

DTIC FILE COPY

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

AD-A225 101

1. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS --	
2. SECURITY CLASSIFICATION AUTHORITY --			3. DISTRIBUTION / AVAILABILITY OF REPORT For Announcement Purposes Only	
3. DECLASSIFICATION / DOWNGRADING SCHEDULE --			5. MONITORING ORGANIZATION REPORT NUMBER(S) A 90-01	
PERFORMING ORGANIZATION REPORT NUMBER(S) --			7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences	
4. NAME OF PERFORMING ORGANIZATION L.L. Thurstone Psychometric Laboratory		6b. OFFICE SYMBOL (If applicable) --	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	
c. ADDRESS (City, State, and ZIP Code) University of North Carolina Chapel Hill, NC 27514			9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-83-K-0347	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 7a.		8b. OFFICE SYMBOL (If applicable) PERI-BR	10. SOURCE OF FUNDING NUMBERS	
8c. ADDRESS (City, State, and ZIP Code) 7b.			PROGRAM ELEMENT NO. 6.11.01B	TASK NO. 74B
			WORK UNIT ACCESSION NO. n/a	WORK UNIT ACCESSION NO. n/a
11. TITLE (Include Security Classification) The Effect of Outcome Desirability on Comparisons of Numerical and Linguistic Probabilities				
12. PERSONAL AUTHOR(S) Brent L. Cohen				
13a. TYPE OF REPORT Announcement		13b. TIME COVERED FROM 85/11 TO 86/11	14. DATE OF REPORT (Year, Month, Day) 1990, July	15. PAGE COUNT 66
16. SUPPLEMENTARY NOTATION Michael Drillings, contracting officer's representative Availability: from L.L. Thurstone Psychometric Laboratory, University of North Carolina.				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	Probability expressions; Subjective probability; Outcome desirability; Vagueness, etc.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This thesis examines the observed phenomenon that when subjects have complete knowledge of a probability distribution, subjective estimates are found to be biased as a function of outcome desirability. The thesis extends that finding to situations in which subjects know the probabilities based only on linguistic expressions. It was found that subjects' comparisons of numerical and linguistic probabilities were biased in various degrees toward the more desirable outcome, based on the particular phrase under consideration. Keywords:				
DISTRIBUTION STATEMENT A Approved for public release; Distribution is unlimited S JUL 30 1990 B D				
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED / UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Drillings			22b. TELEPHONE (Include Area Code) 202/274-8722	22c. OFFICE SYMBOL PERI-BR

90 07 23 092

THE EFFECT OF OUTCOME DESIRABILITY
ON COMPARISONS OF NUMERICAL AND LINGUISTIC PROBABILITIES

by

Brent L. Cohen

A Thesis submitted to the faculty of The University
of North Carolina at Chapel Hill in partial
fulfillment of the requirements for the degree of
Master of Arts in the Department of Psychology.

Chapel Hill

1986

Approved by:

Thomas S. Wallsten
Advisor

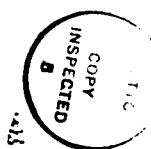
Harvey L. Lumsden
Reader

Samuel T. Little
Reader

BRENT L. COHEN. The Effect of Outcome Desirability on
Comparisons of Numerical and Linguistic Probabilities
(Under the direction of THOMAS S. WALLSTEN.)

ABSTRACT

When subjects have complete knowledge of the probability distribution governing a set of outcomes, their subjective probability estimates have been found to be biased towards the outcome with the relatively favorable desirability. With the use of linguistic probability expressions, this study attempted to extend this finding to situations in which subjects have only vague knowledge or no knowledge of the probability distribution relevant to a set of outcomes. It was found that subjects' comparisons were biased, in varying degrees depending on the particular phrase under consideration, towards the relatively more desirable outcome. This finding indicates that a complete model of judgment under uncertainty must include a factor for context effects such as outcome desirability.



Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By <i>per Form 50</i>	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
<i>A-1</i>	<i>21</i>

ACKNOWLEDGEMENTS

This research was supported by contract MDA903-83-K-0347 from the U.S. Army Research Institute. The views, opinions, and findings contained in this paper are those of the author and should not be construed as official Department of the Army position, policy, or decision.

I would like to thank my advisor, Dr. Thomas S. Wallsten, for all of his advice and encouragement to me as I worked on this project. His comments and constructive criticisms on the earlier drafts of this paper have greatly improved the quality of the final product. Additionally, the financial support from his research grant made it possible to pay subjects and perform the computer analyses. Also, the other members of my committee, Dr. Samuel Fillenbaum and Dr. Marcy Lansman, gave many helpful suggestions and aided my thinking on this project. I also want to thank my fellow students in the Psychometric Lab who helped me with the chore of data management.

