

Fall 1985

JUDGING CONTROL: WHEN CAN WE EXPECT EXPECTATIONS TO PREDICT JUDGMENTS? (CONTINGENCY, CONTRAST EFFECTS)

SUSAN E. NEWMAN

University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/dissertation>

Recommended Citation

NEWMAN, SUSAN E., "JUDGING CONTROL: WHEN CAN WE EXPECT EXPECTATIONS TO PREDICT JUDGMENTS? (CONTINGENCY, CONTRAST EFFECTS)" (1985). *Doctoral Dissertations*. 1461.
<https://scholars.unh.edu/dissertation/1461>

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.
2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again--beginning below the first row and continuing on until complete.
4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

**University
Microfilms
International**

300 N. Zeeb Road
Ann Arbor, MI 48106



8527887

Newman, Susan E.

JUDGING CONTROL: WHEN CAN WE EXPECT EXPECTATIONS TO
PREDICT JUDGMENTS?

University of New Hampshire

PH.D. 1985

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106



JUDGING CONTROL:
WHEN CAN WE EXPECT EXPECTATIONS TO PREDICT JUDGMENTS?

BY

SUSAN E. NEWMAN
B.A., California State University, Fresno
M.A., University of New Hampshire

A DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
The Requirements for the Degree of

Doctor of Philosophy

in

Psychology

September, 1985

This dissertation has been examined and approved.

Victor A. Bennis

Dissertation director, Victor A. Bennis, Associate Professor of Psychology

Grant L. Cioffi

Grant L. Cioffi, Assistant Professor of Education

Ellen S. Cohn

Ellen S. Cohn, Associate Professor of Psychology

Donald M. Murray

Donald M. Murray, Professor of English

Daniel Williams

Daniel Williams, Associate Professor of Psychology

July 29, 1985
Date

ACKNOWLEDGMENTS

Writing a dissertation is a lonely and often neurotic venture. I have been fortunate to have had the support of many people in all phases of this dissertation and, indeed, in my entire education. Space prohibits the public acknowledgment of every person; I hope that each knows that his or her presence has not gone unnoted.

I would like to thank Victor Benassi, my chairperson, for his seemingly undying (and at times relentless) encouragement as well as his keen insights into this material. Ellen Cohn has always been extraordinarily supportive and challenging, both as a committee member and as a friend and professor since my arrival at the University of New Hampshire. Grant Cioffi, Don Murray, and Dan Williams, the remaining members of my committee have also significantly contributed to the completion of this project; I am grateful to them all.

I would like to thank the University of New Hampshire for the summer fellowship and the dissertation-year fellowship, without which I could not have completed this dissertation at this time.

My gratitude is also extended to those who helped conduct these experiments: Debbie Donovan, Michael Emery, Jessie LaCross, and Denise McKenna.

Other people, perhaps without knowing it, also made this process a less trying one. To mention only a few, they include Ernie Landry of the Federal Aviation Agency for providing pleasant and flexible working

hours, Frank White for allowing rent payments at irregular intervals, and Liz Cook for her secretarial skills, grammatical guidance, and wonderful wit.

Many friends, across the continent and beyond, have shared in this process, each in his or her own way. Some, Julia Domino, Pam O'Neill, and Barry Quick, have witnessed the ups and downs of this venture since its inception. I am thankful that they are at the finish line to rejoice with me. Others have joined me along the way: Jim Demosthenes, Berrie Donovan, Joe Gillooly, Ethel Giardi, Cindy Grumbling, Paul Horovath, Beth Howley, Marcy Kraus, Denise McKenna, and Michele Richard. Some have delivered meals to my door when I was buried beneath my papers; some have brought me champagne when I was in the midst of dissertation blues; some have understood; some have now gone different directions. I thank them, especially Berrie, when they were there.

One person, my very significant other, W. Towny Manfull, has never seen me without the shadow of my dissertation. I thank him for his patience and confidence during this time and, especially, for his company and warmth in which I often took refuge.

Finally, I would like to thank my family--my mother, my father, my brothers, and my grandparents--who have provided a veritable wealth of financial, emotional, and spiritual support without which I would not have survived intact. Because of them, my life has been richer.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	iii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
ABSTRACT.....	viii
SECTION	PAGE
INTRODUCTION.....	1
I. OVERALL LITERATURE REVIEW.....	3
Definition of Concepts.....	3
Studies Examining Predictability Judgments of Continuous Data.....	5
Studies Examining Predictability and Control Judgments of Binary Data.....	7
Summary.....	31
II. REVIEW OF LITERATURE SPECIFIC TO PRESENT STUDIES.....	33
Factors that Influence Contingency Judgments.....	33
The Alloy and Tabachnik Model for Understanding Contingency Judgments.....	50
III. THE PRESENT STUDIES.....	58
Experiment 1.....	60
Experiment 2.....	74
IV. DISCUSSION.....	84
The Present Studies.....	84
Directions for Future Research.....	87
LIST OF REFERENCES.....	92
APPENDICES.....	101
APPENDIX A-STATISTICAL COMPUTATIONS FOR CONTINGENCY JUDGMENTS.....	102
APPENDIX B-QUESTIONNAIRE.....	105

LIST OF TABLES

1. Descriptive Statistics for Judged Control in Experiment 1.....	68
2. Mean Percentage of Expected and Obtained Control in Experiment 1.....	69
3. Descriptive Statistics for Judged Control in Experiment 2.....	79
4. Mean Percentage of Expected and Obtained Control in Experiment 2.....	80

LIST OF FIGURES

1. Mean Judgments of Control in Set 1 and Set 2 of Experiment 1.....	70
2. Mean Judgments of Control in Set 1 and Set 2 of Experiment 2.....	81

ABSTRACT
JUDGING CONTROL:
WHEN CAN WE EXPECT EXPECTATIONS TO PREDICT JUDGMENTS?
by

Susan E. Newman
University of New Hampshire, September, 1985

Two experiments were conducted to test the predictive validity of a model put forth by Alloy and Tabachnik (1984) to account for judgments of contingency. The authors contend that such judgments can be accounted for by determining the interaction of information contributed by "one's expectations" and by the objective "situation." It was argued here that those two components, independently or interactively, are not always sufficient to predict judgments; the same objective data may be judged differently in different contexts and sometimes contrarily to one's expectations. That is, contrast effects may occur. This argument was supported using two different paradigms. In the first experiment, 66 female Introductory Psychology students were evenly divided into three groups which each received two sets of 40 trials. In the first set of trials, the response-outcome contingency was high, low, or absent. In the second set, the contingency between response and outcome was absent. The results revealed that subjects who initially received high-contingent response-outcome trials judged Set 2 non-contingent response-outcome trials as having significantly less contingency than subjects who initially received non-contingent response-outcome trials. The second experiment conceptually replicated and extended those findings. Thirty female Introductory and Social Psychology students were evenly divided into two groups who initially received exposure to 40 trials in which

response-outcome contingency was high or absent followed by 40 more trials in which response-outcome contingency was moderate. The Set 2 judgments provided by each group were significantly different in that subjects with initial high-contingent experience judged later medium-contingent response-outcome trials as having significantly less response-outcome contingency than did those participants who had initial exposure to non-contingent response-outcome trials. Alloy and Tabachnik's model apparently cannot account for these findings. A comprehensive review of the literature is included as well as implications for that model and directions for future research.

INTRODUCTION

Perceived control has occupied an important place in the theory and practice of psychologists for decades. Only recently, however, has the accuracy of these perceptions of control emerged as a major focus of research. One's ability to judge accurately the amount of control he/she has over outcomes is fundamentally important in learning adaptive behavior. Thus, this research, within the broader context of contingency judgments, has attracted the attention of psychologists from many diverse backgrounds (e.g., clinical, cognitive, developmental, learning, organizational, personality, and social).

The conclusions of these studies have been equivocal; however, in general, they cast doubt on people's ability to make accurate judgments about contingencies. They reveal that under many circumstances, individuals are not accurate in judging how much control their response exerts over an outcome or in judging the direction and strength of a relationship between two given events.

Many factors that may affect these judgments have been identified: frequency and/or valence of outcome, inappropriate organization, differential use and/or poor recall of information, motivational biases, experimental demands, to name just a few. In what appears to be the first attempt to incorporate these factors into a model, Alloy and Tabachnik (1984) have proposed that the combined influence of one's prior expectations and objective situational information can account for the findings in this area.

The purpose of this dissertation was to test the predictive validity of the Alloy and Tabachnik (1984) model. It is contended that information contributed by "one's expectations" in conjunction with the objective data is not always sufficient to predict contingency judgments of those data. It is argued that the same objective data will be judged differently in different contexts and sometimes contrarily to one's expectations. Specifically, when the same situational information is judged in different contexts, the judgments may differ; they may be displaced away from the context and from what one's expectations would be, based on that context. That is, contrast effects are predicted to occur.

In the next section, an overall review of the literature is presented which discusses definitions of contingency, prediction, and control, the normative (statistical) calculations used to determine prediction and control, and research on judgments of continuous and binary data. Those readers familiar with this literature (and/or prone to narcolepsy) may want to move on to Section II which provides a more specific presentation of the relevant findings in this area as well as a review of the Alloy and Tabachnik (1984) theoretical framework. Finally, in the third and fourth sections, the present studies are presented and discussed.

I. OVERALL LITERATURE REVIEW

Definition of Concepts

Contingency is typically used as a broad term to describe the degree to which two events are related (see Alloy & Abramson, 1979, 1982; Abramson, Alloy, & Kossman, in press; Jenkins & Ward, 1965). The relationship between these events may range from being absent (zero contingency) to completely dependent (perfect contingency). The direction of the contingency may be positive or negative. That is, outcome occurrence may be correlated with response occurrence (positive) or response absence (negative).

When the two events in question are stimuli, their relationship is discussed in terms of predictability (Alloy & Abramson, 1979; Alloy & Abramson, 1982; Abramson, Alloy, & Kossman, in press; Seligman, 1975) or correlation. That is, one event may not predict the other event at all (zero contingency, zero predictability), may predict it to some extent (moderate contingency, moderate predictability) or may predict it completely (perfect contingency, perfect predictability). Similarly, the strength of the correlation may range from zero to perfect correlation; the direction, of course, may be positive or negative.

When the two events of interest are one's response and an outcome, the relationship is described as one of controllability (Alloy & Abramson, 1979; Alloy & Abramson, 1980; Abramson, Alloy, & Kossman, in press; Seligman, 1975). For example, one's response may exert no control over a given outcome (no contingency), some control (some contingency), or complete control (complete contingency). When control

over an outcome is present, the probability of outcome occurrence given one response is different from the probability of the same outcome occurrence given another response (Jenkins & Ward, 1965; Seligman, 1975).

The degree of contingency inherent in a given relationship may be specified by any one of a variety of statistical formulae, some of which are considered more valid than others (see Allan, 1980; Crocker, 1981; Jenkins & Ward, 1965; Wasserman, Chatlosh, & Neunaber, 1983).

For relationships between continuous variables, the determination of the amount of correlation present is made on the basis of Pearson r .

For relationships between binary variables, the degree of predictability or controllability can be specified by any of the following coefficients: chi square (χ^2), phi (ϕ), "delta d" (Δd), and "delta p" (Δp). These techniques differ in their applicability and computational complexity. Delta d, which reflects the difference between the conditional probabilities of an outcome given each response, has been accepted by most researchers in this area as the statistic of choice (e.g., Alloy & Abramson, 1979; Allan & Jenkins, 1983; Jenkins & Ward, 1965; Ward & Jenkins, 1965) and will be employed in the present studies. These statistical computations, in addition to other commonly employed approaches, are discussed in more detail in Appendix A.

Studies Examining Predictability Judgments
of Continuous Data

There have been only a few studies which have examined judgments of correlation of continuous events. Using the Pearson r as the statistical criterion, the results of these studies indicate that subjects tend to produce judgments which are closely related to but not isomorphic with the objective degree of contingency in the data.

In two such studies, Erlick (1966) and Erlick and Mills (1967), subjects were shown two semi-circular dials each with one of five possible values arranged so that 21 different r s ranging from -1.00 to +1.00 were reflected. Estimates of direction and strength of the relationships were linearly related to statistical estimates, but were more accurate in positive relationships. Negative relationships tended to be underestimated.

Beach and Scopp (1966), in a study in which they asked people to estimate only the direction of relationships, presented participants with pairs of numbers ranging from 1 to 10 with actual correlations of $\pm.14$, $\pm.15$, $\pm.30$, $\pm.50$, and $\pm.85$. Similar to the above studies, subjects were generally accurate in judging the direction of the correlation but, in contrast to the above studies, judgment accuracy tended to increase as the magnitude of the correlation increased in either direction.

In a study by Jennings, Amabile, and Ross (1980), (of which only relevant parts are discussed here), participants were asked to estimate the direction and strength of the relationship between unrelated and

positively related pairs of variables in three different sets of data which were qualitatively different and which required different information processing demands. (One set consisted of number pairs; a second set consisted of pictures of men of varying heights with sticks of varying heights; the third set, presented on a tape, was made up of pairs of alphabetical letters and musical notes of varying lengths.) The results revealed that, regardless of the type of data, the task appeared to be difficult. Judgments tended to parallel objective contingencies but were also conservative; only when correlation coefficients reached $+0.85$, did the participants begin to consistently recognize the presence of positive correlations.

Although one must be cautious in discussing the results of only four studies, two points bear emphasis. First, in contrast to binary data contingency judgments, judgments of the contingency of pairs of continuous variables in these studies consistently paralleled objective contingencies except for a tendency to underestimate negative and/or low correlations. Researchers (e.g., Crocker, 1981; Alloy & Tabachnik, 1984) have suggested that these results occur, not because of the type of data, but because the stimuli employed are devoid of any familiarity to the subjects and, therefore, subjects' expectations do not bias the judgments. This contention has some support although more research is certainly needed. For example, at a minimum, more studies are needed which use nonbinary data, particularly data about which participants may hold some expectations.

Second, Crocker (1981) has argued that using the Pearson r as the statistical criterion may be an "unreasonably stringent standard" (p.279). She contends that the concept of r is difficult to understand and that perhaps researchers should accept estimates that are highly correlated with r as normatively accurate. Thus, the estimates found in the aforementioned studies may not be conservative but, rather, may be extremely accurate judgments. The issue of what criteria to accept as reflecting accuracy remains a controversy. The reader is referred to Cohen (1979) for a discussion of normative or statistical criteria models versus intuitive criteria models.

Studies Examining Predictability and Control Judgments
of Binary Data

Most of the contingency judgment work has used two events which are binary (dichotomous) as the focus of judgments. Many events outside of the laboratory are of this nature and, in fact, people often perceive events as binary even though they are actually continuous (Crocker, 1981). The results of these studies are mixed; they range from a portrayal of people as very poor to very accurate contingency judges.

The present review is divided into two parts. In the first part, most of the studies discussed are those in which the binary data are two stimuli; in these experiments, subjects judge the degree to which one event occurrence is correlated with or predicts the occurrence of another event. A series of studies in which the data are discrete but not binary are also presented here. The second section reviews studies in which the binary data are one's response and the occurrence of an

outcome; in these investigations, participants judge the amount of control their response has over an outcome. In either case, the data is typically represented in a 2 X 2 contingency table.

Studies of Predictability

Hake and Hyman (1953), in one of the earliest studies in this area, presented subjects with 240 trials composed of either a random or a probabilistic patterned series of binary events (a horizontal or a vertical array of lights). Unlike subsequent studies in which subjects were asked to judge the presence and strength of any contingency between events, these participants were to predict which of the events would appear on the next trial. The results indicated that the participants did not perceive the probability rules upon which the series of trials was based. Instead, they saw order where none actually existed. For example, people who received the random series tended to perceive that sequence as if it were nonrandom; they believed that subsequent events could be predicted on the basis of previous events. Thus, subjects were not accurate in their predictions and, thus, judgments of contingency.

Similarly, Smedslund (1963), in one of the first studies of actual judgments of contingency, reported that student nurses' judgments of the relationship between two unrelated events bore no resemblance to the objective noncontingency between those two events. Specifically, in two experiments, participants were presented 100 cards, each with letters representing the presence or absence of a symptom and the presence or absence of a disease. The task was to determine whether there was a relationship between the specified symptom and disease.

Even when subjects could arrange the cards as they pleased, review them as often as they wished, and take notes about them (as in Experiment 2), they still stated that there was a relationship between the symptom and the disease when there actually was not one. The nurses were also asked to describe the strategies they used to reach their conclusions. To the extent that a strategy existed, it appeared to be solely reliant on positive confirming cases. Smedslund concluded that "normal adults with no training in statistics do not have a cognitive structure isomorphic with the concept of correlation" (p. 172).

Another group of studies often cited in this literature further corroborates the pessimistic conclusions of people's contingency judgment abilities although the events used in these studies have been discrete not binary. This work has been referred to as "illusory correlation" research and stems primarily from the investigations of the Chapmans (Chapman, 1967; Chapman & Chapman, 1967, 1969) although other researchers have replicated and extended their work (e.g., Golding & Rorer, 1972; Hamilton, 1979; Kurtz & Garfield, 1978; Lueger & Petzel, 1979; Starr & Katkin, 1969).

