	0.0031	0.0025
-137.6959	21.0819	0.0802	-0.0136
-27.5948	11.2910	0.0217	-0.0027
-18.7644	13.5639	0.0322	-0.0010
-21.0282	12.7195	0.0421	-0.0070
51.4957	6.8341	-0.0047	0.0032
-156.2587	21.4108	0.1262	-0.0160
-147.4355	24.2058	0.1424	-0.0137
-193.9107	33.3900	0.1576	-0.0184
-77.5104	14.4137	0.0485	-0.0071
-12.0845	7.5500	0.0305	-0.0019
-11.2191	9.1323	0.0077	-0.0097
38.9145	-3.1206	-0.0405	-0.0015
-62.6509	18.3875	0.0873	-0.0085
26.2881	1.1134	0.0098	-0.0021
62.9745	-9.5363	-0.0625	0.0019
17.8555	-4.1542	-0.0317	-0.0002
-213.2127	37.2798	0.1651	-0.0210
-3.3879	2.4486	0.0406	-0.0014
-13.8324	10.3885	0.0623	-0.0007
-5.2939	5.2980	0.0380	-0.0016
-89.7048	10.7340	0.0813	-0.0074
-4.9474	6.9332	0.0433	-0.0030
-10.7661	3.1760	0.0046	-0.0021
-31.9578	4.8460	0.0107	-0.0016
-56.2856	6.4019	0.1023	-0.0102
-373.5804	44.0575	0.2948	-0.0333
-27.0921	14.2647	0.0518	-0.0052

Group F Regression Coefficients

r	b	Sx^2	Syx^2
18.0533	4.3013	0.0266	-0.0060
-34.5710	10.3997	0.0336	-0.0040
-156.3304	24.2700	0.1068	-0.0135
-102.9364	12.8252	0.0148	-0.0126
26.1943	6.3940	-0.0034	-0.0030
-32.1981	10.2399	0.0358	-0.0119
34.4388	-1.6490	-0.0024	0.0039
-118.6577	24.9850	0.1061	-0.0034
38.0477	1.5672	0.0017	-0.0001
-91.0427	18.5970	0.0900	-0.0101
-49.1714	14.3565	-0.0193	-0.0092
-145.1386	24.3349	0.1015	-0.0157
110.0205	-16.0325	-0.0752	0.0075
113.4421	-18.9941	-0.0637	0.0075
4.9170	0.1282	-0.0034	-0.0008
167.5291	-21.7013	-0.0867	0.0020
20.6279	7.5226	0.0387	-0.0094
98.8095	-4.8979	-0.0492	0.0047
56.1769	4.4537	-0.0061	0.0001
17.4820	-0.9793	-0.0370	-0.0041
38.7768	0.5554	-0.0182	0.0007
155.8130	-17.9009	-0.1054	0.0093
16.4406	0.1150	-0.0107	-0.0017
-66.3905	6.0979	0.0287	-0.0095
91.0442	-4.7532	-0.0789	0.0034
-55.3529	8.7945	0.0500	-0.0030
52.7324	-1.7002	0.0045	-0.0020
47.9067	-1.3594	-0.0183	0.0028
-221.1110	31.2006	0.1427	-0.0206
29.1844	-0.1738	-0.0061	0.0036
-31.1421	3.2336	0.0329	-0.0045
11.7855	0.4319	0.0225	0.0005
-16.8160	11.2452	0.0209	0.0009
-112.1828	15.8777	0.0656	-0.0121
43.1049	-5.8960	-0.0306	-0.0011
-86.6871	29.2507	0.0737	-0.0046
20.4034	4.7838	-0.0056	-0.0050
39.3456	-7.6475	-0.0408	0.0007
164.0249	-19.0798	-0.1093	0.0132
-5.1020	3.1775	-0.0487	-0.0082

REFERENCES

Allan, L.G., and Jenkins, H.M. (1980) The Judgment of Contingencies and the Nature of the Response Alternatives. Canadian Journal of Psychology, 34, 1-11.

Beyth-Marom, R. (1982) Perception of Correlation Re-examined. Memory and Cognition, 10, 511-519.

Brehmer, B. (1974) Hypothesis about Relations Between Scaled Variables in the Learning of Probabilistic Inference Tasks. Organizational Behavior and Human Performance, 11, 1-27.

Crocker, J. (1981) Judgment of Covariation by Social Perceivers. Psychological Bulletin, 90, 272-292.

Einhorn, H., and Hogarth, R. (1986) Judging Probable Cause. Psychological Bulletin, 99, 3-19.

Lane, D., Anderson, C., and Kellam, K. (1985) Judging the Relatedness of Variables: The Psychophysics of Covariation Detection. Journal of Experimental Psychology: Human Perception and Performance, 11, 640-649.

Malmi. R. (1986) Intuitive Covariation Estimation. Memory and Cognition, 14, 501-508.

Pedhauser, E. J. (1982) Multiple Regression in Behavioral Research, 2nd Edition, Holt, Rinehart, and Winston. 232-247.

Wasserman, E.A., Chatlosh, D.L., and Neunaber, D.J.(1983) Perception of Causal Relations in Humans: Factors Affecting Judgments of Response-Outcome Contingenciesunder Free-Operant Procedures. Learning and Motivation, 14, 406-432.

(cont'd)

Well, A., Boyce, S., Morris, R., Shinjo, M., and Chumbley, J.(1988) Prediction and Judgment as Indicators of the Sensitivity to the Covariation of Continuous Variables. Memory and Cognition, 16, (3) 271-280.

Wright, J., and Murphy, G. (1984) The Utility of Theories in Intuitive Statistics: The Robustness of Theory-Based Judgments. Journal of Experimental Psychology: General, 113, 301-322.

Yates, J., and Curley, S., (1986) Contingency Judgment: Primacy Effects and Attentional Decrement. Acta Psychologia, 62, 293-302.