A great debt of thanks is due my parents who constantly encouraged me to finish. Finally, I must thank Marcy Freund, who was a great inspiration and support to me.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	111
LIST OF TABLES	v
LIST OF FIGURES	vi
INTRODUCTION	1
LITERATURE REVIEW	3
PURPOSE AND HYPOTHESES	14
METHOD	16
RESULTS	20
DISCUSSION	27
SUMMARY AND CONCLUSIONS	40
REFERENCES	42
APPENDIX	58

LIST OF TABLES

Table	Page
1. Test-retest correlations	47
2. Desirability effects in terms of mean response values for each phrase and combined over similar phrases	48
3. F-ratios from ANOVAs on individual probability phrases	49
4. F-ratios from ANOVAs on combined probability phrases	50
5. Tests of whether mean responses are significantly below 100 for the neutral phrases vs. 0.50	51
6. F-ratios from ANOVAs for subjects who understood/ did not understand the instructions	52
7. Monotonic (M), single-peaked (SP), and other (O) membership functions for each phrase and desirability level, and averaged over all desirability levels	53

LIST OF FIGURES

Figure	Page
1. Graph of mean responses as a function of probability for individual phrases, ignoring desirability	54
2. Graph of mean responses as a function of probability for phrase pairs, ignoring desirability	55
3. Types of membership functions	56
4. Hypothesized effect of independent outcome desirability on membership functions	57

Introduction

Earlier theories of judgment under uncertainty (e.g. Knight, 1921; Savage, 1954) have assumed that the decision-maker is a rational being, conforming to basic axioms of probability. This study examines an aspect of irrational decision-making; namely, the ways in which judgments about uncertain outcomes are affected by the independent desirability of those outcomes. The independent desirability of an outcome refers to the value of the outcome itself, independently of one's prediction. This study extends the finding (Marks, 1951; Irwin & Snodgrass, 1966; Irwin & Metzger, 1966; Slovic, 1966) that independent outcome desirability affects subjective probability, in direct opposition to a fundamental axiom of rationality.

The work of Irwin, et al., on independent outcome desirability was motivated in part from Irwin's interest in the types of factors affecting people's expectations of outcomes. He held that there were cognitive factors, which would lead to realistic expectations, as well as affective factors, which would lead to unrealistic expectations (Irwin, 1944). Underlying this research may have been Irwin's intent to develop a more complete model of judgment than the subjectively expected utility (SEU) model. The results found by Irwin, et al. indicate that, at the least,

it may be necessary to control for independent outcome desirability in SEU models.

Uncertainty is a broad concept, including all states of knowledge other than perfect about the occurrence of events. As an attempt to segment the construct of uncertainty, Ellsberg (1961) described three states of knowledge regarding the probabilities of future events. He labeled these conditions "risk", "ambiguity", and "complete ignorance", representing precise knowledge of probabilities, vague or imprecise knowledge of probabilities, and absolutely no information about probabilities, respectively. If the probabilities are known to be either exactly 0 or exactly 1, however, this corresponds to a situation of certainty.

As will be documented below, it is already known that under conditions of risk, subjective probabilities are positively related to independent outcome desirability. However, people commonly make decisions based on information that is ambiguous or imprecise, or even on the basis of no information at all, which corresponds to Ellsberg's states of ambiguity and complete ignorance, respectively. This study examines the effect of independent outcome desirability on judgments under ambiguity and complete ignorance.

There are at least two major ways to express uncertainty in an ambiguous, vague, or imprecise manner. One way is to use a range of numerical probability values to describe the probability of an outcome, while the other

is to use verbal expressions of uncertainty. The latter was used in the present study, in order to further investigate the factors that affect the meanings of probability terms.

Literature Review

Rationality, Uncertainty, and the Ellsberg Paradox

For people to be rational decision-makers, they must conform to certain basic axioms of rationality. One of these axioms holds that choices are completely ordered according to the judged values of the consequences of those choices. Another important axiom is the sure-thing principle (Savage, 1954), according to which, people's ordering of alternative choices should be unaffected by adding a constant to the value of the consequences of each choice. A third principle requires people to be free of "wishful thinking" (Radner, 1972). That is, one's ranking of probabilities of events should be independent of the consequences of those events.

These axioms of rationality are not predictive of actual behavior, however. Ellsberg (1961) has shown that certain commonly obtained choice patterns are in violation of the first two of these axioms. The present study is designed to examine violations of all three of these axioms, but the primary hypothesis of this experiment is that decision-makers will consistently violate the third principle listed above.

According to Ellsberg, a condition of risk exists when a decision-maker knows with certainty the probability distribution over a set of outcomes. A state of complete ignorance exists when the decision-maker has absolutely no information about the probability distribution over outcomes. Ellsberg showed that a person's ordering of gambles depends upon whether that person is operating under a condition of risk or a condition of ignorance.

Between the extremes of risk and complete ignorance lies the state of ambiguity, in which the decision-maker can rule out at least one probability distribution. However, under ambiguity, more than one probability distribution remains under consideration, and the decision-maker does not have enough information to determine which probability distribution is operative. One of the purposes of this study is to extend Ellsberg's results to situations of ambiguity represented by probability phrases such as likely, doubtful, good chance, and so forth.

The importance of distinguishing the three types of uncertainty becomes apparent in light of the Ellsberg paradox. This paradox can be described by considering a situation in which there exist two urns containing 100 red and black balls each. The proportion of red and black balls in Urn I is completely unknown, whereas, Urn II contains exactly 50 red balls and 50 black balls. Thus, Urn I corresponds to a state of nearly complete ignorance, while Urn II corresponds to a state of risk. Let RedI, BlackI, RedII, and BlackII represent the event of drawing a

red ball from Urn I, a black ball from Urn I, a red ball from Urn II, and a black ball from Urn II, respectively. Then, the paradox lies in typical answers to the following four questions:

1. Which do you prefer to bet on, RedI or BlackI, or are you indifferent?
2. Which do you prefer to bet on, RedII or BlackII, or are you indifferent?
3. Which do you prefer to bet on, RedI or RedII, or are you indifferent?
4. Which do you prefer to bet on, BlackI or BlackII, or are you indifferent?