"Illusory correlation" is defined as:

the report by an observer of a correlation between two classes of events which in reality (a) are not correlated, or (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction than that which is reported (Chapman & Chapman, 1967, p. 194).

Mainly, this concept has been used to explain the persistence of clinicians in observing relationships which have not been empirically

substantiated between certain projective test responses (e.g., from the Draw A Person (DAP) test and the Rorschach test) and clinical symptoms as well as the remarkably similar persistence of laboratory subjects in observing the same relationships despite their empirical absence. In fact, even when, for example, there were negative relationships and/or when a monetary incentive to make more accurate judgments was introduced in the laboratory studies, the inaccuracies were only somewhat mitigated (Chapman & Chapman, 1967).

Chapman and Chapman (1967), for example, conducted six studies of illusory correlations using the DAP test, a commonly employed test which has never been empirically validated. After establishing which test responses (e.g., atypical eyes, large head, etc.) clinicians associated with which symptoms (e.g., manliness, suspiciousness), the authors randomly paired 45 DAP drawings with two of each of six symptoms. The subjects examined each pair for 30 seconds and were then asked to report which test responses had been most often associated with each symptom. The results indicated that for all but one of the test responses, naive subjects "rediscovered" the same associations as the clinicians had reported even though there were no objective relationships between the responses and symptoms.

In subsequent experiments, subjects were provided repeated experience with the stimulus materials (Experiment 2), exposed to responses and symptoms which were actually negatively correlated (Experiment 4), offered \$20.00 for the most accurate judgments (Experiment 5), and allowed to take notes and rearrange the DAP cards in addition to being offered \$20.00 for the most accurate judgments

(Experiment 6). Although illusory correlations were reduced in Experiments 4 and 6, they were not eliminated.

Starr and Katkin (1969) conducted a similar study, but one in which less information processing was required. In a design in which five sentences from the Rotter Incomplete Sentences Blank (ISB) were paired equally often with each of five symptoms, participants reported the same illusory correlations as those stated by clinicians. Also, Lueger and Petzel (1979) found that if you increase the amount of information to be processed (e.g., by increasing the number of cards), illusory correlation reports increase. Kurtz and Garfield (1978) attempted to reduce illusory correlations by pretraining subjects ("warning" them to be on guard against such biases and showing examples of illusory correlations); their efforts were not successful in reducing the inaccuracy in subjects' perceptions.

In sum, the studies of "illusory correlation" indicate that people are consistently inaccurate in their contingency judgments, even under seemingly optimal circumstances. Other studies, however, have indicated that people are not inaccurate in judging the contingency between two events; rather, they are quite accurate.

Inhelder and Piaget (1958), in one of the earliest works in this area, found that by about age 15, most children make accurate judgments about the presence of contingency between two binary events. At this point, individuals are also said to begin using appropriate judgment strategies. (For example, rather than relying on cell a frequencies or a comparison of cell c and cell b frequencies, people begin to compare diagonals.)

Inhelder and Piaget (1958) are not explicit about their procedure, but, in general, subjects were shown subsets from a total of 40 cards, each with a face drawn on it. There were two types of eye color (brown, blue) and hair color (brown, blonde). Subjects were to determine, for example, if there was a relationship between eye color and hair color, the strength of the relationship (e.g., "maximum" or "weak"), and which cards needed to be removed in order to change the strength of the relationship (e.g., to make it stronger or weaker). Apparently, the participants were able to arrange the cards in stacks, if they so desired. As stated above, the results of their work portrays humans who are older than 15 as accurate judges of contingency.

However, as Jenkins and Ward (1965) pointed out, the above task was more facile than those in other studies. The cell frequencies were small (e.g., 3,6,6,4; 4,2,2,4; etc.) and the instances could be simultaneously viewed and arranged by the participants into relevant cells. In contrast, for example, in Smedslund's (1963) second experiment in which judgments were not generally accurate, subjects also viewed cards and were able to arrange them as they pleased; however, there were 100 cards which resulted in much larger cell frequencies (e.g., 37, 33, 17, 13) which were virtually impossible to simultaneously view. Further, because at least one of the marginal frequencies were equal (in Inhelder & Piaget's study), "Delta d," a less complex judgment strategy could be employed rather than "Delta p." (Although "Delta d" was considered appropriate in Smedslund's study, perhaps it was not enough to counteract the large cell frequencies.)

A study which sought to further delineate the conditions under which judgment accuracy would be more likely to occur was conducted by Ward and Jenkins (1965). These researchers examined the effect of the mode of presentation of information. Participants were asked to judge the amount of contingency between cloud seeding (present or absent) and rainfall (present or absent). Three groups of people received information about the relationship in one of the following ways: only serially (trial-by-trial), only an organized summary table, or serially followed by an organized summary table. There were eight problems, all of which reflected a positive contingency between cloud seeding and rainfall. (A second purpose of this study was to use variables which had some inherent element of chance in order to counteract a possible failure by subjects to view favorable events as occurring by chance; rainfall certainly can occur by "chance.")

The results indicated that judgments were most accurate when information was presented only in a summary table. In addition, it is interesting to point out that the majority of subjects in the summary table group (75%) used an appropriate judgment rule (e.g., Delta d) as compared to 17% of the participants in the trial-by-trial group. The authors propose that the judgment accuracy of the summary table group was probably due to the organization of the information rather than the simultaneous availability of that information.

Three other studies also lend support to Ward and Jenkin's conclusions. Shaklee and Mimms (1982) and Arkes and Harkness (1983) found that when subjects are able to fill in the cells of a 2 X 2 matrix as the trials proceed, more complex rules are used and, where

assessed as in the former experiment, judgment accuracy is higher. These researchers attributed their findings to reduced memory requirements. However, Wasserman and Shaklee (1984), who followed a suggestion by Ward and Jenkins (1965) that information should be presented in a time-line format to eliminate memory demands, found that summary tables still produced the most accurate judgments. (In a time-line format, sampling intervals are graphically and linearly arranged. This maintains the sequential nature of the events found in trial-by-trial formats, and makes the memory demands more comparable to those of summary tables.) Further, memory demands required by a trial-by-trial format.) Further, people who transferred time-line information into summary tables provided judgments that were still not as accurate as those given by subjects with only summary tables. Thus, Ward and Jenkins may be correct in their contention that the memory demands which are freed by the simultaneous presentation of information cannot entirely account for the high level of accuracy exhibited by subjects in the summary table only conditions.

In relating the Ward and Jenkins (1965) findings to those of Inhelder and Piaget (1958), one might argue that the simultaneous availability of information, when in conjunction with low cell frequencies and the opportunity for subjects to impose organization on the data, provides the same circumstances as a summary table and, thus, enables accurate judgments.

In another study, Seggie (1975), designed primarily to examine the presence of the Piagetian concept of correlation, more light is cast on factors which appear to play a role in judgment accuracy. In this

study a method was used which combined aspects of the previously described studies by Smedslund, Inhelder and Piaget, and Ward and Jenkins.

In the first phase, adults were given a pack of 100 cards, each marked with two letters representing, respectively, whether or not a soldier had been sent to the hospital for his disease and whether or not he recovered. Six groups of subjects each received a different set of information (pack of cards), reflecting a relationship between the variables that was negative, positive, or absent. The participants had 10 minutes to examine the cards as well a pencil and paper with which to take notes. In the second phase of the study, all of the participants viewed information from the previously described different sets, only the information was now presented in the form of a summary table (taken from Ward & Jenkins). In both phases, the primary task goal was to decide whether or not to send the sick man to the hospital.

The researchers found that, regardless of mode of presentation, when a contingency (positive or negative) was present between hospitalization and recovery, subjects' decisions were correct, thus implying that they accurately detected the presence and direction of the contingency. But, when there was no relationship between the two variables, decisions appeared to have been made at random, thus implying a lack of recognition and/or understanding of noncontingency.

To the extent that these decisions and judged contingencies are related, it may be instructive to compare the methods and results of some of the experiments reported so far. Seggie's design was much like the third group of the Ward and Jenkins (1965) study: subjects first

received information in a strict serial fashion (Ward & Jenkins) or a fairly serial (Seggie) presentation mode and then in a summary table. In a comparison of only positively contingent events (since all problems in Ward & Jenkins were of this nature), judgments were not accurate in the Jenkins and Ward study, but were in the Seggie study. Jenkins and Ward argued that "to be effective, ... the organization must be provided when subjects first begin to assimilate the information. It has little influence on judgments when it is provided after they have followed a trial by trial presentation" (p. 240). Yet, this information was effective in Seggie's experiment. Because the method sections of either study are not detailed, it is unclear what may have led to this apparent discrepancy. One difference is that Seggie's subjects could use a pencil and paper; however, it is not reported whether and to what extent they were used.

Seggie's experimental procedure, unlike that of Inhelder and Piaget (1958), did not employ small cell frequencies nor could the information be viewed simultaneously. And yet, like Inhelder and Piaget's results, subjects' judgments (of contingent relationships) were accurate. The less complex "Delta d" rule, though, could be successfully employed in both cases.

In regard to Smedslund's second experiment, the procedure is similar to Phase 1 of Seggie's study: a deck of 100 cards were used, cell frequencies were identical in some cases, simultaneous viewing was not possible, note-taking was possible, and the "Delta d" rule could be appropriately employed. But, in contrast to the judgments of Smedslund's subjects, Seggie's subjects gave accurate judgments when

contingency was present. When it is absent, as in Smedslund's experiment, judgments were apparently less likely to be accurate.

In a comparison of only noncontingency conditions (since Experiment 2 of Smedslund's study only used noncontingent situation), in both cases, subjects provided inaccurate estimates.

It has been suggested that noncontingency is difficult for subjects to recognize (e.g., Allan, 1980; Allan & Jenkins, 1980; Alloy & Abramson, 1979). Most studies addressing this issue have been conducted using a response-outcome control format rather than a stimulus-stimulus predictability paradigm. An exception is a study conducted by Peterson (1980).

Peterson suggested that subjects may fail to recognize noncontingency because they do not expect to see it. Because of the "special characteristics of the psychology experiment" (Peterson, 1980, p.728), participants do not entertain hypotheses of randomness. Using a task similar to that first employed by Hake and Hyman (1953) to assess predictability judgments, he manipulated expectations (i.e., the availability of randomness as an hypothesis) in two ways. Instructions were included which explicitly stated that randomness was a viable possibility and a prior comparison sequence which was patterned was added. These conditions were crossed with control conditions which involved no such instructions and a prior comparison sequence which was random.

The results partially supported Peterson's contention. If the experimenter-coded versions of the subjects' verbal descriptions of the pattern of the sequence are used as the dependent measure, the manipulations had an effect. When both experimental instructions and prior patterned sequence occurred, subjects viewed the sequence as random. When participants viewed the random sequence without either of the experimental manipulations, judgments were nearly identical to the mean score of the subjects observing the patterned set (2.88 vs. 2.85 on a scale of 1-3). But if probability estimates are used as the dependent variables, only differences between the test sequences are found, indicating that the random sequence was (accurately) perceived as more random than the patterned sequence.

Two points are worth noting here. First, in contrast to Hake and Hyman's (1953) study, subjects in Peterson's study did, overall, differentiate between random and patterned sequences. It is unclear why this difference occurred. Second, Peterson's conclusions would be more convincing if other researchers had not also stated explicitly that no control was a possibility. Jenkins and Ward (1965), for example, told their subjects that among the three states of affairs that might be found in the problems was that "response choices do not affect the outcome" (p. 6). They add "the possibility that the correct judgment for a given problem might be one of zero control was stated explicitly" (p. 6). As discussed in the next section, participants in Jenkins and Ward's study made judgments that were completely unrelated to actual contingency.

Studies of Control

The second group of studies which have investigated contingency judgments of dichotomous variables are those in which the response-outcome design is employed. Following the format of Alloy and Tabachnik (1984), these studies may be divided into those which indirectly and those which directly assess judgments. Again, the results of these investigations are mixed, although the direct assessment method has yielded somewhat more encouraging findings about people's contingency judgment abilities (Alloy & Tabachnik, 1984). Since the present series of studies deals with people's direct judgments of the control their response exerts over an outcome, attention will be primarily devoted to those studies.

Indirect Assessments of Control Judgments. This research has focused on people's behavior in situations in which outcomes are typically not contingent on outcomes or are only contingent on one aspect of the situation. Studies fall into two categories: those of the learning tradition which have primarily studied "superstitious" behavior and those of the social psychology orientation which have investigated, among other areas, attribution of causality, "illusions of control," and "learned helplessness." The reader should bear in mind that Jenkins and Ward (1965) stated that "although it is not clear exactly how to relate performance under noncontingent reinforcement to a judgment of contingency made after the performance, the experiments are suggestive" (p. 4).

Research on superstitious behavior stems from the seminal work by Skinner (1948) who demonstrated that simply delivering food to a hungry animal is sufficient to condition certain behaviors. For example,

pigeons who had access to food every 15 sec independent of their behavior exhibited stereotypic behavior such as pecking and turning in circles that eventually became temporally correlated with food delivery. Although there has been some reinterpretation of the original work (see Staddon & Simmelhag, 1971, for example), the finding remains a robust one. Studies of superstitious behavior in humans have contributed to a further understanding of contingency judgments.

Bruner and Revusky (1961), for example, used an operant conditioning paradigm to investigate subjects' behavior and perceptions of what was the behavior necessary to be reinforced. Participants were placed before four telegraph keys which they were to press in order to attempt to collect reinforcement. Although reinforcement could only be obtained by pressing the third key and then allowing a certain amount of time to pass before pressing it again, all subjects perceived and behaved as if reinforcement were dependent on a pattern of responses. No subject perceived that the passage of time was related to reward.

The amount of reinforcement and its temporal relationship to the subjects' behavior also play roles in establishing superstitious behavior. Wright (1962) found that superstitious response preferences and patterns are more likely to develop with high levels of noncontingent reward. Catania and Cutts (1963) found that superstitious behavior is more likely to occur when that behavior is closely related temporally to the noncontingent reward. Thus, it appears that people are more likely to perceive that their behavior controls a given outcome (i.e., develop superstitious behavior) if reinforcement occurs at a high frequency and in close temporal

contiguity to the behavior even though there may be no causal relationship at all.

Social psychologists have also observed that people often do not respond to noncontingent response-outcome situations as if they had no control; rather, they may perceive their response as having control over a given outcome (e.g., Kelley, 1967; Wortman, 1975). Conversely, people may also respond to response-outcome controllable situations as if their response had no control over the outcome (e.g., Abramson, Garber, & Seligman, 1980; Abramson, Seligman, & Teasdale, 1978).

Langer (1975; Langer & Roth, 1975) has investigated people's perceptions of control in chance situations. She introduced factors from skill situations (e.g., competition, choice, familiarity, involvement) into obviously chance-determined situations and found that subjects exhibited "illusions of control" (Langer, 1975). That is, they had "an expectancy of a personal success probability inappropriately higher than the objective probability would warrant" (p.311). However, it is unclear whether these findings indicate that subjects actually misperceived the degree of contingency in these situations or whether they simply reflect the expected stability of causes involved in the initial situation (Abramson & Alloy, 1980; Alloy & Abramson, 1980; Alloy & Seligman, 1979; Weiner, 1974).

Other studies have documented that people may act as if they had no control over an outcome when they actually do. Most well-known is the work of Seligman and his colleagues (Maier & Seligman, 1976; Miller & Seligman, 1976; Seligman, 1975) who have proposed that when subjects (both animal and human) are exposed to situations in which outcomes are

independent of their responses, they may learn to expect that outcomes cannot be controlled. In other words, they learn that they are helpless and exhibit the concomitant deficits in cognition, motivation, and emotion. These deficits may persist despite the fact that in other situations, their responses and outcomes may be related.

The prototypical study of human learned helplessness is that conducted by Hiroto (1974). Using a triadic design, subjects received initial exposure to one of the following conditions: "escapable" in which termination of noise was contingent on button pressing; "inescapable" in which noise termination was not related to button pressing but instead was yoked to the "escapable" group; and a control group which received "no noise." All groups were then exposed to noise which could be escaped/avoided in a human shuttle box. In comparison to the "escapable" and "control" groups, the group which initially received inescapable noise did not learn to effectively escape/avoid the later noise. These results were said to support the contention that deficits in motivation and cognition. Other studies (see Abramson, Seligman, & Teasdale, 1978; Seligman, 1975; Seligman, Klein, & Miller, 1976) have shown that "depression" may also occur as a result of initial exposure to uncontrollable outcomes.

Hiroto's study and others in the learned helplessness tradition do not actually support the contention that subjects have a cognitive representation of contingency (see Alloy & Abramson, 1979) nor that, if they do possess such a representation, that they had perceived no contingency in the initial sequence of the experiment (see, e.g., Tennen, Drum, Gillen, & Stanton, 1982; Oakes & Curtis, 1982).

Researchers have suggested that the observed performance decrements may reflect only motivational deficits and not cognitive deficits (Alloy & Seligman, 1978; Costello, 1978).