Most people answer the first two questions with indifference responses, implying that they feel that $\text{Pr}(\text{RedI}) = \text{Pr}(\text{BlackI})$ and $\text{Pr}(\text{RedII}) = \text{Pr}(\text{BlackII})$. Since $\text{Pr}(\text{RedI}) + \text{Pr}(\text{BlackI}) = 1$ and $\text{Pr}(\text{RedII}) + \text{Pr}(\text{BlackII}) = 1$, these indifference responses further imply that $\text{Pr}(\text{RedI}) = \text{Pr}(\text{RedII}) = \text{Pr}(\text{BlackI}) = \text{Pr}(\text{BlackII}) = 0.50$. A paradox occurs because most people will then respond to questions (3) and (4) with RedII and BlackII, respectively. The response of RedII in question (3) implies that $\text{Pr}(\text{RedII}) > \text{Pr}(\text{RedI})$, from which it logically follows that $\text{Pr}(\text{BlackII}) < \text{Pr}(\text{BlackI})$, due to the fact that $\text{Pr}(\text{RedI}) + \text{Pr}(\text{BlackI}) = 1$. However, this is contradicted by the choice of BlackII in question (4), which implies that $\text{Pr}(\text{BlackII}) > \text{Pr}(\text{BlackI})$. This pattern of responses violates the first two axioms of rationality described above.

From the above discussion, it appears that people prefer to bet on events with known rather than unknown probabilities. Such betting preferences may depend on how probabilities of events are judged under ambiguity, but they may also depend on an aversion to gambling under conditions of ignorance. For research purposes, it is useful to employ a technique that distinguishes judgments under uncertainty from avoidance of ambiguity in gambling scenarios. In the present study, subjects responded by indicating their degree of confidence in their judgment that one event is more likely to occur than is another event. Other studies have examined aspects of the Ellsberg paradox (Fellner, 1961; Becker & Brownson, 1964; Yates & Zukowski, 1976; Einhorn & Hogarth, 1985; Einhorn & Hogarth, in press). In these studies, however, the values of outcomes were dependent upon subjects' responses, as opposed to the present study's use of independent outcome values.

Ambiguity and Probability Terms

One can describe the uncertainty of an outcome in an ambiguous fashion by defining a range of numerical probability values relating to the event in question. An alternative approach toward ambiguity can be found in the use of probability words or phrases. Probability terms, such as likely, doubtful, good chance, etc., are generally used by people to describe uncertain events.

There are many possible reasons that people frequently prefer to use non-numerical estimates of probabilities.

One reason suggested by Wallsten, Budescu, Rapoport, Zwick, and Forsyth (in press) is that people prefer to express imprecise opinions in imprecise terms, such as probability phrases. Secondly, Zimmer (1983) held that most people deal more effectively with verbal than with numerical expressions of uncertainty. Note that children learn how to talk long before they learn how to perform arithmetic and calculate percentages. Additionally, Zimmer noted that a mathematical treatment of probability was not conceived until 1660, while verbal expressions of uncertainty existed in most languages for a long time before that. Another reason for the common use of verbal expressions of uncertainty may stem from people's desire to use a verbal rather than numerical style of written and spoken communication. There are many real-world events for which a numerical probability estimate would seem less appropriate than a verbal probability expression. For example, it seems much more sensible to say "it was possible but not likely", rather than "there was a 10-15% chance" with reference to the event that Shakespeare was thinking of Ann Hathaway when he wrote his twelfth sonnet.

Beyth-Marom (1982) suggested other reasons for the use of non-numerical probability estimates. First, they prevent one's estimate from being judged. Secondly, they allow one to avoid making explicit derivations to defend the estimate. Thirdly, many people believe that a probability phrase is no more ambiguous than is a range of

numerical probability values. Because most people use probability phrases to express degrees of uncertainty, it is important to study the factors that affect the meanings of these probability terms.

Measurement of Vagueness

A problem with studying ambiguity in verbal rather than numerical expressions is that the measurement of ambiguity is not as straightforward. It involves the formulation of a reasonable means of transforming verbal expressions of uncertainty into measurable quantities. Although it will not be necessary to actually measure the meanings of terms used in this study, it is still appropriate to discuss measurement techniques for two reasons. First, prior measurement of each of the terms used in this study (except for the term, unspecified) was required to determine important parts of the experimental design. Secondly, the present research may provide knowledge about factors that should be considered when measuring the meanings of vague terms.

One approach toward the measurement of the meanings of probability words involves gathering numerical probability estimates for probability phrases from a large number of subjects. From these data, information about the central probability value and the variability in probability values for each term can be determined. This technique has been used often in previous research on the meanings of probability or frequency terms (Simpson, 1944, 1963; Stone & Johnson, 1959; Lichsteinstein & Newman, 1967; Hake, 1968;

Beyth-Marom, 1982). All of these studies found considerable variability over subjects in the assignment of numbers to probability phrases. However, the probability phrases did have fairly consistent numerical meanings to individual subjects over time (Beyth-Marom, 1982). A problem with this technique is that it does not address the issue of whether the probability phrases, themselves, have vague meanings to individuals, or whether people merely differ in their interpretation of probability values (Budescu & Wallsten, 1985).