Direct Assessments of Control Judgments. Probably the most widely cited series of experiments in this area was conducted by Jenkins and Ward (1965). In three experiments, each varying several factors, subjects were asked to judge the degree of contingency between outcomes which were contingent and noncontingent upon responses. In all conditions, judgments were correlated with the number of trials in which reward occurred and not at all related to the objective degree of contingency.

In Experiment 1, two levels of instructions ("to score" or "to control") were crossed with two levels of participation (active or passive). The procedure required subjects to choose between one of two response buttons after which one of two stimuli occurred. There were five problems of 60 trials each. In the first level of the first condition, the stimuli were "score" or "no score;" in the second level, the stimuli were a lighted circle or a lighted square. In addition, participants who received "control" instructions were required to indicate which of the outcomes he/she was attempting to bring forth. The two levels of participation were one in which subjects actually participated and another in which they simply observed another subject's participation. (The latter condition was somewhat akin to the stimulus-stimulus studies described earlier.) Subjects were asked to rate the amount of control they thought they had on a 0 to 100 scale.

The results revealed that instructions did have a strong effect on judgment although they are difficult to interpret since in some problems the "control" instructions led to higher control estimates while the reverse was also true. In general, the level of participation had no significant effect on judgments. The most robust finding was that the judgments of all subjects paralleled the frequency of successful trials.

Experiment 2, using basically the same method with only active subjects investigated the possibility that subjects may not understand that the word "control" means "the ability to alter what you get" rather than "getting what you want" (Jenkins & Ward, 1965, p. 9). The results confirmed that subjects may not have an accurate understanding of the concept of control, so the third experiment was designed to further address this issue.

In the final experiment, employing basically the same procedure with only active subjects and "control" instructions, participants were provided with pretraining along with correct estimates of the amount of control actually present. One group received examples in which the number of successful trials in a noncontingent series was correlated with the amount of control in the contingent series. A second group received examples in which the number of successes was not correlated with the amount of control. The third group was the same as the second group except that they were also given information about how judgments should be calculated. As predicted, the most accurate judgments were found in the third group, followed by the second; however, no significant correlation emerged between the actual amount of control

and judged control. (It is also interesting in light of experiments described earlier, that questionnaire results revealed, among several findings, that the influence of successful trials was apparent even when data were presented in 2 X 2 summary tables.)

In sum, Jenkins and Ward concluded that subjects' judgments of control were dependent on the number of successful outcomes rather than on the actual amount of control. This finding occurred regardless of the level of participation, type of instruction, presence of pretreatment, and, in contrast to studies described earlier, even when judgments were based on summary tabulations of trials. Further, like others (e.g., Smedslund, 1963; Chapman & Chapman, 1967, 1969; Nisbett & Ross, 1980), they concluded that the subjects in their experiments did not possess "an abstract appreciation of statistical contingency" (p. 15). Although the factors that may have contributed to the poor performance of Jenkins and Ward's subjects are discussed in detail in another section, it is instructive to presently review three that are identified by the authors.

First, Jenkins and Ward state that, unlike natural conditions, the response and outcome in their experiments had no temporal variations as a function of contingency. This is in contrast to the real world where responses and outcomes that are highly contingent generally occur with closer temporal contiguity than responses and outcomes that are not contingent. While this may be a difference between the conditions in their experiments and the natural environment, it is not a difference between their experiment and those described earlier in which subjects were accurate.

The second factor also contributes to making these experiments distinct from most natural conditions but is also common to all other experiments on contingency judgment. Unlike "perceptions" of contingency, a judgment implies that two inputs may occur together on some number of trials even though they are not contingently related. Subjects appeared to be making judgments based on the premise that two given events would not occur by chance (a zero-baseline). Although this may be a reasonable approach in many real world situations, it leads to inaccurate judgments in situations in which two given events may occur together simply by chance. [It was this error that led Ward & Jenkins (1965) to use cloud seeding and rainfall as stimuli in their experiment. They thought that subjects would expect some rainfall to occur by chance and, therefore, abandon a zero base-line. However, this did not appear to happen.]

The third factor involves the representation of the variables. Briefly, because there were two discrete symmetrical responses and two discrete symmetrical outcomes, there were actually three states involved for each. For example, there was Response 1, Response 2, and a resting state in between. In contrast, other experiments have used discrete responses of the "on-off" type, resulting in situations in which there is one response and a resting or neutral state. This factor has also been identified by other researchers (Allan & Jenkins, 1980, 1983; Alloy & Abramson, 1979; Beyth-Marom, 1982) and is discussed at length in a later section. It is one of several factors put forth to explain the difference between these results and those of Alloy and Abramson (1979) who, in a similar experiment (Experiment 1), found that subjects' contingency judgments were more accurate.

Alloy and Abramson (1979) conducted four studies of contingency judgments in which they were primarily interested in any differences that may occur between the contingency judgments of depressed people and those of nondepressed people. (Depression is defined on the basis of Beck Depression Inventory scores of 9 or above.) In all experiments, a modified version of the Jenkins and Ward (1965) study was employed. In general, subjects were asked to determine the amount of control they exerted (by pressing or not pressing a button) over an outcome (the presence or absence of a green light). Thus, there was only one response that could be made and only one outcome which could occur in contrast to the Jenkins and Ward (1965) study in which, as described above, there were two possible responses and outcomes. Other differences were that each subject received only one problem (compared to five) consisting of 40 trials (compared to 60) in which the intertrial interval (ITI) was an average of 14 seconds (compared to a 2 second ITI).

In Experiment 1, depressed and nondepressed college students were asked to judge the amount of control they had in one of the following conditions: 1) 25% control; 2) 50% control; or 3) 75% control. Because Jenkins and Ward (1965) had found such a robust frequency effect, Alloy and Abramson negatively correlated reinforcement frequency and the actual degree of contingency in order to counteract that effect. In this way, if the participants based their judgments on frequency of reinforcement, they would give inaccurate estimates.

Based on the learned helplessness model of depression, the following general predictions were put forth: "According to the strong prediction, depressed individuals should underestimate the degree of contingency between their responses and outcomes relative to the objective degree of contingency. According to the weak prediction, depressed individuals merely should judge that there is a smaller degree of contingency between their responses and outcomes than nondepressed individuals should" (Alloy & Abramson, 1979, p. 441).

The results did not support the predictions. Instead they revealed the depressed and nondepressed subjects did not differ in their judgments of control and that judgments of both groups were fairly accurate. Only at 25% control, where reinforcement frequency was highest, did both groups somewhat overestimate the amount of control they had. Although it is unclear which, if any, of the aforementioned factors may account for the differences in findings between this experiment and those of Jenkins and Ward (1965), much attention has been focused on the single versus multiple response-outcome difference.

Because previous research (e.g., Chapman & Chapman, 1967, 1969; Langer, 1975; Seggie, 1975; Smedslund, 1963) has indicated that noncontingency was difficult to detect, in Experiment 2, depressed and nondepressed subjects were asked to judge the amount of contingency present in noncontingent response-outcome situations. Here, also, the frequency of reinforcement was varied such that half the subjects received reinforcement on 25% of the trials while the other half received reinforcement on 75% of the trials. Based on the authors' interpretation of learned helplessness theory, it was predicted that

nondepressed subjects would overestimate the amount of control present relative to actual control (strong prediction) and/or relative to depressed students (weak prediction). The procedure was identical to that used in Experiment 1.

The results indicated that, although depressed students put forth fairly accurate judgments in both conditions, nondepressed subjects were only fairly accurate in the 25% reinforcement condition and they overestimated the amount of control in the 75% reinforcement condition. Thus, the predictions were partially supported: nondepressed people gave estimates of control which were greater than actual control and greater than depressed subjects' responses, but only in the 75% condition. So, it appears that frequency of reinforcement may become a biasing factor in high density situations for nondepressed but not depressed people.

Experiment 3 was conducted primarily to determine whether frequency of reinforcement or valence of outcome led to the observed overestimations of nondepressed people in the 75% reinforcement condition of Experiment 2. Depressed and nondepressed students provided judgments of contingency in a task identical to those of earlier experiments except that the green light onset was associated with either gain (win) or loss (lose) of money. Responses and outcomes were noncontingently related; reinforcement occurred on 50% of the trials.

The authors stated that the results supported the view that valence rather than frequency of outcome leads to overestimates of control in nondepressed people. That is, nondepressed people reported

more control than depressed people only in the "win" condition. However, it should be noted that since the frequency of reinforcement was held constant at 50%, one cannot conclude that valence is solely responsible for overestimates of control in nondepressed people.

In Experiment 4, the effect of valence on contingently related responses and outcomes was investigated. It was suggested that depressed people would underestimate the degree of control they wielded if contingent outcomes were positive. In this experiment, subjects received a "win" or a "loss" problem, in both of which there was a 50% degree of response-outcome contingency.

The results indicated that depressed people did not underestimate control in either condition; they were accurate. But nondepressed people underestimated control in the "lose" situation and appeared to overestimate control in the "win" situation.

In summing the results of the Alloy and Abramson (1979) studies, the conclusions most pertinent to the present review are that people do have a concept of contingency, but that it is affected by factors such as frequency and/or valence of reinforcement, the degree of contingency, and mood (depressed or nondepressed).

At the same time Alloy and Abramson (1979) were replicating and extending various aspects of Jenkins and Ward's (1965) work, Allan and Jenkins (1980) were also attempting to further understand the 1965 research. Their primary goal was to determine if having two active responses versus one active response differentially affected subjects' judgments of contingency. A secondary purpose was to explore the

effects of language; specifically, to see if "influence" versus "connection" differentially affected judgments.

The experiment consisted of three groups. Two groups had access to two active responses; each group received instructions to determine whether he/she influenced the outcome or his/her response was connected to the apparatus. The third group had access to only one active response in the "influence" instruction condition. Each group received ten problems of 100 trials each (five contingent and five noncontingent problems).

The results revealed that instructions had no effect on judgments and, more importantly, that the nature of the response alternatives played an important role in determining contingency estimates. When only one response was available, judgments more closely paralleled actual contingencies than when two active responses were available. The frequency of reinforcement affected both conditions, but was more prominent in the "two-active-response" condition. Thus, this factor may, at least partially, explain the differences in the findings of Jenkins and Ward (1965) and Alloy and Abramson (1979). Why the number of responses should play a role in making a judgment was speculated by Allan and Jenkins in the 1980 paper as well as a later report (Allan & Jenkins, 1983), which will be discussed in the next section.

Summary

The preceding review of literature is not exhaustive. It is intended to familiarize the reader with the most well-known studies in this area--those which have spawned other research and which have brought to light issues which have been explored more intensively in

later work. Other contingency judgment studies which were primarily conducted to further investigate factors which may play a role in making these judgments are discussed below, along with aspects of the previously reviewed studies which are pertinent to an understanding of these factors.

Two points should be clear by now. First, the results of studies in this area are mixed; in some cases, subjects appear to be fairly accurate while in other cases, subjects appear to be very inaccurate in their contingency judgments. Second, there are a multitude of factors which have been different among these experiments; some of these may help to account for the presently mixed bag of results. These factors are presented in the following section.

II. REVIEW OF LITERATURE SPECIFIC TO PRESENT STUDIES

Factors that Influence Contingency Judgments

There are no less than 25 different factors that may be identified as being influential in the process of making a contingency judgment. Most of these factors are irrelevant or only indirectly relevant to the computational task itself but, nonetheless, exert impact on the contingency judgment process. Tversky and Khaneman (1981) refer to the process that may be involved in some cases as "task framing," meaning that, like a picture, a frame may affect one's perception of the enclosed item, even though the item itself has not been changed.

The factors may be divided into two overlapping categories: 1) Experimental Factors and 2) Subject Factors. The former refers to features of the task, data, or setting. The latter category refers to characteristics of the subject which are brought into the situation by the subject, as an individual or as a member of the human species. Some of the factors that will be discussed have also been reviewed by other researchers (Arkes & Harkness, 1983; Alloy & Tabachnik, 1984; Beyth-Marom, 1982; Crocker, 1981).

The influence of one or more of these variables may occur at one or more of the steps that are proposed to occur when making a contingency judgment (see Crocker, 1981). There are five steps which are briefly outlined here. It should be noted, however, that these steps are based on a normative or statistical model. Whether or not subjects' cognitive processes actually mirror these steps remains to be

seen (see Alloy & Abramson, 1981; Alloy & Tabachnik, 1984; Cohen, 1979; Einhorn & Hogarth, 1978). At this point, the proposed steps serve satisfactorially as an organizational guideline.

First, the observer must decide what type of information to collect. For example, in assessing the relationship between two dichotomous variables, one must collect data to fill all four cells of the 2 X 2 matrix.

Second, the individual must collect a sample of relevant data from the population. This should be accomplished using a random procedure so that the sample will be representative of the population.

Actually, the observer in a contingency judgment study can typically bypass the aforementioned steps since most experimenters present the subject with the appropriate data. In real-life settings, though, both steps are subject to error and bias.

Third, the observed data must now be interpreted and classified as confirming or disconfirming information. When the data are distinctive (e.g., blue eyes vs. brown eyes), this step is fairly objective. However, most data do not afford such an objective process and the classification is, therefore, subject to error and bias.

Fourth, the perceiver must now estimate the contents of the classes, that is, the frequency of confirming and disconfirming cases. When this information has not been recorded, one must rely on memory and is, again, open to error and bias.

Steps three and four are also bypassed in some studies, such as those in which summary tables of information are presented.

Fifth, the observer must now employ some strategy to integrate the collected data and determine the degree of contingency present. As discussed above, there are a number of statistical techniques that may be used; there are also a number of techniques that are typically, but erroneously, used by subjects. This step is also susceptible to a number of factors which may lead to inaccurate conclusions.

Experimental Factors

One of the most discussed experimental factors is outcome frequency. For example, Jenkins and Ward (1965) found that contingency judgments were highly positively correlated with the frequency of the outcome. Alloy and Abramson (1979, Experiment 1), in contrast, reported that they did not find an outcome frequency effect (although a closer examination of their results reveals that that conclusion is debatable). Benassi, Knoth, and Mahler (in press) have reviewed the outcome frequency effect in response-outcome noncontingent cases and have concluded that, outcome frequency effects are not universal. It appears that the effect of this factor is dependent on other factors.

One of the factors which may interact with outcome frequency to affect judgments of contingency between binary variables may be the way those variables are represented (Allan & Jenkins, 1980; 1983; Alloy & Abramson, 1979; Beyth-Marom, 1982). As stated above, this factor probably plays a major role in explicating the contradictory results of Jenkins and Ward (1965) and Alloy and Abramson (1979).

Beyth-Marom (1982) refers to the representation of binary variables as asymmetrical (event-nonevent) or symmetrical (event-event). For example, in regard to response-outcome designs, the required response may involve, for example, pressing a lever or not pressing a lever (event-nonevent) or it may entail, for example, pressing the lever to the right or pressing the lever to the left (event-event). Similarly, the outcome may be the presence or absence of a stimulus (event-nonevent) or either of two changes in the stimulus, such as its movement up or down (event-event). In regard to stimulus-stimulus designs, for example, one stimulus may be a symptom that is present or absent (event-nonevent). Another stimulus, for example, may be a prognosis that is improved health or death (event-event). Since most research on this topic has been conducted using a response-outcome format, that work will be presented first.

The two levels of responses (R) and outcomes (O) may be factorially combined to yield the following four cases: 1R/1O, 2R/1O, 1R/2O, 2R/2O (Allan & Jenkins, 1980; 1983). Recent research has revealed that these cases are not judged equally accurately and that outcome frequency, in particular, seems to differentially interact with each case to produce varying degrees of accuracy.

Specifically, both Allan and Jenkins (1980) and (1983) found that 1R/1O cases yielded more accurate judgments than 2R/1O cases and subjects used more valid indexes (e.g., delta p or delta d) to make their estimates for the 1R/1O cases. Additionally, although outcome frequency led to higher estimates of contingency in both cases, the latter case was significantly more effected. Allan and Jenkins (1980)

found these results only in noncontingent cases while in 1983 (Experiment 1), they observed the aforementioned findings in both noncontingent and contingent cases. Furthermore, in 1983, they included 1R/20 and 2R/20 cases and found that subjects in the former condition had less accurate judgments than subjects in the 1R/10 and 2R/20 situations but were more accurate than those subjects in the 2R/10 case. Subjects in the 2R/20 case were approximately equal to subjects in the 1R/10 case.

It is unclear why this differential accuracy occurs. Allan and Jenkins (1983; see also, Allan & Jenkins, 1980) have suggested that it is the causal compatibility of the input-output representation that may be responsible. In the 2R/10 case, "the causal compatibility of event-event occurrences leads some subjects to judge influence as though they assume that in the absence of an event-like input, the event-like output would not have occurred; in other words, they appear to be victims of the logical fallacy known as the denial of the antecedent" (Allan & Jenkins, 1983, p. 404). The 2R/10 case is described by those authors as "standing apart from others" since it is especially susceptible to "upward distortion of judged influence based on the frequency of the event-like outcome" (p. 404). There is some supportive evidence.