In other research, it has been suggested that the meaning of a probability term can be represented by a membership function on the $[0,1]$ interval (Zadeh, 1978; Watson, Weiss & Donnell, 1979). A membership function indicates the degree to which an event, object, or value "fits" into a vague description or category. In the case of probability terms, a membership function would show how well each probability value is described by a particular term. These membership functions can take on any shape whatsoever, but continuous single-peaked or monotonic functions are the easiest to analyze and make the most sense conceptually.

The specific numerical probability value (or values) that are best described by a particular phrase would take on the maximum value of the membership function (usually one). Those probability values that are not at all described by a probability phrase would take on the minimum

value of the function (usually zero). Other probabilities with intermediate degrees of membership in the concept represented by the phrase would generally take on intermediate values between zero and one in this function.

Wallsten, et al, (in press) utilized a modified pair-comparison procedure to measure the vagueness of probability phrases. In this procedure, subjects judged pairs of probability values with respect to how much better one rather than the other was described by each of various probability terms. If the properties of the resulting data satisfy the axioms of an additive difference structure (Krantz, Suppes, Luce, & Tversky, 1971), then it is legitimate to use certain scaling procedures to derive the membership function for each probability term. These scaling techniques allow one to make predictions about subjects' responses in the pair-comparison task. It was found that observed and predicted responses were correlated in the range of 0.75 to 0.77. Independent judgments in a related modified pair-comparison task were also found to be predicted by the responses on the first task. Further, responses were found to be consistent for subjects over time (average correlation of 0.88 between sessions). The shapes and locations of the membership functions gathered in that study revealed large individual differences, however.

Factors Affecting the Meanings of Probability Terms

Environmental factors may play a large role in people's interpretations of probability phrases. The

research problem could be expressed in terms of determining the individual difference and context factors that affect the judged meanings of probability terms, regardless of the way in which the meanings are derived. Factors that might affect the meaning of a probability phrase to an individual include, among other things: (1) the reliability of the information source(s); (2) the number of sources of information; (3) the number of words in one's vocabulary of probability phrases; (4) general semantic context; (5) the importance of the event being analyzed; (6) the base rate of occurrence of the event. The effects of some of these context factors on probability or frequency phrases have been the focus of previous research (Pepper & Prytulak, 1974; Beyth-Marom, 1982; Wallsten, Fillenbaum & Cox, 1986).

This study is primarily concerned with a specific context effect -- the desirability of the outcomes of the event being analyzed by the decision-maker. Desirability is somewhat related to the importance of the event, although outcomes can have differing desirabilities while importance is held constant. It seems reasonable to assume that a continuum of desirability exists at each level of importance of events. For example, both flunking a test and nuclear war could be rated by a subject as highly undesirable, but certainly the latter event is more important. Importance was held constant in this study, while desirability was manipulated by the winning or losing

1 of points on each trial.

Outcome Desirability

The effects of outcome desirability on subjective probabilities under conditions of risk have been studied in depth (Marks, 1951; Scheibe, 1964; Morlock & Hertz, 1964; Pruitt & Hoge, 1965; Irwin & Snodgrass, 1966; Irwin & Metzger, 1966; Slovic, 1966; Irwin & Graess, 1968). In these studies, subjects' choices or estimates of the likelihoods of events were positively related to the independent values of the events. The independent outcome values (IOs) are to be distinguished from the dependent outcome values (DOs), which depend on the accuracy of the person's prediction.

Irwin & Snodgrass (1966) gauged subjective probabilities by means of a betting procedure, in which the frequency of subjects' positive bets was the dependent variable. Subjects made bets within a prescribed range on whether or not a marked card would be drawn from a deck with a known number of marked and unmarked cards. In addition to winning or losing the bet, subjects won or lost money if a marked card was drawn, regardless of their prediction. The bet thus determined the DO and the additional winnings or losses attributable to the drawing of a marked card became the IO. It was found that positive IOs led to a higher frequency of positive bets than did negative IOs. Further, this effect was strongest for objective probabilities (proportions of marked cards) closest to 0.50. In a related study, Irwin and Metzger

(1966) found that the positive relationship between IO and estimated likelihood was maintained even when IO was probabilistic. In this study, the IOs had well-defined positive or negative expected values, but on any given trial subjects did not know with certainty whether they would win or lose money for drawing a marked card.

A slightly different paradigm was utilized by Slovic (1966) to test the effect of desirability on judged likelihood. In his study, subjects gave direct numerical estimates of subjective probabilities under conditions of varying IO. Again, it was found that estimated likelihoods of events were positively related to desirabilities of IOs. Slovic termed this effect "partial optimism" because it related only to the sign of the IO, not to its magnitude. Also, Irwin & Graee (1968) were unable to find a consistent significant relationship between IO and choice probabilities for paired events when the IOs for the two events under consideration had the same sign, but different magnitudes. There was, however, a significant effect of IO magnitude at some levels of objective probability. Crandall, Solomon, & Kellaway (1955) found a significant overall interaction between IO and choice probabilities when both IOs had the same sign. When this effect was analyzed at each level of objective probability, however, it was not significant.

In all cases described above, the events for which subjects estimated probabilities were precisely defined.

14

)

That is, the information on which subjects based their subjective probabilities was either full knowledge of the probability distribution of events or sample information relating to a specific probability distribution of events. Thus, subjects were responding under conditions of risk, as described by Ellsberg. If subjects had had only vague or ambiguous information on the distribution of outcomes, or perhaps no information at all (creating an Ellsberg paradox situation), their subjective probabilities might have been affected differently by IO desirability.

Purpose and Hypotheses

The main purpose of this experiment was to examine the effect of independent outcome desirability on judgments made under conditions of ambiguity and ignorance, as represented by verbal expressions. A secondary goal of the experiment was to determine if different probability phrases show different effects due to IO desirability.

)

On every trial, subjects were given information about the likelihood of occurrence of two outcomes. Each outcome had a certain IO associated with it. For one outcome, subjects were given objective probability information, and for the other outcome, they were provided either ambiguous information (a probability term) or no information at all. The latter case represented a condition similar to the Ellsberg paradox. Subjects then indicated their degree of confidence that one outcome was more likely to occur than the other outcome. It was hypothesized that subjects'

responses in this experiment would be positively biased toward the outcome with the relatively favorable desirability factor. Further, this bias should be strongest in cases where an objective probability of 0.50 was matched with the probability terms, tossup and unspecified (the Ellsberg paradox condition), according to the results of Irwin & Snodgrass (1966). Thus, when the objective probability of 0.50 has associated with it a negative desirability and the probability phrase, unspecified, has associated with it a positive desirability, the verbal expression should be interpreted with a subjective value greater than that accorded the objective probability. This would be a contradiction of Ellsberg's contention that people's intuition usually leads them to accord a higher subjective value to an objective probability of 0.50 (risk) than to an event for which people were completely ignorant of probabilities. This bias due to desirability could be construed as: (1) a change in the judged meaning of a given probability phrase (perhaps a change in the shape or location of the membership function for that phrase), (2) a change in the subjective value accorded the objective value of 0.50, (3) a type of ambiguity aversion, or (4) some combination of the above.

In this experiment, one could have asked subjects to bet on either the risky event or on the event with the vaguely specified probability, while manipulating the desirability factor. The drawback to that design, however,

is that any effects due to desirability might be confounded with subjects' desire to avoid ambiguity in a gambling scenario. The present design controls for this factor by asking subjects to give their best judgments of the relative likelihoods of the two paired events, the consequences of which are independent of their judgments.

Method

Subjects

Twenty graduate students recruited with notices distributed in the social science departments, the Business School, and the School of Journalism at the University of North Carolina at Chapel Hill served as paid subjects. At the end of the experiment, a cash bonus of \$10 was paid to those ten people whose point totals were the highest.

Procedure

Subjects performed the experiment individually while seated in a small cubicle. Stimuli were presented on the color monitor of an IBM PC, and responses were made by means of a joystick connected to the computer.

The computer screen was, in effect, split in half from top to bottom. One side of the screen showed a probability spinner with a radial portion of the spinner colored white and the remainder colored red. Subjects were informed that the displayed proportion of white in the spinner represented the true probability of that spinner's landing on white.

On the other side of the screen, subjects received information relating to the likelihood of another spinner's landing on white, but were not able to actually view this other spinner. Instead, they saw one of six phrases. There were five probability terms (doubtful, improbable, tossup, likely, or good chance) which were each followed by the phrase, "... to land on white." These phrases described the likelihood of that spinner's landing on white. (Hereafter, this will be referred to as the invisible spinner.) In one condition, subjects saw the phrase, "Unspecified to land on white," which they understood to mean that absolutely no information was available about the invisible spinner. Except for the case involving the word, unspecified, the actual probability distribution that governed the likelihood of an invisible spinner's landing on white was represented by a single probability value that was based on previous research on the measurement of the meanings of probability terms (Wallsten, et al, in press). The probability terms used in this experiment were assigned probability values as follows: doubtful: 0.18; improbable: 0.32; tossup: 0.50; good chance: 0.64; and likely: 0.78. Unspecified was assigned a probability value of 0.50, based on the reasonable assumption that complete ignorance of probability information should lead one to assume an attitude of indifference between any two mutually exclusive and exhaustive outcomes.

On a single trial, a probability spinner appeared on a

randomly selected side of the screen, and one of the six phrases appeared on the other side. The proportion of white in the visible spinner was varied among four values uniquely predetermined for each of the six phrases with which it was matched. The phrases "likely to land on white" and "good chance to land on white" were matched with visible spinners that had proportions of white of 0.45, 0.62, 0.79, and 0.95. Phrases "doubtful to land on white" and "improbable to land on white" were matched with visible spinners that had proportions of white of 0.05, 0.22, 0.39, and 0.55. The phrases "tossup to land on white" and "unspecified to land on white" were matched with visible spinners with proportions of white of 0.30, 0.50, 0.50, and 0.70.

To incorporate the concept of outcome desirability in the experiment, each spinner's landing on white had associated with it an independent point value. These values were either +500, -500, or 0 points. In all cases, the value of a spinner's landing on red was 0 points. There were three different conditions of desirability presented. One condition involved a desirability factor of +500 points associated with the visible spinner and -500 points associated with the probability phrase. In the second condition the signs of these desirability factors were reversed. The third condition had desirability factors of 0 points associated with both the visible spinner and the probability phrase.

At the end of the experiment, subjects answered a few questions concerning their response strategies and their understanding of the instructions. This manipulation check consisted of a brief written multiple choice questionnaire.