Beyth-Marom (1982) has examined the effect of the representation of variables in stimulus-stimulus designs. She found that less valid rules were used when the contingent or noncontingent variables were asymmetrical (analogous to 1R/10). These results seem to contrast with the work of Allan and Jenkins (1983) which indicated that asymmetrical

(i.e., 1R/10) cases yielded judgments approximately the same as symmetrical (i.e., 2R/20) cases. This difference may stem from the use of response-outcome versus stimulus-stimulus designs. The finding is also interesting in light of other work which has revealed that multivalued correlational cases (which are analogous to symmetrical cases) are typically judged accurately.

Finally, the reader may have recognized that it may be difficult to distinguish between asymmetrical and symmetrical cases in both real-life and laboratory situations. For example, subjects in Jenkins and Ward's (1965) study had the choice of making one or the other of two responses (press one button or press the other) which was followed by one of two outcomes (a "score" light or a "no score" light or, in another condition, a lighted square or a lighted circle). Clearly, there are two responses and the square and circle are two outcomes, making for a 2R/20 case. The score-no score condition, however, has been considered both a 20 (Allan & Jenkins, 1980) and a 10 (Allan & Jenkins, 1983; Beyth-Marom, 1982) case. Even the example that Allan and Jenkins (1983) provide to illustrate a 20 case is ambiguous; they suggest that "a car that stops or starts" is a 20 case. One must ask what the prevailing state of such a car would be.

Still, this factor clearly plays a role in determining judgments and warrants further investigation. Especially important are studies which control for other variables that have differentiated and confounded a clean comparison of experiments in this area. They include the number of trials received, between- versus within- subjects designs, intertrial interval (ITI) and related to ITI, discrete- versus

continuous-trials versus free operant experiments.

The nature of the contingency also affects the accuracy of judgments. It has already been noted that noncontingency is generally difficult to detect (Allan, 1980; Allan & Jenkins, 1980; Alloy & Abramson, 1979; Chapman & Chapman, 1967; Langer, 1975; Peterson, 1980; Seggie, 1975; Smedslund, 1963). Its detection appears to be influenced by several factors (e.g., mood, instructions, variable representation). The sign of the covariation also makes a difference: positive contingencies are typically judged higher than negative contingencies of the same magnitude (Allan & Jenkins, 1983; Erlick, 1966; Erlick & Mills, 1967; Jennings et al., 1980; Wasserman & Shaklee, 1984). Causal compatibility has been put forth (Allan & Jenkins, 1983) to explicate this differential accuracy. However, in Wasserman and Shaklee's (1984) study, negative contingencies were not always judged less accurately leading them to propose that the mode of information presentation differentially affected the judgments of positive and negative contingencies. It may be that some modes of presentation are such that subjects' biases in identifying more positive contingencies than negative contingencies are not as prominent.

The mode in which the information is presented does appear to affect the accuracy of contingency judgments (see, Beyth-Marom, 1982; Seggie, 1975; Smedslund, 1963; Ward & Jenkins, 1965; Wasserman & Shaklee, 1984). Until recently, information was typically presented over a series of trials (e.g., Alloy & Abramson, 1979; Shaklee & Mims, 1982; Ward & Jenkins, 1965) or in a summary table (e.g., Seggie, 1975; Smedslund, 1963; Ward & Jenkins, 1965). It may be recalled, for

example, that Ward and Jenkins, who employed a serial presentation, a summary table presentation, and both types of presentations, found significant differences in judgment accuracy as well as differential use of judgment rules among those groups. The condition in which only the summary table was presented yielded the most accurate scores, prompting researchers to examine the roles of memory demands and of memory organization.

A series of studies (Wasserman & Shaklee, 1984) was recently conducted to investigate the roles that these factors may play in determining differences in judgment accuracy among methods of information presentation. They included a presentation format (previously suggested by Ward & Jenkins, 1965) in which information was given to subjects in a time-line. In this way, the memory demands required are the same as those in a summary table, but the organization of the material is different in the two formats. As in Ward and Jenkins' work, the Wasserman and Shaklee series of studies indicated that the latter factor rather than memory demands may be responsible for the overall increased accuracy consistently found when information is presented only in a summary table. This may mean that the fourth step of Crocker's normative model for making contingency judgments may be more complex than simply recalling and estimating the frequencies of confirming and disconfirming cases or that there is substantial compartmentalization of information.

Memory demands do, however, influence judgments. Several studies have recently confirmed that increased memory demands are associated with decreased judgment accuracy (see, e.g., Arkes & Harkness, 1983;

Shaklee & Mims, 1982). It appears that when memory demands are increased (e.g., by using a trial-by-trial procedure, including distracting stimuli, etc.) judgment accuracy is primarily affected in one of two ways; 1) Subjects revert to simpler judgment strategies and 2) Subjects become less accurate in their frequency estimates.

The simpler but less accurate approaches to making judgments of contingencies (see Appendix A for review) include the "cell-a" or the "a-versus-b" strategies or a shift to one of these during the assessment. Frequency estimates may be inaccurate because very low cell frequencies tend to be overestimated (Arkes & Harkness, 1983) and/or subjects tend to move from left to right and top to bottom such that cell a (which is probably attended to more, anyway) is most accurate and cell d is least accurate (Shaklee & Mims, 1982).

Other biases that have been found to affect the recall of events, especially familiar ones or those in real life situations are availability (Tversky & Kahneman, 1973), representativeness (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974), and confirmatory biases (Snyder & Swann, 1978; Wason & Johnson-Laird, 1972) (see also, Nisbett & Ross, 1980). Subjects, of course, carry expectations into the experimental situation which will also bias their recall of events as well as other processes involved in judgment.

It has been suggested that judgmental strategy and accuracy are dependent on the instructions that subjects receive (Allan & Jenkins, 1980; Beyth-Marom, 1982; Crocker, 1981, 1982; Peterson, 1980; Shaklee & Tucker, 1980). Aspects of instructions that may be particularly important are the definition of contingency (that may or may not be

provided), the question format that may (inadvertantly) lead the subjects to selectively attend to certain information, and the presence or absence of information that noncontingency is a possibility. There is some (albeit inconsistent) support for the contention that these factors play a role.

The effect of pretraining on judgmental accuracy has also been examined (Alloy & Abramson, 1982; Jenkins & Ward, 1965; Peterson, 1980; Shaklee & Tucker, 1980). The findings are inconclusive as to the role of this factor, but this may at least partially stem from the use of different kinds of pretraining and test-trials as well as the variety of procedures employed.

A factor which has been found to influence subjects' expectations of control is the introduction of aspects characteristic of skill into chance situations (Langer, 1975). When competition, choice, or practice are introduced to such situations, subjects' expectancies of control increase.

Finally, there are "characteristics of the data" (Crocker, 1981) that have been shown to influence judgment accuracy. Some of these have already been reviewed such as continuous versus binary data, stimulus-stimulus versus response-outcome data, single response-outcome versus multiple response-outcome data, positively versus negatively related sets of data, contingently versus noncontingently related data, and the frequency and valence of the outcome.

Other characteristics include the sequence of outcomes, temporal contiguity between response and outcome, and theory- versus data-driven data. Langer and Roth (1975), for example, found that when successes come early in a task and failures come late, subjects tend to be more likely to expect success than when the outcome order was reversed. Events that occur in close temporal contiguity are more apt to be judged as contingently related than are nontemporally related events (Catania & Cutts, 1963). Lastly, judgments of data which are theory-laden are more likely to be inaccurate than are judgments of data which are unrelated to any theory (Alloy & Tabachnik, 1984; Chapman & Chapman, 1967; Jennings et al., 1980).

In sum, a great deal of variety in the setting, task, and data has been found in the study of contingency judgments. It appears that many of these experimental factors lead to differences in judgment strategy and/or accuracy, but few have been directly compared or experimentally examined to determine the precise aspects of those factors that influence judgments. As Wasserman & Shaklee (1984) point out, such laboratory activity may not closely resemble everyday activity, but it may "... more effectively elucidate the psychology of causation than more naturalistic methods, just because their artificiality introduces greater experimental control over the relevant variables and because they are more readily modified in ways that permit isolation of particular cognitive processes" (p. 285).

Subject Factors

Contingency judgments may also be influenced by factors related to the individual making that judgment. These factors may have to do with the person him or herself (e.g., mood, expectations) or, more generally, with the human species (e.g., judgment strategies and heuristics). Since the general factors have been discussed, this section will focus on more particularistic variables.

The subject factor which has attracted most attention in recent years is mood, especially depressed mood. As reviewed in an earlier section, Alloy and Abramson (1979) found that depressed and nondepressed individuals sometimes make different judgments of response-outcome contingency. (Depression is defined on the basis of Beck Depression Inventory scores of 9 or above.) Other research has also confirmed differences in judgments between the two groups (e.g., Alloy & Abramson, 1982; Alloy, Abramson, & Musson, 1983; Alloy, Abramson, & Viscusi, 1981; Martin, Abramson, & Alloy, 1984). But, the emergence of differences and the direction of those differences appears to depend on certain experimental factors (e.g., whether outcomes are contingent upon responses, whether or not the outcome used in the judgmental task is desirable, undesirable, or neutral, whether or not an observer is present, and whether or not the study involves controllability or predictability).

Specifically, Alloy and Abramson (1979) found that when neutral outcomes were contingent upon responses, no between-group differences arose. However, when outcomes are noncontingently related to responses and occur frequently or are desirable, nondepressed people overestimate the degree of control relative to depressed people's

judgments which are fairly accurate. Also, when contingent outcomes are undesirable, nondepressed subjects underestimate the degree of contingency relative to the estimates of depressed people which were fairly accurate.

These judgment differences between groups are not maintained, though, when additional factors are introduced. Benassi and Mahler (in press) found that the relative accuracy of depressed persons' judgments is not apparent (at least in noncontingent cases in which neutral outcomes occur frequently) when an observer is present in the same room as the subject; in fact, nondepressed subjects were more accurate than depressed subjects. When the experimental task involves predictability rather than controllability, it appears that, regardless of whether the events were contingently or noncontingently related or the predicted outcome was frequent or desirable, the differences between the two groups do not occur; both groups provided fairly accurate estimates of predictability (Alloy, Abramson, & Kossman, in press).

The question arises as to why the differences in response-outcome judgments of depressed and nondepressed people should occur at all. It has been suggested (e.g., Abramson & Alloy, 1980; Alloy & Abramson, 1979; Alloy, Crocker, & Tabachnik, 1980) that nondepressed people have strong expectations that they will generally control a given event and/or that they have self-serving attribution biases to perceive control in order to protect or improve their self-esteem in contrast to depressed subjects who appear to lack both of these characteristics. Although these explanations are consistent with the Alloy and Abramson (1979) and Alloy et al. (1980) findings, they are not consistent with

Benassi and Mahler (in press) findings. In fact, it would seem that nondepressed subjects would be more motivated to protect or enhance their self-esteem in the presence of another person and thus would continue to overestimate the amount of judged response-outcome control. More research is needed to answer this question.

At the other end of the continuum lie manic individuals. Langer (1975) has speculated that, in contrast to depressives, they may be prone to an illusion of control. This remains to be empirically studied.

Another variable which may be related to judgments of contingency is locus of control. This variable refers to people's generalized expectancies about whether reinforcements are contingent on their own behavior (internal) or factors such as luck, fate, and other people (external) (Lefcourt, 1972; Phares, 1976; Rotter, 1966). In a study involving a task similar to contingency judgments, Coppel and Smith (1980) divided subjects into two groups based on their locus of control and assessed the length of time it took them to detect a contingency between two stimuli in a predictability (S-S) or a controllability (R-O) situation. They found no between-group differences in the ability to identify contingencies overall, but they did find an interaction of locus of control by type of problem. In line with their predictions, internals (who would expect to perceive a relationship between their response and outcome) detected R-O relationships more quickly than externals; but externals (who would expect to perceive a relationship between outcomes and external events) detected the S-S relationship more rapidly than internals.

Research on depression and judged control casts an interesting light on the relationship between locus of control and accuracy of contingency judgments. Depression has been found to be positively correlated with external locus of control. Based on the thesis of Coppel and Smith, one would expect that externals would not be as proficient judges of response-outcome contingency as internals, but, as reported above, it is known that, in many cases, depressed subjects (and perhaps externals) are more accurate than nondepressed subjects (and perhaps internals) and that there are no differences in judgment accuracy of predictability relationships. It may be that precisely because depressed subjects lack strong response-outcome control expectancies (an internal characteristic) that they, in many cases, are more accurate judges but slower detectors. Some support for this contention stems from a study conducted by Benassi, Sweeney, and Drevno (1979) which revealed that internals provided higher estimates of control than externals in a chance-related task (die tossing) in which subjects were actively involved (threw the die). These differences did not occur when subjects were not actively involved (observed the experimenter throw the die).

Another factor which has received a lot of attention and which is a major component in the Alloy and Tabachnik (1984) model is the expectations a subject has (see also, Crocker, 1981; Nisbett & Ross, 1980). As alluded to above, the Coppel and Smith study actually involves subjects' expectations. Internals were predicted to more readily detect response-outcome contingencies than externals because they expect to perceive a relationship between their behavior and outcomes. Conversely, externals were predicted to more readily detect

stimulus-stimulus contingencies because they expect outcomes to be contingent on external events. The role of expectations in the contingency judgments of depressed and nondepressed subjects has also been discussed.

The "illusory correlation" studies also illustrate the importance of one's expectations. In those studies, subjects reported that a contingency existed even if none did or if a negative one was present. The most common explanation for these findings is that the subjects, like the clinicians, expected to see certain contingencies.

Prompted by the Chapmans' research, Jennings et al. (1980) compared judgments of information which was relatively "theory-free" with that which was "theory-driven" (i.e., information about subjects would have strong expectations). Part of this study was reported above (p. 8). In contrast to judgments of the "theory-free" material which were very conservative, judgments of "theory-driven" material were "far higher than any data derived from first-hand experience could justify" (Nisbett & Ross, 1980 p. 101). Thus, one's expectations appeared to influence judgments.

Subjects may also hold expectations about the experimental situation itself as Peterson's (1980) work suggested. He found support for his contention that subjects often do not recognize noncontingency because they often do not expect it. If subjects' expectations are modified to expect noncontingency, it is more readily detected.

People may also have more general expectations about how much control their responses typically have over outcomes. According to Abramson and Alloy (1980) these expectations are established on the basis of past experiences in which, for example, people learn to expect that when positive outcomes occur frequently and in close temporal contiguity to the response, they have control or that when characteristics such as practice and competition are apparent, they typically have control. Hence, when those features occur in the experimental situations, subjects are likely to judge that their response is controlling the outcome. However, based on locus of control research, these conclusions should be qualified; perhaps they are only accurate for internal subjects.

Finally, it should be noted that sex differences have been recorded in judgment accuracy (Shaklee & Hall, 1983; Wasserman et al., 1983; Wasserman & Shaklee, 1984) and in strategy use (Shaklee & Hall, 1983). It appears that females may use simpler less accurate strategies (e.g., the a vs. b rule) than males and may tend to make less accurate contingency judgments than males. Sex has also interacted with other variables in other studies (Alloy & Abramson, 1979; Alloy et al., in press). For example, females tended to overestimate the degree of contingency when noncontingent outcomes occurred frequently (in both controllability and predictability designs). It has been suggested that these sex differences may stem from the different mathematical backgrounds typically experienced by men and women (Shaklee & Hall, 1983), but certainly more research is needed which directly examines this issue.

The Alloy and Tabachnik Model
for Understanding Contingency Judgments

In an effort to integrate the findings in this area, Alloy and Tabachnik put forth a theoretical framework in which they purport to be able to interpret "all of the studies that examine people's covariation detection abilities" (p. 123). This model will be critiqued and, in the following section, two studies are presented to test this model.

Alloy and Tabachnik propose that the judgment of the degree of contingency between two events is based on the interaction of two pieces of information: 1) "situational information about the objective contingency between the events provided by the current environment" and 2) "the organism's prior expectations or beliefs about the event covariation in question" (p. 114). The authors continue that whether or not a judgment is accurate is a function of the relative strength of these two components. Strength, in regard to situational information, is defined as the amount of information about the covariation available to the person and, in regard to expectations, refers to the extent to which the person already has beliefs about the particular event covariation.

Although Alloy and Tabachnik emphasize that strength of these variables should be seen as lying on continua, for illustrative purposes, they have dichotomized the strengths of the variables and adopted a 2 X 2 table used by Metalsky and Abramson (1981). High and low strengths of current situational information are crossed with high and low strengths of prior expectations, resulting in four possible conditions under which contingency judgments may be made.

Predictions of covariation judgments may be straightforwardly determined for three of the conditions. When the strength of both components is low (Cell 1), the person will probably not make a covariation judgment or will make one with low confidence. When the strength of the situational information is low and that of prior expectations is high (Cell 2), the individual will make a judgment following his/her expectations. Whether or not the judgment is accurate, according to the model, depends on the veridicality of his/her beliefs. Conversely, when the strength of the situational information is high and that of prior expectations is low (Cell 3), a judgment will be made in line with the situational information. This judgment should be accurate, according to the model.

When the strength of both factors is high, as in Cell 4, contingency judgments may not be predicted without more information. There are actually two conditions that may arise in this cell: 1) both variables lead to the same covariation judgment (congruous) or 2) the variables imply contrary judgments (incongruous). In the first case, the resultant judgment should be accurate and made with high confidence. In the second case, the judgment cannot be predicted. According to Alloy and Tabachnik, the person in Case 2 is in a "cognitive dilemma;" they may disregard current situational information and make the judgment based on his/her beliefs and expectations or disregard the latter information and make a judgment in line with the situational information. They add that people tend to opt for the former resolution in which case judgments follow beliefs and expectations and are not likely to be accurate.

Alloy and Tabachnik sum up by the following: "Judgments of covariation are relatively accurate when people lack strong beliefs about the event relationship in question (Cell 3) or when the situational information concerning the objective correlation between the events is congruent with people's preconceptions about the event relationship (Case 1 of Cell 4). When objective data and preconceptions are incongruent (Case 2 of Cell 4), judgments of covariation are frequently erroneous and biased in the direction of initial expectations" (p. 123). They add that "all of the studies that examine people's covariation detection abilities can be interpreted in these terms" (p.123).

Although much of the work in this area may be "interpreted in these terms," the interpretation is post hoc (a drawback recognized by Alloy and Tabachnik) and, of course, such interpretation is not definitive with respect to the validity of the model. Further, it is argued here that "all of the studies that examine people's covariation detection abilities" cannot be "interpreted in these terms." Rather, Alloy and Tabachnik presented a selective review of studies that can be subsumed by their model.

In regard to Alloy and Tabachnik's interpretation of past studies, since the researchers have not conducted experiments to actually test their suppositions, alternative interpretations cannot be ruled out. The following examples taken from the authors' (1984) article illustrate this point.

Alloy and Tabachnik cite studies by Beach and Scoop (1966), Erlick (1966), and Erlick and Mills (1967) as support for their model. In these studies, subjects made relatively accurate judgments of the degree of contingency between two sets of continuous stimuli. Alloy and Tabachnik point out that the pairs of stimuli used in these studies were devoid of any familiarity to subjects and, hence, "subjects rely on the available situational information in making their judgments and, thus, subjective judgments of correlation mirror objective correlations" (p. 123). This may be correct, but since no studies have been conducted in which familiar stimuli have been paired together in a similar way, other factors may be responsible. Representation of variables (i.e., symmetrical or asymmetrical), for example, may have affected judgments (see Allan & Jenkins, 1980, 1983).

Ward and Jenkins' (1965) study is also said to support their model. This work used three modes of presentation and involved subjects' judgments of contingency of cloud seeding and rain. Alloy and Tabachnik state that "it is common knowledge that cloud seeding is at least partially effective in producing rain, so it is reasonable to assume that Ward and Jenkins' subjects expected cloud seeding to be followed by rain. Thus, the use of expectation confirming cases offers a plausible account of Ward and Jenkins' data" (p.123). What Alloy and Tabachnik fail to mention is that when the data were presented (only) in a summary table form, judgments were relatively accurate. One may argue that such data were "stronger" than that seen in a serial presentation, but it cannot be convincingly argued, given the description of situational information provided by Alloy and Tabachnik, that this information was "stronger" than the same information coupled

with a serial presentation. At a minimum, additional information would need to be incorporated into the model in order to account for these findings.

Other experiments are also discussed in support of their model, but as stated above, since no studies have empirically tested their model, other explanations cannot be ruled out. Even the work of Chapman and Chapman (1967, 1969) which is probably the work most readily interpretable by the Alloy and Tabachnik framework is only consistent with the model; it does not actually support the model since neither expectations nor situational information are manipulated.

Finally, Alloy and Tabachnik's contention that "all studies" may be interpreted by their model is supported--albeit tenuously--only in their restricted review of the literature. For example, conspicuous by its absence in their review is the work of Allan and Jenkins (1980, 1983) whose work has suggested that the number of response and outcome alternatives (mode of representation) affects judgment accuracy. It is difficult to understand, for example, why expectations of control would be different due to the number of response alternatives or why the situational information would be considered different in those cases. Studies which have shown that memory affects judgments (e.g., Arkes & Harkness, 1983; Shaklee & Mims, 1982) cannot be accounted for by their model and are not reviewed by those authors.

Thus, it appears that this model serves effectively as a heuristic tool which enables the organization and conceptualization of the contingency judgment research on a general level. Support for the explanatory and predictive validity of the model has been weak but,

heretofore, no studies have disconfirmed the model either. It is contended here, though, that information contributed by one's expectations in conjunction with the objective data as defined by Alloy and Tabachnik, is not always sufficient to explain and predict contingency judgments of those data. The present experiments tested this contention. Before discussing these studies, several definitional weaknesses are reviewed which have made testing the validity of the model a conceptually difficult task.

The major components of this model do lack definitiveness and, to the extent that they are defined, there is no clear objective method of assessment. For example, it is not stated whether situational information refers to objective or subjective information. One would assume the authors are referring to objective information since subjective information, by its very nature, would already have been affected by various biases.

The question then arises as to the level of objective information about which they are referring. A judgment differs from a perception in that a series of events is typically observed and there probably are several steps preceding a judgment whereas a perception refers to an immediate and single experience (Jenkins & Ward, 1965). What step in Crocker's (1981) model are Alloy and Tabachnik referring to in their use of the term "situational information"? The distal stimuli perceived at Step 3 or the recalled number of confirming and disconfirming cases at Step 4? This is important because Step 4 information is probably not an accurate reflection of Step 3 information due to errors of a purely cognitive nature (e.g., memory).

Further, the boundaries of the situational information are not defined. Again, are they referring to the actual stimuli about which a judgment is being made or to those stimuli within a broader context of time and space? Since they refer to the "current environment," one might assume that they are referring to the specific situation that is being judged.

The meaning of expectations is also unclear as Alloy and Tabachnik acknowledge in a footnote. The major question centers on the specificity of expectations: "Are prior beliefs about only the specific events involved in an event covariation (e.g., button pressing and nickels) sufficient or are expectations about abstractions of these events (e.g., one's own responses and positive outcomes) also relevant?" (p.114).

One of the key ideas in this model is the "strength" of each component which, in both cases, is vaguely defined. (Both definitions were presented earlier.) In the case of situational information, the authors elaborate on their definition by stating that "current situational information can be unavailable or weak because it is insufficient in quantity to support a covariation perception (e.g., the organism has had little experience with the events in the current situation) and/or because it is ambiguous (e.g., it is not very diagnostic)" (p. 114). When these conditions are apparent is not made clear. When expectation strength may be "unavailable or weak" is not described. In both cases, the notion of strength is conceptually somewhat appealing but it is not readily amenable to empirical assessment and, therefore, not explanatorially or predictively useful

information.

The issue of the definitiveness of the terms, situational information and expectations, is important because lack of it makes testing the validity of the model difficult. These definitions may be broadened in an attempt to account for all findings; although, such action would increase post hoc explanatory power while sacrificing predictive power. For purposes of testing the validity of this model in the present studies, the statements made by Alloy and Tabachnik about these factors are employed as they are stated.

III. THE PRESENT STUDIES

The purpose of the present studies was to empirically examine the predictive validity of the model put forth by Alloy and Tabachnik (1984) to account for contingency judgments. In general, it is maintained here that the combined influence of expectations and situational information will not always be sufficient to account for judgments of contingency. Specifically, there is a large body of empirical and theoretical literature which has revealed that the response to a stimulus is a function of the stimulus itself as well as other stimuli that make up a context, or frame of reference. Therefore, the same stimulus or situational information may be judged differently in different contexts. If this contention is correct, then the Alloy and Tabachnik model would not have predictive validity.

The contrast effect is a classic example (see Eiser & Stroebe, 1972). It refers to the judgment of a stimulus (or series thereof) along a specified dimension which is displaced away from a reference or anchor point on that scale.

For example, Manis and Paskewitz (1984) found that subjects exposed to "high pathology" definitions of vocabulary words subsequently rated midscale definitions as less pathological than those subjects who were initially exposed to "low pathology" definitions. This pattern of results has emerged in a wide range of psychophysical and psychosocial areas, such as in judgments of weight and color perception (Helson, 1964), line length (Krantz & Campell, 1961), temporal stimulus duration (Postman & Miller, 1945), the distance between two lines (Parducci, 1968), facial expression (Manis, 1971),

and severity of different crimes and judicial sentences (Pepitone & DiNubile, 1976). Contrast effects have also been observed in the performance of animals who, for example, are shifted from large to small reward and then perform at a lower rate (e.g., run slower) than animals who had initially had been maintained on the same small reward (e.g., Bower, 1961; Collier & Marx, 1959; Ferrell & Shanab, 1975).

Although this effect is common and widespread, there is no consensus as to what a contrast effect is theoretically or to what the origins of this effect are (see Eiser & Stroebe, 1972). Although the purpose of the project was not to disentangle either of these issues, the major positions are that 1) a central mechanism or perceptual process is involved (Helson, 1964) or that 2) a peripheral mechanism or linguistic process is responsible (Volkman, 1951). Proponents of the first position argue that changes in judgment result from actual changes in the stimulus itself. Proponents of the second position contend that shifts in judgment stem from differential use of the scale itself. There is support for both views (see Eiser & Stroebe, 1972).

In contingency judgment research, context could refer to previous experience with the same type of stimuli varied along the dimension of degree of control. So, judgments of control of the same situational information may be different and contrary to expectations based on previous experience because of the initial experience with situational information of the same type, but differing in degree of control. Alloy and Tabachnik's model apparently could not account for such findings. The present series of experiments examined the role that expectations and contrast effects play in determining judgments of contingency.

Experiment 1

In this experiment, judgments of noncontingent response-outcome trials were examined as a function of initial exposure to a set of trials with one of three different degrees of control (High, Low, or None). Each subject received two sets of response-outcome contingency trials. The first set consisted of trials in which there was a high-, low-, or non-contingent relationship between responses and outcomes. The second set of trials, identical to the aforementioned third condition, was given to all subjects.

On the basis of a contrast effect perspective, it was hypothesized that if subjects have initial experience with a response-outcome situation in which they have high control (HC), then their judgments about later noncontingent response-outcome situations (NC) (which are typically overestimated) will be lower than those of subjects who have initial experience with response-outcome situations in which they have low (LC) or no control. Because the difference between LC and NC trials in group LC-NC is not large and because there is no difference between the two sets of trials in group NC-NC, those groups should not have significantly different judgments between Set 1 and Set 2 trials.

Alloy and Tabachnik's model could not account for such findings. Instead, their model would have to predict one of three different outcomes. Before presenting those predictions, two points should be emphasized. The "situational information about the objective contingency between the events provided by the current environment" (Alloy & Tabachnik, p. 114) should be precisely the same for each of the three conditions. However, based on the model, the subjects'

"prior expectations or beliefs about the event covariation in question" (Alloy & Tabachnik, p. 114) should differ for each condition. The three predictions for judgments of the second set of data, deduced from Alloy and Tabachnik's model, are presented below.

Prediction 1: If both sources of information were of equal "strength" (and this is not possible to determine or quantify), subjects in the first two groups (HC-NC, LC-NC) would be in a "cognitive dilemma," and, according to Alloy and Tabachnik, would probably make a judgment in line with their expectations. Subjects in the third condition (NC-NC) would be in a "fortunate" position in which expectations and situational information match one another. Thus, it is predicted that judgments of the second set of data would be in descending order for the three conditions respectively. This prediction would also be made if the situational information were considered not as "strong" as the expectations.

Prediction 2: If, on the other hand, situational information were considered "stronger" than expectational information, judgments for each of the three conditions would be predicted to be the same and in line with situational information (noncontingency). However, there is no reason to assume that situational information would be "stronger."

Prediction 3: If both sources of information were considered weak, it is not possible to predict judgments, according to the model, but, it would seem probable that they would not differ among groups.

Because the strength of these components cannot be empirically assessed, one can only guess at what may occur. But, under no circumstances, would the model predict the same findings as that based on contrast effects. That is, judgments of the second NC set would be lowest for subjects with previous exposure to HC.

Method

Subjects. Sixty-six female students from Introductory Psychology courses at the University of New Hampshire participated in this experiment as part of their course requirements. (Subjects for all experiments were restricted to women because judgment variability is high in this area of research and, since sex differences have been observed, one way of reducing this variability was to include only one gender. Women were chosen because they have been typically employed in other similar studies, thus facilitating comparisons with those studies.)

Design. A three-group design was employed in which there were three levels of context (High-, Low-, or No-Control on the first set of trials). In the High-Control (85-40) condition, subjects had approximately 45% control; in the Low-Control (80-60) condition, they had approximately 20% control; and in the No-Control (70-70) condition, they had no control. Following the procedure of Alloy and Abramson (1979), the first number in the parentheses referred to the percentage of trials in which the outcome (the blue light) occurred when the subject responded (pressed the button); the second number referred to the percentage of trials in which the outcome occurred when the subject did not respond (did not press the button). Based on the "delta p"

coefficient, the degree of control can be determined by the difference between the two numbers. (The terms, high, low, and no control, are accurately descriptive in Experiment 1. The values that these terms represent are not comparable to those represented by the terms in Experiment 2.)

The outcome frequency for each group was as similar as possible, given the parameters of the apparatus. The outcome occurred on approximately 68% (28) of the trials for the LC and NC conditions and on about 62% (25) of the trials for the HC condition. Any effect that this difference may have would work contrary to the hypothesis since outcome frequency has been shown to be positively correlated with judgments of control (Jenkins & Ward, 1965).

Apparatus. The experiment was conducted in two separate rooms. In one room a standard probability generator with conventional relay circuitry was located. It was programmed to determine outcome presentation and to record subjects' responses. In the other room the apparatus which the subjects used in the experimental task was located. It consisted of a 15 X 50 X 5 cm white cardboard box with two 2 cm round lights situated atop. The left light which, when on, signified the start of a new trial, was yellow and the right light which, when on, was the outcome of interest and was blue. To the right of the box was an 8 cm lever which the subject pressed when she chose to respond.

Materials. One "Judgment of Control" scale and one questionnaire were used. On the scale, taken from Alloy and Abramson (1979), subjects were asked to indicate "the degree of control that you believe your responses (pressing and not pressing the button) had over the

appearance of the blue light." The scale ranged from 0 (no control) to 100 (complete control), marked off in intervals of 5 units.

On the questionnaire, subjects were asked to estimate the number of times they pressed and received the outcome, pressed and did not receive the outcome, did not press and received the outcome, and did not press and did not receive the outcome (i.e., the information necessary to fill in a 2 X 2 contingency table). Other questions centered on the evidence that they used to determine how much control their response exerted over the outcome.

Procedure. Participants received two problem sets of 40 3-sec trials each. A yellow light signaled the onset of a trial at which time the subject had the option of pressing or not pressing the lever. At the end of 3 sec, the blue light would appear or not appear, depending on the subject's response and the condition in which she was assigned. The inter-trial-interval (ITI) was 5 sec.

The actual procedure that was followed was adapted from Alloy and Abramson (1979). After the subject was greeted, she was seated at a table facing the task apparatus. All subjects were then read the following instructions, adapted from Alloy and Abramson.

Now, in this problem-solving experiment, it is your task to learn what degree of control you have over whether or not this blue light comes on. Each time the yellow light comes on indicates the start of a new trial, the occasion to do something. For each trial, after the yellow light comes on, you have the option of either making a button press response or not making a button press response. A button press response consists of pressing this button once and only once immediately after the yellow light comes on. Not making a button press response consists, of course, of doing nothing when the yellow light comes on. If you do intend to press the button on a given trial, you must press within three seconds after the yellow comes on; otherwise the trial will be counted as a no press trial. So, in this experiment there are only two possibilities as to what you can do on each of the trials: either press the button within three seconds after the yellow

light comes on, or else, just sit back and do nothinAny questions so far?

You may find that the blue light will go on, on some percentage of the trials on which you do make a button press response. You may also find that the blue light will go on, on some percentage of the trials when you do not make a button press response. Alternatively, you may find that the green light will not go on, on some percentage of the trials on which you do make a button press response. And, you may find that the green light will not go on, on some percentage of the trials when you do not make a button press response. So there are four possibilities as to what may happen on any given trial: 1) you press and the blue light does come on; 2) you press and the blue light does not come on; 3) you don't press and the blue light does come on; 4) you don't press and the blue light does not come on. Since it is your job to learn how much control you have over whether the blue light comes on, as well as whether the blue light does not come on, it is to your advantage to press on some trials and not on others, so you know what happens when you don't press as well as when you do press. (When it is clear that the subject understands the outline of the task, she will then be shown the Judgment of Control scale and the concept of control will be discussed briefly.)