The sequence of events on a single trial was as follows:

1. A randomly selected screen was displayed to the subject. On one side of the screen was a spinner and the desirability value associated with that spinner's landing on white. On the other side of the screen was one of the six probability phrases referring to the likelihood of an invisible spinner's landing on white and the desirability value associated with that spinner's landing on white. Traversing the width of the screen at the bottom was a line segment with an arrow positioned in the middle.

2. The subject indicated which spinner (visible or invisible) he or she judged to be more likely to land on white by using a joystick to position the arrow along the line segment. Moving the arrow toward one or the other end of the line segment indicated stronger confidence in the judgment that the spinner on that side of the screen would land on white. The arrow could be positioned at any one of 200 points along the line segment, but there were only three hash marks visible to the subject--one at each end and one in the middle of the line segment.

3. After the arrow was positioned, the subject pushed a button on the joystick assembly that caused the spinners to "spin". The subject did not actually see the spinners

) spinning and was not told the outcome (white or red) of an individual spinner. However, a noise effect simulated spinning while the word "spinning" appeared at the top of each side of the screen. The outcome of each spinner was based on the probability distributions described above.

4. If either or both spinners landed on white, the corresponding desirability values were added to or subtracted from the subject's cumulative total of points. This total was displayed and updated on every trial in a window near the top of the screen

5. The subject pushed a button on the joystick assembly to initiate the next trial.

Design

There were six probability phrases matched with four different probability spinners each. Also, each spinner-phrase pair had associated with it one of three conditions of desirability. This created a total of 72 trials which constituted a block. Subjects performed one block of trials in random order on each of two days. Sessions were separated by 24 to 48 hours. There was no practice, but subjects read detailed instructions while viewing the first randomly selected screen of their session.

Results

Reliability

) The first step was to assess the reliability of the data by calculating the linear correlation between each subject's responses over the two sessions. Table 1 shows

both a simple Pearson correlation of each subject's responses on the two days, as well as a pooled correlation. Pooling was achieved by standardizing the responses so that they had a mean of zero for each of the six probability phrases. As can be seen in the table, this adjustment did not significantly change the correlations. The mean correlation over all subjects between responses for session one and session two was 0.61 for both the simple and pooled cases. The simple correlations were above 0.80 for four subjects and above 0.70 for ten subjects.

Analyses on Individual Phrases

The three main factors in the experimental design for each of the six probability phrases were (a) the probability of the visible spinner landing on white (four levels), (b) the outcome desirability combination (three levels), and (c) session (two levels). Figure 1 shows mean responses for each phrase as a function of probability, ignoring desirability level. Responses are scaled so that 0 means certain that the invisible spinner is more likely to land on white, 200 means certain that the visible spinner is more likely to land on white, and intermediate values designate intermediate degrees of certainty. As expected, responses increase with visible spinner probability. Mean responses at each desirability condition for each phrase, ignoring probability level, are shown as the first six rows of Table 2.

The main analyses sought to determine the effect, if

any, of the desirability values on subjects' responses. The initial analyses involved six distinct $4 \times 3 \times 2$ (Probability x Desirability x Session) repeated measures ANOVAs, one for each probability phrase. The ANOVAs were performed in the manner recommended by Winer (1971), in which the error term for each effect was the interaction of that effect with the factor of subjects.

Subjects' responses, scaled such that a higher value of the response indicated a stronger confidence that the visible spinner would land on white, was the dependent variable. It was hypothesized that mean responses would increase as the level of spinner probability increased, as they do (see Figure 1). No other response pattern would be appropriate if the subjects believed that the spinners were fair. The main hypothesis of the experiment, however, was that mean response values would be lowest for the negative spinner/positive phrase desirabilities, and greatest for the positive spinner/negative phrase desirabilities.

The results of these analyses are shown in Table 3. Session was not a significant factor for any of the six probability phrases. As expected, the probability of the visible spinner landing on white was significant in the appropriate direction for all six probability phrases. From Tables 2 and 3, it can be seen that level of desirability was significant in the hypothesized direction for the probability phrases, doubtful, good chance, and likely. It was significant, but with somewhat unexpected mean response values, for the probability phrase,

improbable. Desirability was not significant, but had mean response values in the hypothesized direction for the probability phrases, tossup and unspecified.

Analyses on Phrase Pairs

In order to increase power, the phrases were combined into three pairs--high probability (good chance and likely), middle probability (tossup and unspecified), and low probability (doubtful and improbable)--and a 2 x 4 x 3 (Phrase x Probability x Desirability) ANOVA was performed on each pair. There was no hypothesized pattern of mean responses for the probability phrase factor, so it was not expected to be significant. The mean responses for each phrase pair at each level of desirability, ignoring probability level, are shown as the last three rows of Table 2. Mean responses for each phrase pair as a function of probability, ignoring desirability level, are plotted in Figure 2.

The results of these ANOVAs, shown in Table 4, reveal that probability phrase was not significant in any of the three ANOVAs, while the probability of the visible spinner landing on white was significant in the hypothesized direction in every case. Desirability was significant with the predicted response pattern for both the high and the low probability phrases, but not for the middle probability phrases. The interaction of probability with desirability was not significant in any of these analyses.

As can be seen in Figures 1 and 2, the probability

factor had a generally large effect. In contrast, however, when the IO desirability factor was significant, the effect was rather small, as shown in Table 2. In general, the high probability terms, good chance and likely, showed much larger effects of both desirability and objective probability than did the low probability terms, doubtful and improbable. The middle probability terms obviously showed small effects of desirability, since they were non-significant, but showed rather large effects of objective probability.

The Ellsberg Paradox

Ellsberg (1961) claimed that people commonly violate a fundamental axiom of rationality by preferring to bet on outcomes stated under conditions of risk rather than on outcomes stated under conditions of ignorance or high levels of ambiguity, even though the best estimate of the probabilities of all the outcomes under consideration was 0.50. This finding, known as the Ellsberg paradox, would be obtained in the present study if subjects accorded a higher subjective probability to the objective probability of 0.50 than to the phrases, tossup and unspecified, or in other words if subjects gave mean responses significantly above the neutral response of 100 when these two phrases were each matched with a visible spinner with a 0.50 probability of landing on white. In contrast, a major hypothesis of this study was that independent outcome desirability biases the subjective probability toward the outcome with the relatively favorable desirability, thereby

overwhelming any effects due to ambiguity avoidance.

Simple t tests, comparing the appropriate mean responses to 100, are suitable for testing this hypothesis. The results of this analysis, shown in Table 5, reveal that only with the phrase, tossup, under negative spinner/positive phrase desirability, was the mean response significantly below 100. Thus, in this instance, IO desirability appeared to produce a bias toward the outcome with the relatively favorable desirability and the results predicted from the Ellsberg paradox failed to hold. Notably, however, in no case were the results expected on the basis of those from Ellsberg obtained at a significant level.

Analysis of the Questionnaire Data

The results of the questionnaire given to subjects at the end of the experiment showed that eighteen out of twenty subjects understood that the spinners were fair and that their responses had no effect on the probability of either the visible or invisible spinner landing on white. Six questions, phrased like the first one in the Appendix, but with a slightly different situation in each case, were used to assess subjects' understanding of this point. Four questions, phrased like the second one in the Appendix, revealed that nine subjects out of the twenty thought that their responses had some effect on whether or not they won points. Additionally, a final question, which was identical to the third question in the Appendix, verified subjects' understanding of the two main ideas of the

instructions, and explicitly asked subjects to explain their response strategy when presented with the probability term, unspecified.

Given the above results, it was decided to perform separate analyses on the eleven subjects who seemed to understand the instructions perfectly and on the nine who seemed not to. The $2 \times 4 \times 3$ ANOVAs, as described above for the three pairs of probability phrases, were performed separately for the two groups of subjects, for a total of six analyses, with results shown in Table 6.

Probability phrase was significant only for the phrase pair, tossup and unspecified, for those who understood the instructions, while the probability of the visible spinner landing on white was significant in the predicted direction in every case. Desirability was significant in the hypothesized direction for the high probability phrase pair, good chance and likely, for both groups of subjects. For the low probability phrase pair, doubtful and improbable, desirability was significant in an inappropriate pattern for those subjects who did not understand the instructions, and was not significant, but in the hypothesized direction for those subjects who did understand. Desirability was not significant for either group of subjects for the middle probability phrase pair, tossup and unspecified. The subjects who did not understand the instructions had mean responses in the appropriate pattern, however, while those subjects who did understand the instructions had an inappropriate pattern of

mean responses. The interaction of probability and desirability was not significant in any of these cases.

Reanalyses on Most Reliable Subjects

Analyses, similar to the ones described above, were performed on the ten subjects with between session correlations of responses at or above the median level of 0.70. Individual phrase analyses showed that, for every phrase, the probability of the visible spinner landing on white was significant in the hypothesized direction. However, only for the phrase, likely, was the spinner desirability effect significant, and this effect was not precisely in the hypothesized direction. No other effects or interactions were significant.

For the analyses on phrase pairs, the probability of the visible spinner landing on white was significant in the hypothesized direction in every case. For the low probability phrases, spinner desirability was significant in the hypothesized direction, while for the high probability phrases, it was significant, but not exactly in the hypothesized pattern. Spinner desirability was not significant for the middle probability phrases. In short, these analyses did not produce any better results than had been obtained previously on all twenty subjects.

Discussion

In general, the results show that IO desirability affects judgments of relative probability when the

) uncertainty is represented by verbal expressions. The probability phrases, doubtful, good chance, and likely, showed this effect clearly when analyzed separately. The phrase, improbable, was also significantly affected by desirability, but did not show precisely the hypothesized pattern of responses. The middle phrases, tossup and unspecified, showed the expected pattern of mean responses, but the results were not statistically significant.

The results of the analyses on pairs of phrases showed that desirability was significant in the hypothesized direction for the high pair, good chance and likely, as expected. The increased power of this analysis also produced a significant desirability effect in the appropriate direction for the low pair, doubtful and improbable. The middle pair, tossup and unspecified, still failed to achieve significance for the desirability effect, although the pattern of responses was in the hypothesized direction.

Differential Effects

The results of this experiment indicate a clear difference in the effects of desirability on the low, middle, and high probability phrases. Any explanation of the overall effect of desirability on judgments under ambiguity and ignorance must take this finding into account.

) Judged meanings of phrases. The low phrases, doubtful and improbable, showed smaller effects of both objective probability and desirability than did the high

phrases, good chance and likely. Other research (Wallsten, Fillenbaum, & Cox, 1986) has also shown that low probability phrases are less affected by context than are neutral and high probability phrases. Context in that case referred to the base rate of occurrence of events. A possible reason that context has less effect on low than on high phrases might be that the meanings of low phrases are further from the center of the probability interval than are the meanings of high phrases. This would give the meanings of high probability phrases more freedom to undergo shifts when influenced by factors such as desirability or objective probability value.