Forty trials will constitute a problem. You will have two sets of forty trials, that is, two problems. After each problem, you will be asked to indicate your judgment of control by putting an 'X' someplace on this scale: at 100 if you have complete control over the onset of the blue light, at 0 if you have no control over the onset of the blue light, and somewhere between these extremes if you have some but not complete control over the onset of the green light. Complete control means that the onset of the blue light on any given trial is determined by your choice of responses, either pressing or not pressing. In other words whether or not the blue light goes on is totally determined by whether you choose to press or to just sit back and not press. No control means that you have found no way to make response choices so as to influence in any way the onset of the blue light. In other words, the onset of the blue light has nothing to do with what you do or don't do. Another way to look at having no control is that whether or not the blue light comes on, on any given trial, is totally determined by factors such as chance or luck, rather than by your choice of pressing or not pressing. Intermediate degrees of control means that your choice of responses, either pressing or not pressing, influences the onset of the blue light even though it does not completely determine whether the blue light goes on or not. In other words, what you do or don't do matters to some extent but not totally. Another way to look at having intermediate control is that one response, either pressing or not pressing, produces the green light onset more often than does the other response, but the other response may also produce the blue light onset. So, it may turn out that you will have no control, that is, your responses will not affect the onset of the blue light, or it may turn out that you will have some degree of control, either complete or intermediate, that is, one response produces blue light onset more often than does the other. Any questions before we begin?

The experimenter then left the room and initiated the start of the first problem set. Upon completion of 40 trials the experimenter returned and asked the subject to record her judgment on the Judgment of Control scale. All subjects were then told that they would receive another problem set in which they "may or may not have the same amount of control as they did in the first set."

The experimenter then left the room again and initiated the second problem set. Upon completion of the second set of forty trials, the experimenter returned and asked the subject to fill out the Judgment of Control scale. When the subject finished with that task, she was given the questionnaire and asked to "read the following questions and answer them as accurately as possible for the last set of trials only."

At that point, the experiment was over and subjects were debriefed.

Results

Hypothesis Testing. In order to test for a contrast effect, a one-way analysis of variance (ANOVA) was conducted to determine the effect of amount of control on judgments of control in the second set of trials. If a contrast effect occurred, subjects with initial experience with HC trials should provide judgments for the NC trials in the second set that are lower than judgments of those subjects who had initial experience with LC or NC trials. (Results pertinent to groups LC-NC and NC-NC are presented later.)

The predicted contrast effect was reflected by the direction of the Set 2 means for judgments of control in the HC-NC and NC-NC groups (see Table 1), but it was not substantiated by a one-way ANOVA [$F(2,63) = 2.12, p = 0.12$]. The only significant difference was between the Set 2 judgments of groups LC-NC and HC-NC, as revealed in the post-hoc tests [$t(41.8) = 2.06, p = 0.05$]. (These t-tests are two-tailed.)

However, an examination of the descriptive statistics of Set 2 judgments of control showed that the distribution of control judgments of groups HC-NC and NC-NC were positively skewed, which suggested that a median test for a contrast effect would be a more appropriate analysis. This test revealed that the Set 2 judgments of the two groups were significantly different from one another [$\chi^2(1) = 25.99, p < .001$]. That is, there were significantly more judgments below the overall median in group HC-NC than in group NC-NC, indicating that a contrast had occurred.

Manipulation Checks. Several other analyses were conducted and are presented below. They focused primarily on manipulation checks.

First, as can be seen in Table 2, the empirical delta p's for the LC and NC groups in Set 1 and all groups in Set 2 were somewhat higher than the expected delta p's. T-tests revealed that these differences were significant in Set 1 [$t(21) = 4.66, 7.03, p < .01$ for LC, and NC, respectively] and in Set 2 [$t(21) = 4.06, 7.01, 7.75, p < .01$ for HC-NC, LC-NC, and NC-NC, respectively]. In Set 1, the delta p's for the HC condition did not differ from one another [$t(21) = .83, p > .05$]. (T-tests for all comparisons involving HC and LC were two tailed; those involving NC were one-tailed.)

Table 1
Descriptive Statistics for Judged Control in
Experiment 1 (n=22)

<u>Set 1</u>						
<u>Condition</u>	<u>Mean</u>	<u>Mode</u>	<u>Median</u>	<u>SD</u>	<u>Skew</u>	<u>SE</u>
HC-NC	44.55	45.00	47.50	24.39	-0.44	5.20
LC-NC	46.73	0.00	49.67	24.45	-0.69	5.21
NC-NC	32.50	0.00	27.50	28.82	0.18	6.14

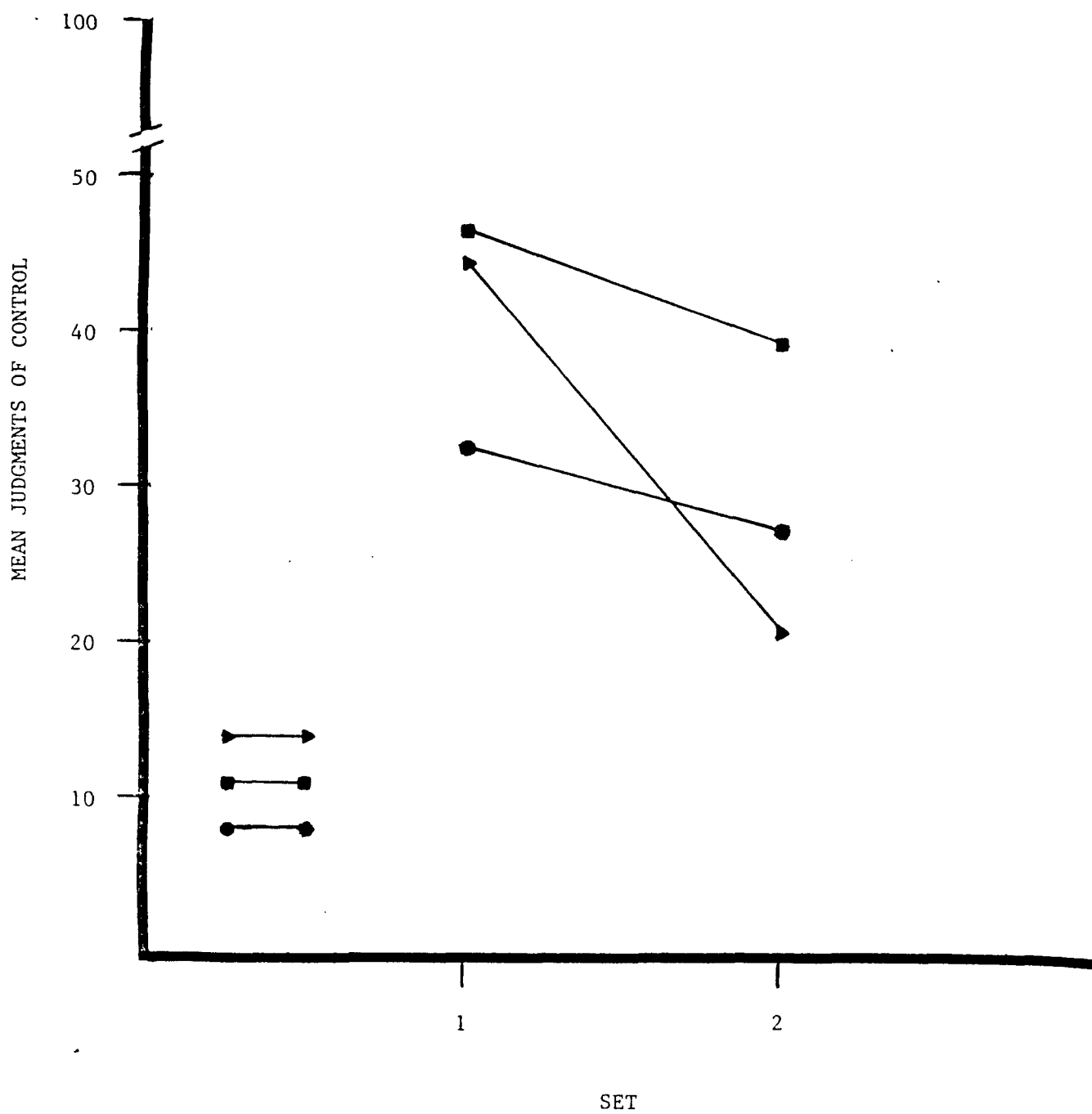
<u>Set 2</u>						
<u>Condition</u>	<u>Mean</u>	<u>Mode</u>	<u>Median</u>	<u>SD</u>	<u>Skew</u>	<u>SE</u>
HC-NC	20.91	0.00	2.50	28.44	0.95	6.06
LC-NC	39.23	0.00	45.00	30.59	-0.00	6.52
NC-NC	27.27	0.00	22.50	28.77	1.18	6.14

Table 2
Mean Percentage of Expected and Obtained Control in
Experiment 1

	<u>Set 1</u>		
	<u>HC-NC</u>	<u>LC-NC</u>	<u>NC-NC</u>
Expected Delta p	45.00	20.00	0.00
Empirical Delta p	47.71	31.32	10.26

	<u>Set 2</u>		
	<u>HC-NC</u>	<u>LC-NC</u>	<u>NC-NC</u>
Expected Delta p	0.00	0.00	0.00
Empirical Delta p	15.83	13.97	12.09

Figure 1. Mean Judgments of Control in Set 1 and Set 2 of Experiment 1



Second, in Set 1, as planned, there were significant differences in the expected direction in the empirical delta p's among the groups [$F(2,63) = 57.05, p = 0.00$]. It is interesting to note, though, that the control judgments provided by subjects for these empirical delta p's did not differ among groups [$F(2,63) = 1.92, p = 0.16$]. In Set 2, as planned, there were no differences in the empirical delta p's among groups [$F(2,63) = 0.48, p = 0.62$]. It has already been noted that the control judgments provided by subjects for these empirical delta p's did differ among groups.

Third, because outcome frequency has been found to be related to judgments of control, the role of that factor was also assessed. In Set 1, a one-way ANOVA revealed that there were significant differences among groups [$F(2,63) = 5.41, p = 0.01$]. A Scheffe test showed that groups NC-NC and LC-NC each had the same outcome frequency, but that they each had a higher outcome frequency ($p < .05$). These findings are in accord with the amount of planned outcome frequency and, as stated earlier, work contrary to the hypothesis. Importantly, in Set 2, there were no significant outcome frequency differences among groups [$F(2,63) = 0.89, p = 0.42$].

In order to determine whether outcome frequency differed across trials, three within-subject t-tests were conducted for each group. The amount of outcome frequency did not change from Set 1 to Set 2 in group LC-NC [$t(21) = 1.12, p = 0.28$] or in group NC-NC [$t(219) = 0.74, p = 0.47$]. However, outcome frequency in group HC-NC, as would be expected, increased from Set 1 to Set 2. This finding also operates contrary to the hypothesis.

Finally, a split-plot ANOVA was conducted to determine the effect of levels of control (HC, LC, and NC) and set (Set 1, Set 2) on judgments of control. This revealed a significant trial by condition interaction [$F(2,63) = 3.67, p = 0.03$] due to group HC-NC. Judgment of control scores for that group were significantly lower in Set 2 than in Set 1 [$F(1,3) = 20.35, p < .05$]. The judgments of groups LC-NC and NC-NC, as predicted, remained the same across trials. (See Figure 1.)

Discussion

The primary hypothesis was that subjects who have initial exposure to HC trials in Set 1 will judge the amount of control in NC trials in Set 2 as lower than will subjects who have had initial experience with NC trials and are then shifted to the Set 2 NC trials. Although this was not confirmed by the conventionally employed analysis (a one-way ANOVA), it was confirmed by a median analysis, which is probably a more appropriate analysis for these data.

Also, as expected the judgments of groups LC-NC and NC-NC did not differ from Set 1 to Set 2. Judgments of subjects in group HC-NC did, as expected, drop significantly from Set 1 to Set 2.

The results indicate that the context in which a set of trials is presented does affect judgments of that set of trials. The same situational information (a set of NC trials) was judged differently in different contexts. Specifically, the amount of control in NC trials, when preceded by HC trials was judged lower than when preceded by LC trials or the same NC trials. Alloy and Tabachnik's model, which proposes that judgments of control are determined on the basis of an interaction of information from the situation and from one's

expectations, does not appear to be able to account for the Set 2 judgments of group HC-NC in relation to group NC-NC.

It should be noted that because the empirical delta p's in Set 2 were significantly higher than zero, the difference between HC trials in Set 1 and NC trials was reduced, thus perhaps even decreasing the strength of the contrast effect.

In addition, the fact that outcome frequency actually increased from Set 1 to Set 2 and yet judgments of control significantly decreased across trials further supports the role of previous experience with different levels of control on later judgments of control. In other studies, outcome frequency has been positively correlated with judgments of control. It may well have been that Set 2 judgments of group HC-NC would have been even lower if outcome frequency had remained the same across trials. If outcome frequency is viewed as a context variable, it may have acted to increase scores in Set 2. It is not known whether this effect occurred since it is not known whether the outcome frequency difference between the two sets was large enough to lead to a contrast effect.

Returning to the predictions derived from the Alloy and Tabachnik model, Set 2 judgments of group HC-NC would be expected to be the same as those of group NC-NC (situational information wielding strength) or higher than those of group NC-NC (expectations wielding strength or the components being of equal strength). It would seem that under no circumstances, would NC judgments of group HC-NC be predicted to be lower than those of group NC-NC.

It is important to reinforce the fact that the aforementioned finding did not occur because of an anomalous rise in the Set 2 judgments of group NC-NC. As both Alloy and Tabachnik and a contrast effect orientation would predict, Set 2 judgments remained the same as Set 1 judgments in group NC-NC.

Thus, Alloy and Tabachnik's model could predict the NC-NC as well as the LC-NC Set 2 judgments. However, whether or not the underlying mechanisms they put forth to account for these findings are valid remains to be confirmed. In addition, if the underlying mechanisms are valid, the question then arises as to which one of the underlying mechanisms accounts for these findings (e.g., are these findings due to stronger situational information, stronger expectations, or equally strong components?). But, if neither component, independently or interactively, can predict the HC-NC finding, why, then, should these components--alone or in conjunction with one another--predict the NC-NC or LC-NC findings? It seems that the Alloy and Tabachnik model is not always sufficient to predict contingency judgments.

In order to test the reliability of the contrast effect, a second experiment was conducted which was a conceptual replication and extension of the present experiment.

Experiment 2

Like the previous experiment, the purpose of Experiment 2 was to test the predictive validity of Alloy and Tabachnik's model by examining judgments of control of two sets of trials as function of initial exposure to different sets of trials. Experiment 2, however, employed another paradigm which has also been used to investigate

contrast effects (e.g., see Eiser & Stroebe, 1972; Manis & Paskewitz, 1982, 1983). It set the stage for a conceptual replication of the earlier findings as well as an extension of those findings.

As in Experiment 1, each subject received two sets of 40 trials. However, in the present study, there were only two groups. One group received initial experience with HC trials; the other group was first exposed to NC trials. Then, in Set 2, both groups were shifted to 40 trials in which there was a medium degree of response-outcome contingency (MC). The two groups are referred to as HC-MC and NC-MC, respectively.

It was hypothesized that if subjects have initial experience with HC, then their judgments about subsequent MC trials will be lower than those of subjects with initial exposure to NC trials. If this hypothesis is confirmed, the previously observed contrast effect will be replicated using an alternative design. Furthermore, the previous findings will be extended by demonstrating that judgments may increase (as well as decrease) across trials to provide a contrast effect.

According to the predictions derived from Alloy and Tabachnik's model, the aforementioned results should not occur. Instead, on the basis of their model, they would appear to have to put forth one of the following predictions.

Prediction 1: If both sources of input were of equal "strength" (which is not possible to determine or quantify), subjects would be in a cognitive dilemma which would probably be resolved by following their expectations. Thus, it is predicted that HC subjects will judge the

Set 2 trials as higher in response-outcome contingency than will NC participants. The same prediction would be proposed if situational information were considered not as "strong" as the expectations.

Prediction 2: If, situational information was considered "stronger" than expectational information, Set 2 judgments for the two groups would be predicted to be the same and in accord with situational information (MC). But, as stated in Experiment 1, there is no a priori reason to assume that situational information would be "stronger."

Prediction 3: If both sources of information were considered weak, it is not possible to predict Set 2 judgments, but, it would seem likely that the judgments would not differ.

As stated in the first experiment, since the strength of these components cannot be empirically determined, one does not know which premise to accept and, therefore, which prediction to put forth. But, most importantly, under no circumstances, would the model predict the same results as those proposed on the basis of a contrast effect orientation.

Method

Subjects. Thirty female students from Introductory and Social Psychology courses at the University of New Hampshire participated in this study as part of their course electives.

Design. A two group design was used in which there were two levels of context (High- or No-Control on the initial set of trials). The second set of trials (Medium-Control) was judged within one of the above contexts. The actual amount of control received by these groups

differed from that in the respective groups in Experiment 1, but the same procedure used in that experiment to establish and describe the amount of control was maintained in this experiment. In HC (70-0), participants had approximately 70% control; in MC (55-10), subjects had approximately 45% control; and in NC (35-35), they had approximately 0% control.

The outcome frequency was as similar as possible for each condition, given the parameters of the apparatus. The outcome occurred on about 35% (14) of the trials for HC and NC and on approximately 31% (13) of the trials for MC.

Apparatus and Materials. Refer to the respective sections in Experiment 1, as these remained the same.

Procedure. The procedure remained the same as the previous experiment with one exception. Four individuals were trained to assist the present researcher in conducting this experiment.

Results

In order to determine whether any experimenter effects were present, two one-way ANOVAs on Set 1 and Set 2 judgments of control were conducted. This yielded no significant differences among experimenters [$F(3,26) = 0.59, p = 0.63, F(3,26) = 0.01, p = 0.99$, respectively]. Thus, the data were collapsed across this factor.

Hypothesis Testing. In order to test for a contrast effect, a one-way ANOVA was carried out on judgments of control in the second set of trials. This revealed that a contrast effect had occurred since Set 2 judgments of group HC-MC were significantly lower than those

judgments of the same trials made by group NC-MC [$F(1,8) = 9.13, p = 0.01$]. (See Figure 2.) Unlike the Set 2 judgments in Experiment 1, these data were not skewed (see Table 3).

Manipulation Checks. First, t-tests were conducted to determine whether the expected delta p's mirrored the actual delta p values. In Set 1, the actual values were significantly higher than the expected delta p's for both group HC-MC [$t(14) = 8.53, p < .01$, using a two-tailed test] and group NC-MC [$t(14) = 18.83, p < .01$, using a one-tailed test]. In Set 2, the expected delta p values were reflected by the empirical delta p values [$t(14) = -1.61$ and $1.48, p > .05$, for HC-MC and NC-MC, respectively, using two-tailed tests]. (See Table 4.)

Second, the empirical delta p's for each group within each set were examined. As planned, the HC empirical delta p was significantly higher than the NC empirical delta p [$F(1,28) = 304.52, 0.00$]. Set 1 judgments of control were also significantly different [$F(1,28) = 21.87, p = 0.00$]. In Set 2, as planned, the empirical delta p's did not differ between groups [$F(1,28) = 0.31, p = 0.58$]. As stated above, the judgments of control for Set 2 trials did differ between groups, thus producing the contrast effect.

Third, outcome frequency differences between groups and across trials were examined. A one-way ANOVA indicated that in Set 1, the outcome occurred more frequently in group HC-MC ($M = 18.33, SD = 3.54$) than in group NC-MC ($M = 12.40, SD = 3.02$) [$F(1,28) = 24.40, p = 0.00$]. Importantly, though, there were no between-group differences in Set 2 (HC-MC: $M = 14.67, SD = 4.32$; NC-MC: $M = 16.07, SD = 3.45$) [$F(1,28) = 24.40, p = 0.00$]. The reader may recognize that these values are

Table 3
Descriptive Statistics for Judged Control in
Experiment 2 (n=15)

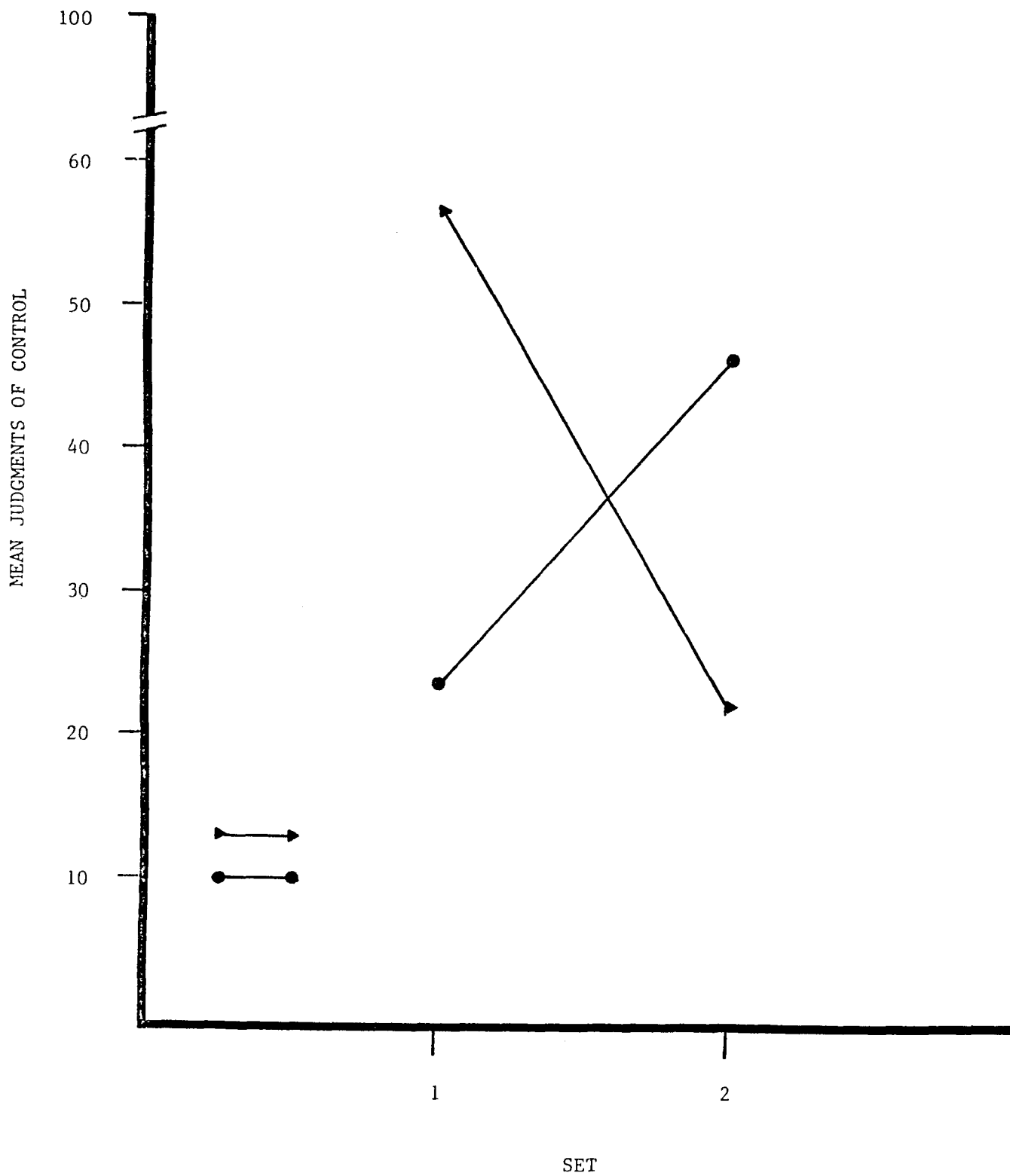
<u>Set 1</u>						
<u>Condition</u>	<u>Mean</u>	<u>Mode</u>	<u>Median</u>	<u>SD</u>	<u>Skew</u>	<u>SE</u>
HC-MC	56.67	75.00	68.33	23.04	-0.63	5.95
NC-MC	23.67	10.00	20.00	14.69	0.79	3.79
<u>Set 2</u>						
<u>Condition</u>	<u>Mean</u>	<u>Mode</u>	<u>Median</u>	<u>SD</u>	<u>Skew</u>	<u>SE</u>
HC-MC	22.00	0.00	16.25	19.25	0.59	7.97
NC-MC	46.33	25.00	53.75	24.53	-0.28	6.33

Table 4
Mean Percentage of Expected and Obtained Control in
Experiment 2

	<u>Set 1</u>	
	<u>HC-MC</u>	<u>NC-MC</u>
Expected Delta p	70.00	0.00
Empirical Delta p	75.39	12.77

	<u>Set 2</u>	
	<u>HC-MC</u>	<u>NC-MC</u>
Expected Delta p	45.00	45.00
Empirical Delta p	42.92	46.18

Figure 2. Mean Judgments of Control in Set 1 and Set 2 of Experiment 2.



different than those which were planned (14, 13, and 14 for conditions HC, MC, and NC, respectively)].

Two within-subject t-tests were also conducted to determine whether outcome frequency differed across trials. As expected from examining the outcome frequency means, these tests revealed that there was a significant downward shift from HC to MC [$t(14) = 2.85, p = 0.01$] and a significant upward shift from NC to MC [$t(14) = -2.87, p = 0.01$].

Discussion

The hypothesis that Set 2 judgments of control would be greater for group NC-MC than for group HC-MN was confirmed. That is, a contrast effect occurred.

The present results corroborate the findings observed in Experiment 1 which demonstrate that the context in which a set of trials is judged has an effect on those judgments. The same situational information (a set of MC trials) was judged differently in different contexts. Specifically, MC trials were judged to have less response-outcome contingency when preceded by HC trials than when preceded by NC trials.

In addition, the present findings show that a contrast effect may be obtained using sets of trials in which there is a nonzero amount of contingency and, thus, a situation in which judgments must increase (rather than decrease as in Experiment 1) to obtain a contrast effect.

Alloy and Tabachnik's model apparently cannot account for the judgments of control provided by subjects in Set 2. The judgments should be expected to be the same (situational information being

stronger) or different from one another in the reverse direction (i.e., the HC-MC group having higher Set 2 judgments than the NC-MC group) (expectations being stronger or the strength of the two components being equal). In any case, the present findings should not be predicted.

Because the Set 1 empirical delta p's were higher than the planned values, the difference between Set 1 and Set 2 trials for each group was altered from what was planned. In the HC-MC group, the difference between Set 1 and Set 2 trials was increased; in the NC-MC group, the difference between Set 1 and Set 2 trials was decreased. Nonetheless, a contrast effect occurred.

The fact that outcome frequency did not remain the same across trials adds another factor that, in addition to a shift in amount of control, may have contributed to the contrast effect. Along with a shift in degree of response-outcome control across trials, there was also a parallel shift in amount of outcome frequency. Specifically, subjects in group HC-MC not only received less control in Set 2, but fewer outcomes. Likewise, subjects in group NC-MC not only received more control in Set 2, but more outcomes. So, it is unclear which factor led to the contrast effect; but, in either case, both factors (degree of control and amount of outcome frequency) are part of the context. And, importantly, the amount of outcome frequency did not differ between groups in Set 2; the situational information was the same between the two groups. Hence, Alloy and Tabachnik's model still apparently cannot account for the findings.

IV. GENERAL DISCUSSION

To review, the purpose of the present experiments was to test the predictive validity of a model (Alloy & Tabachnik, 1984) put forth to account for contingency judgments. It was argued here that while the model does improve the organization and conceptualization of this research, it does so on a very general level and devoid of any real explanatory or predictive power. It was proposed that the components of this model, independently or interactively, cannot apparently account for the effects caused by the context in which a judgment is made. Specifically, the same situational information (a set of trials) was hypothesized to be judged differently in different contexts (varying degrees of control in initial sets of trials).

This hypothesis was supported using two different paradigms. The situational information that was judged was also different in each experiment. In the first experiment, noncontingent response-outcome trials were used which, when preceded by high-contingent response-outcome trials, were judged to be lower in contingency than those preceded by low- or the same non-contingent trials. In the second experiment, medium-contingent response-outcome trials were used which were judged to have a lower degree of inherent control when preceded by high-contingent response-outcome trials than when preceded by non-contingent response-outcome trials.

As explained above, the contrast effects apparently cannot be explained or predicted by the Alloy & Tabachnik (1984) model. The results of the NC-NC group in Experiment 1, which received two sets of the same noncontingent response-outcome trials and provided

approximately the same judgments in each set, could have been predicted, but it is not clear which explanation Alloy and Tabachnik would have put forth to account for them. For example, were Set 2 judgments the same as those of Set 1 because of stronger situational information, stronger expectations, or equally strong components? This question cannot be answered since there is no reliable method to empirically assess the strength of any one of those components. Thus, any predictive or explanatory power of Alloy and Tabachnik's model is reduced.

It appears that the Alloy and Tabachnik model is not sufficient to explain or predict judgments of contingency in all cases and, in those cases which they can account for, the explanations are post-hoc.

Several alternatives are available to Alloy and Tabachnik. One, the model could be abandoned. Two, the present definition of situational information could be expanded to include the context in which that information is embedded. Three, a separate component focusing on the context could be added.

The abandonment of this model has some appeal in that it seems that the number of factors which could potentially affect judgments of contingency is enormous. Thus, it may be unreasonable to expect that these factors could be encompassed by the two proposed components. However, although the model may not be valid in all cases, as stated earlier, it does play an effective heuristic and organizational role. In a relatively young area of research, this role is welcomed.

If the model is to continue to be useful, though, two changes are necessary. First, the components, situational information and expectations, must be more clearly delineated. Alloy and Tabachnik recognize that the definition of the expectation component is not clear in regard to the specificity of expectations. The definition of situational information is less clear. As stated earlier, they need to determine the level of information processing to which they are referring as well as the boundaries of that information.

Second, the situational information component needs to be expanded or a third component needs to be adopted in order to account for context effects. It is in the clarification of situational information boundaries that the effects of context could best be introduced. Presently, the authors refer to the "current environment," which implies that they are referring to the specific situation that is being assessed rather than that information within the larger context of time and space.

As was demonstrated in the present two experiments (and in research in other areas as well), information is not judged within a vacuum. Instead, it is judged within a context. In some instances, a contrast effect may occur, as shown in the present experiments (e.g., groups HC-NC, HC-MC, and NC-MC); in other instances, also illustrated in the present experiment (e.g., LC-NC), such an effect did not occur. If situational information were defined as "that information within the broader context of time and space," then the validity of the Alloy and Tabachnik model would increase.

The Alloy and Tabachnik model would then be similar to one proposed by Manis and Paskewitz (1984) to account for judgments made within a context that is different (along a particular dimension) from the stimuli being judged. On the basis of their research on contrast effects, Manis and Paskewitz have suggested that judgments are mediated by two fairly independent "paths," one path leading to judgments based on expectations and the second one leading to judgments based on contrast effects.

Further research is still needed to determine the circumstances under which one component (or pathway) is stronger than the other and thus determines judgments. For example, Manis and Paskewitz suggest that the passage of time may be one circumstance under which expectations may be more likely than contrast effects to determine judgments.

Such changes in the Alloy and Tabachnik model would certainly increase its explanatory power. For example, the results of group HC-NC could now be explained by recognizing that NC trials were judged with the context of HC trials which led to a contrast effect. As stated above, it would still not be possible to make valid predictions since it is not possible to assess the strength of each component.

Directions for Future Research

In the area of contingency judgments, research needs to move down both macro and micro avenues. Several ideas for such research are highlighted here.

It should be clear by now that people are not always accurate in judging how much control their response has over an outcome or in judging the direction and strength of a relationship between two events. Research oriented toward specific findings has uncovered numerous factors, revolving around the situation and the person, that may affect judgments of contingency. Research oriented toward integrating these findings has not been as common or as fruitful.

In regard to research of a macro nature, more studies are needed that are directed toward integrating the factors that may affect contingency judgments. To this end, researchers must conduct more multifactored studies as well as recognize that potentially confounding variables are ubiquitous and must to be controlled. Such studies would reduce the equivocality of conclusions of studies and set the stage to more readily integrate research findings.

For example, as discussed in Section 11, it appears that outcome frequency interacts with several variables to affect judgments of contingency and, in fact, probably contributes to the contradictory findings of Jenkins and Ward (1965) and Alloy and Abramson (1979). Research to confirm this possibility is needed. The representation of variable is another factor that is not always controlled. Allan and Jenkins (1983) point out that Shaklee and Tucker (1980) ignored this factor in their study of strategies for contingency judgments; they combined the results of nine problems which had used a total of at least three different modes of representation. The use of within- and between- subject designs is also not always controlled. For example, Benassi (unpublished work) has pointed out that within- versus

between-subject design type has confounded a comparison of discrete-versus continuous-trials procedures-- two procedures which otherwise appear to yield judgments of differential accuracy.

It is recommended that each of the factors identified in Section II should be examined within a broader framework and recognized as a potentially extraneous factor.

The Alloy and Tabachnik (1984) model, notwithstanding its drawbacks, is one example of the type of theoretical macro work that can be done. Hopefully, work of this nature will stimulate other experiments, such as the present two.

In regard to the present studies, several issues of a minor nature were raised which bear further attention.

First, it would be informative to repeat Experiment 2 using the same amount of outcome frequency across sets of trials. As discussed earlier, it is unclear whether it was the context of control or of outcome frequency that led to the Set 2 judgments of control.

Second, in order to further understand the conditions under that expectations or contrast effects will have the most influence, another experiment should be conducted which manipulates the amount of time between sets of trials. If Manus and Paskewitz (1984) are correct, it may be that as this time period increased, the influence of expectations would become stronger.

Third, to address the same issue presented in the previous paragraph, an experiment in which the number of initial trials is varied might provide fruitful findings. It would seem that Alloy and Tabachnik (1984) would have to argue that expectations would positively covary with the number of Set 1 trials. Conversely, a contrast effect perspective might contend that the context would be clearer, thus increasing any contrast effects.

Fourth, previous research (Pepitone & DiNubile, 1976) has indicated that contrast effects do not occur when the initial experience (the context) is not publicly recorded. Those authors suggested that this result may stem from a lack of "ego involvement" with the task. It may also be that subjects have no initial scale value with which they wish to be consistent. In any case, another experiment might focus on the effects of recording judgments of control in the first set of trials versus not recording those judgments.

Finally, it is important to replicate the present experiments while varying the potential influential factors presented in Section II (e.g., positive vs. negative contingencies, mode of representation and presentation, etc.).

It is also important to note that more research is needed on the theory of contrast effects. As stated in Section III, this issue was not the center of the present research. Indeed, even researchers in that area cannot agree upon the origins of contrast effects. Still, if it is influencing judgments of contingency, it would be important to establish whether, for example, judgment strategies of a particular set of trials are the same when that set is embedded in a context versus

occurring alone.

As stated at the outset, the ability to accurately judge the degree of control one has over outcomes is fundamentally important in effectively functioning in this world. It is hoped that the present experiments will contribute to understanding how judgments are made by identifying factors important to the judgmental process and by helping to improve a model used to integrate findings in this area. In the future, hopefully, the results from research of this nature will be applied to improving judgments that are made in natural settings.

LIST OF REFERENCES

LIST OF REFERENCES

- Abramson, L. Y., & Alloy, L. B. (1980). Judgment of contingency: Errors and their implications. In A. Baum & J. Singer (Eds.), Advances in environmental psychology (Vol. 2). Hillsdale, New Jersey: Erlbaum.
- Abramson, L. Y., Alloy, L. B., & Kossman, D. (in press). The judgment of predictability in depressed and nondepressed students. In J. B. Overmier & R. F. Brush (Eds.), Affect, conditioning, and cognition: Essays on the determinants of behavior. Hillsdale, New Jersey: Erlbaum.
- Allan, L. G. (1980). A note on measurements of contingency between two binary variables in judgment tasks. Bulletin of the Psychonomic Science, 415, 147-149.
- Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of the response. Canadian Journal of Psychology, 34, 1-11.
- Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. Learning and Motivation, 14, 381-405.
- Abramson, L. Y., Garber, J., & Seligman, M. E. P. (1980). Learned helplessness in humans: An attributional analysis. In J. Garber & M. E. P. Seligman (Eds.), New York: Academic Press.
- Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. (1978). Learned helplessness in humans: Critique and reformulation. Journal of Abnormal Psychology, 87, 49-74.
- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? Journal of Experimental Psychology: General, 108, 441-485.
- Alloy, L. B., & Abramson, L. Y. (1980). The cognitive component of human helplessness and depression: A critical analysis. In J. Garber & M. E. P. Seligman (Eds.), Human helplessness: Theory and application. New York: Academic Press.
- Alloy, L. B., & Abramson, L. Y. (1982). Learned helplessness, depression, and the illusion of control. Journal of Personality and Social Psychology, 42, 1114-1126.
- Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. Psychological Review, 91, 112-149.

- Alloy, L. B., & Viscusi, D. (1981). Induced mood and the illusion of control. Journal of Personality and Social Psychology, 41, 1129-1140.
- Alloy, L. B., Crocker, J., & Tabachnik, N. (1980, August). Depression and covariation judgments: Expectation-based distortions in information search and recall. Paper presented at the meeting of the American Psychological Association, Montreal, Canada.
- Alloy, L. B., Seligman, M. E. P. (1979). On the cognitive component of learned helplessness and depression. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 13). New York: Academic Press.
- Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. Journal of Experimental Psychology: General, 112, 117-135.
- Beach, L. R., & Scopp, T. S. (1966). Inferences about correlations. Psychonomic Science, 6, 253-254.
- Benassi, V. A., & Mahler, H. I. M. (in press). Contingency judgments by depressed college students: Sadder but not always wiser. Journal of Personality and Social Psychology.
- Benassi, V. A., Knott, R. L., & Mahler, H. I. M. (in press). Detection of noncontingency in free-operant situation. Personality and Social Psychology Bulletin.
- Benassi, V. A., Sweeney, P. D., Drevno, G. E. (1979). Mind over matter: Perceived success at psychokinesis. Journal of Personality and Social Psychology, 37, 1377-1386.
- Beyth-Marom, R. (1982). Perception of correlation reexamined. Memory and Cognition, 10, 511-519.
- Bower, G. H. (1961). A contrast effect in differential conditioning. Journal of Experimental Psychology, 62, 196-199.
- Bruner, A., & Revusky, S. H. (1961). Collateral behavior in humans. Journal of Experimental Analysis of Behavior, 4, 349-350.
- Catania, A. C., & Cutts, D. (1963). Experimental control of superstitious responding in humans. Journal of Experimental Analysis of Behavior, 6, 203-208.
- Chapman, L. J. (1967). Illusory correlation in observational reports. Journal of Verbal Learning and Verbal Behavior, 6, 151-155.

- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. Journal of Abnormal Psychology, 72, 193-204.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. Journal of Abnormal Psychology, 74, 271-280.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? The behavioral and brain sciences, 4, 317-370.
- Collier, G., & Marx, M. H. (1959). Changes in performance as a function of shifts in the magnitude of reinforcement. Journal of Experimental Psychology, 57, 305-309.
- Coppel, D. B., & Smith, R. E. (1980). Acquisition of stimulus-outcome and response-outcome expectancies as a function of locus of control. Cognitive Therapy and Research, 4, 179-188.
- Costello, C. G. (1978). A critical review of Seligman's laboratory experiments on learned helplessness and depression in humans. Journal of Abnormal Psychology, 87, 21-31.
- Crocker, J. (1981). Judgment of covariation by social perceivers. Psychological Bulletin, 90, 272-292.
- Eiser, J. R., & Stroebe, W. (1972). Categorization and Social Support. London and New York: Academic Press.
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 85, 395-416.
- Erlick, D. E. (1966). Human estimates of statistical relatedness. Psychonomic Science, 5, 365-366.
- Erlick, D. E., & Mills, R. G. (1967). Perceptual quantification of conditional dependency. Journal of Experimental Psychology, 73, 9-14.
- Ferrell, H. J., & Shanab, M. E. (1975). Contrast effects as a function of shifts in delay of water reward. Bulletin of Psychonomic Society, 5, 417-420.
- Golding, S. L., & Rorer, L. G. (1972). Illusory correlation and subjective judgment. Journal of Abnormal Psychology, 80, 249-260.
- Hake, H. W., & Hyman, R. (1953). Perception of the

st
Jo

Hamil
an
in
Aa

Helso
an

Hiro:
he
18

Inhel
th
Bo

Jenki
cor
Mor

Jennir
cov
juc
(Ec
Can

Kahner
pre

Kelley
psy
Mot
Pre

Kellog
Psy

Kurtz,
cor
Jou

Krantz
per
abs
35-

Langer
Per:

- statistical structure of a random series of binary symbols. Journal of Experimental Psychology, 45, 64-74.
- Hamilton, D. L. A. (1979). A cognitive attributional analysis of stereotyping. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 12). New York: Academic Press.
- Helson, H. (1964). Adaptation level theory: An experimental and systematic approach to behavior. New York: Harper & Row.
- Hiroto, D. S. (1974). Locus of control and learned helplessness. Journal of Experimental Psychology, 102, 187-193.
- Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books.
- 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., & Ross, L. (1982). Informal covariation assessment: Data based versus theory based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge, England: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), Nebraska Symposium on Motivation (Vol. 15). Lincoln: University of Nebraska Press.
- Kellog, W. N. (1949). "Superstitious" behavior in animals. Psychological Review, 56, 172-175.
- Kurtz, R. M., & Garfield, S. L. (1978). Illusory correlation: A further exploration of Chapman's paradigm. Journal of Consulting and Clinical Psychology, 46, 1009-1015.
- Krantz, D. L., & Campell, D. T. (1961). Separating perception and linguistic effects of context shifts upon absolute judgments. Journal of Experimental Psychology, 62, 35-42.
- Langer, E. (1975). The illusion of control. Journal of Personality and Social Psychology, 32, 951-955.

- Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. Journal of Personality and Social Psychology, 32, 951-955.
- Lefcourt, L. (1972). Recent developments in the study of locus of control. In B. A. Maher (Ed.), Progress in experimental personality research (Vol. 6). New York: Academic Press.
- Lueger, R. J., & Petzel, T. P. (1979). Illusory correlation in clinical judgment: Effects of amount of information to be processed. Journal of Consulting and Clinical Psychology, 47, 1120-1121.
- Maier, S. F., & Seligman, M. E. P. (1976). Learned helplessness: Theory and evidence. Journal of Experimental Psychology: General, 105, 3-46.
- Manis, M. (1971). Context effects in communication. In M. H. Appley (Ed.), Adaptation-level theory. New York: Academic Press.
- Manis, M., & Paskewitz, J. R. (1984) Judging psychopathology: Expectation and contrast. Journal of Experimental Social Psychology, 20, 363-381.
- Manis, M., & Paskewitz, J. R. (1984). Specificity in Contrast Effects: Judgments of psychopathology. Journal of Experimental Social Psychology, 20, 217-230.
- Martin, D., Abramson, L. Y., & Alloy, L. B. (1984). The illusion of control for self and others in depressed college students. Journal of Personality and Social Psychology, 46, 125-136.
- Metalsky, G. I., & Abramson, L. Y. (1981). Attributional styles: Toward a framework for conceptualization and assessment. In P. C. Kendall & S. D. Hollon (Eds.), Assessment strategies for cognitive behavioral interventions. New York: Academic Press.
- Miller, W. R., & Seligman, M. E. P. (1973). Depression and the perception of reinforcement. Journal of Abnormal Psychology, 82, 62-73.
- Neunaber, D. J., Wasserman, E. A., & O'Hara, M. W. (1982). Judgment of contingency and operant responding in depressed and nondepressed college students. Unpublished manuscript.
- Nisbett, R. E., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. New Jersey: Prentice Hall.

- Oakes, W. F., & Curtis, N. (1982). Learned helplessness: Not dependent upon cognitions, attributions, or other such phenomenal experiences. Journal of Personality, 89, 110-112.
- Parducci, A. (1963). Range-frequency compromise in judgment. Psychological Monographs, 72, (2, Whole No. 565).
- Pepitone, A., & DiNubile, M. (1976). Contrast effects in judgments of crime severity and the punishment of criminal violators. Journal of Personality and Social Psychology, 33, 448-459.
- Peterson, C. R. (1980). Recognition of noncontingency. Journal of Personality and Social Psychology, 38, 727-734.
- Phares, E. J. (1976). Locus of control in personality. Morristown, New Jersey: General Learning Press.
- Postman, L., & Miller, G. A. (1945). Anchoring of temporal judgments. American Journal of Psychology, 58, 43-53.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80, (1, Whole No. 609).
- Seggie, J. (1975). The empirical observation of the Piagetian concept of correlation. Canadian Journal of Psychology, 29, 32-42.
- Seligman, M. E. P. (1975). Helplessness: On depression, development and death. San Francisco: W. H. Freeman, 1975.
- Seligman, M. E. P., Klein, D. C., & Miller, W. R. (1976). Depression. In H. Leitenberg (Ed.), Handbook of behavior modification and behavior therapy. Englewood Cliffs, New Jersey: Prentice-Hall, 1976.
- 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. (1981). Development of rule use in judgments of covariation between events. Child Development, 52, 317-325.
- Shaklee, H., & Tucker, D. A. A rule analysis of judgments of covariation between events. Memory and Cognition, 8, 459-467.
- Skinner, B. F. (1948). "Superstition" in the pigeon. Journal of Experimental Psychology, 38, 168-172.
- Smedslund, J. (1963). The concept of correlation in adults.

Scandinavian Journal of Psychology, 4, 165-173.

- Staddon, J. E. R., & Simmelhag, J. L. (1974). The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. Psychological Review, 78, 3-43.
- Starr, B. J., & Katkin, E. S. (1969). The clinician as an aberrant actuary: Illusory correlation and the sentence blank. Journal of Abnormal Psychology, 74, 670-675.
- Stern, G. S., & Berrenberg, J. L. (1979). Skill-set, success outcome, and mania as determinants of the illusion of control. Journal of Research in Personality, 13, 206-220.
- Tennen, H., Drum, P. E., Gillen, R., & Stanton, A. (1982). Learned helplessness and the detection of contingency: A direct test. Journal of Personality, 50, 426-442.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the rationality of choice. Science, 211, 453-458.
- Volkman, J. (1951). Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), Social psychology at the crossroads. New York: Harper & Row.
- Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. Canadian Journal of Psychology, 19, 231-241.
- Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans: Factors affecting judgments of response-outcome contingencies under free-operant procedures. Learning and Motivation, 14, 406-432.
- Wasserman, E. A., & Shaklee, H. (1984). Judging response-outcome relations: The role of response-outcome contingency, outcome probability, and the method of information presentation. Memory and Cognition, 12, 270-286.
- Weiner, B. (Ed.), (1974). Achievement motivation and attributional theory. Morristown, New Jersey: General Learning Press.
- Werner, Wortman, C. B. (1975). Some determinants of perceived control. Journal of Personality and Social Psychology, 31, 282-294.
- Wortman, C. B., Costanzo, R. P., & Witt, J. R. (1973). Effects of anticipated performance on the attributions of

causality to self and others. Journal of Personality and Social Psychology, 27, 372-381.

Wright, J. C. (1962). Consistency and complexity of response sequences as a function of schedules of noncontingent reward. Journal of Experimental Psychology, 63, 601-609.

APPENDICES

APPENDIX A, STATISTICAL COMPUTATIONS FOR CONTINGENCY JUDGMENTS

The degree of contingency between binary variables can be specified by any of the following coefficients: chi square (χ^2), phi (ϕ), delta d (Δd), and delta d (Δp). These techniques differ in their applicability and computational complexity, but are all based on the four cell frequencies of the 2 X 2 matrix. Each cell of this matrix reflects the frequency of one combination of the two events (e.g., response/outcome, response/no outcome, no response/outcome, no response/no outcome, representing cells a, b, c, and d, respectively).

In the statistical literature, χ^2 and ϕ are probably used more often than the other coefficients to represent the degree of contingency between two binary variables (Allan, 1980). These, however, are mathematically cumbersome and have not been traditionally used in psychological judgment research.

Delta d (also known as the "sum of the diagonals") was used in some of the early judgment studies, most notably those conducted by Inhelder and Piaget (1958). This statistic is obtained by comparing the difference between the sum of the confirming cases and the sum of the nonconfirming cases. Jenkins and Ward (1965) and Ward and Jenkins (1965) have stated that delta d is a valid index of contingency only when the two row frequency totals and/or the two column frequency totals are equal. If this requirement is not met, delta d may reveal a nonzero value even when there is no contingency present. If this requirement is met, delta d is an accurate assessment of contingency.

Delta p reflects the difference between the conditional probabilities of an outcome given each response. Unlike delta d , it does not require equal marginal frequencies. It is less mathematically cumbersome than X and o and yet it is highly correlated with the latter coefficients (see Allan, 1980; Alloy & Abramson, 1979). As stated in the text, it has been accepted by most researchers in this area as the statistic of choice (e.g., Alloy & Abramson, 1979; Allan & Jenkins, 1983; Jenkins & Ward, 1965; Ward & Jenkins, 1965) and was employed in the present experiments.

Finally, it should be noted that while these statistical computations provide the most accurate estimates of contingency, they are not necessarily employed by most individuals. In fact, research has revealed that 33-38% of college student samples use the conditional probability rule (Shaklee & Mims, 1981; Shaklee & Tucker, 1980). But, prior to the development of logical reasoning (Inhelder & Piaget, 1958) or when the stimulus situation becomes complex (e.g., when memory demands are increased), subjects tend to use simpler but less accurate methods of estimation (Arkes & Harkness, 1983; Shaklee & Mims, 1982).

Two such methods are referred to as the "cell a" and the "a-versus-b" strategies. The former, which is the least sophisticated and most error-producing strategy, involves basing judgment on the frequency of occurrences in cell a. This rule was identified by Smedslund (1963) and Nisbett and Ross (1980) as being the most commonly employed strategy of adults, but this contention has been disconfirmed (e.g., Shaklee & Tucker, 1980). The a-versus-b strategy entails comparing the frequency of entries in cell a with those in cell b.

Because it incorporates somewhat more information, it is less error-prone than the cell approach, but it, too, generally yields inaccurate estimates.

APPENDIX B, QUESTIONNAIRE

1. Please write the total number of times you believe the blue light appeared, regardless of whether you pressed or did not press the button. (Your answer should be between 0 and 40)

2. Please write the number of times that you believe you pressed the button and the blue light appeared. (0-40)

3. Please write the number of times that you did not press the button and the blue light did appear. (0-40)

4. Please write the number of times that you believe that you pressed the button and the blue light did not appear. (0-40)

5. Please write the number of times that you believe you did not press the button and the blue light did not appear. (0-40)

6. Do you feel that your responses (pressing and not pressing the button) are related to the outcome (appearance of the blue light)?

_____ No (go to question b)
 _____ Yes (go to question a)

- a. Please describe the evidence that convinced you that a relationship existed.

- b. Please describe the evidence that convinced you that there was no relationship.

7. Did you use your experience with the first set of trials to help you judge the second set of trials?

_____ No _____ Yes. Please describe how you used that information.