The data from this experiment are not suitable for precisely determining the best judged meanings of the phrases. By examining the data for each subject averaged over all levels of desirability, however, one can note at what approximate point the phrase and probability were judged equal (i.e. where the mean response was 100). Out of 40 such determinations for the low terms, 26 were located below the probability value of 0.22. Eleven of these 26 determinations were for doubtful and the remaining 15 were for improbable. In contrast, the high terms had only 3 out of 40 determinations above the probability value of 0.79, all three occurring for the phrase, good chance. This indicates that the best judged meanings for high probability phrases are less extreme, on average, than they are for low probability phrases.

Precision of meaning. A reason for the lack of significant results for the middle probability terms may stem from the relative precision of their meanings. Tossup was the most specific of all the terms used in this study, with a relatively precise meaning of 0.50. In contrast, unspecified would seem to be the least specific of all the terms used in this experiment. However, the results of the questionnaire (see Appendix, question 3) indicated that unspecified was also treated as having a relatively precise meaning of 0.50 probability. Thirteen of the twenty subjects explicitly mentioned the 0.50 probability in the visible spinner as the cutoff point for deciding whether the visible or the invisible spinner was more likely to land on white. The meanings of both tossup and unspecified may therefore have been more impervious to the effects of independent outcome desirability than the meanings of the other phrases used in this study, resulting in the small, non-significant results reported here.

Unfortunately, the above result and interpretation also imply that the phrase, unspecified, may not truly represent Ellsberg's condition of "complete ignorance". It may be more closely related to a condition of ambiguity.

Shifts in subjective value. Previous research (Irwin & Graess, 1968; Morlock & Hertz, 1964; Crandall, Solomon, & Kellaway, 1955) showed that IO desirability caused a significant shift in the subjective values accorded to objective probabilities, with the largest shifts found near the middle of the probability range. In the present study,

however, this pattern of results did not occur. Note that it cannot be determined whether the significant effects of desirability found at the low and high ends of the continuum are due to shifts in the subjective values accorded the visible probabilities or to changes in the meanings of the phrases. However, the lack of any significant effects due to desirability in the middle range of probability would seem to indicate that only very small shifts, if any, occurred in the subjective probability accorded to the visible spinner in that range.

That the shifts in subjective value accorded the visible spinner probability were not significant (for the middle range of probability, at least) may relate to the fact that this experiment is fundamentally different from the earlier studies of independent outcome desirability. That is, subjects were never before required to compare the relative likelihoods of probabilities and probability phrases. The task of making this type of comparison may interact with desirability or produce effects that add to those of desirability. One such effect, discussed further below, is ambiguity aversion. For example, since there is some evidence that the middle probability phrases were the most precise in meaning, and thus the least ambiguous, it should not be surprising if ambiguity aversion effects were smallest for these terms.

Asymmetry of effects. One interesting facet relating to this interpretation of the results is that the IO

desirability effect was generally not symmetric. That is, for four of the six phrases (all except doubtful and unspecified), the effect was greater for negative spinner/positive phrase desirability than it was for the opposite type of desirability. This implies that either the judged meaning of the phrase increased more, or that the subjective value accorded the objective probability decreased more for this condition than for the opposite one. One interpretation of this result is that people normally have a bias against ambiguity, and towards firm probabilities, as previous research has shown (Ellsberg, 1961; Becker & Brownson, 1964; Yates & Zukowski, 1976; Einhorn & Hogarth, 1986). Further evidence for this point is given in Table 2, where the mean responses at zero desirability are all greater than 100, indicating an avoidance of ambiguity. Therefore, attempts to make the ambiguity-avoidance bias stronger will show fairly small effects. In contrast, attempts to induce the opposite type of bias may show larger effects.

Membership functions. Another explanation for the results of this experiment incorporates the concept of membership functions. There are two commonly encountered shapes of membership functions for probability terms: (a) single-peaked, and (b) monotonic (Wallsten, et al, in press; Rapoport, et al, in press). Figure 3 shows some examples of the different types of membership functions one might encounter in this experiment.

In order to link the results from this experiment to

the concept of membership functions one must posit some fairly strong assumptions. First, it must be assumed that unique membership functions exist for each probability phrase and that these functions are generally of a monotonic or single-peaked shape. Secondly, it must be assumed that these membership functions have a maximum, which may occur at either a single point or over a range of values. The third assumption, which is the most tenuous, is that in judging how much more likely one spinner rather than the other is to land on white, the subject focuses only on how well the displayed objective probability is described by the phrase associated with the invisible spinner. In other words, the response depends only on the value of that objective probability in that phrase's membership function. Thus, if an objective probability with the value 1.0 in a phrase's membership function were paired with that phrase, the subject would consider both the visible and invisible spinners equally likely to land on white, and would leave the arrow in the middle of the response line.

Recall that the response line is scaled 0 (definite that the invisible spinner is more likely to land on white) to 200 (definite that the visible spinner is more likely to land on white). Given the above assumptions, responses can be transformed to membership function values by the following formula:

$$m(p) = 1 - [(100 - x) / 100] ,$$

where $x \in [0, 200]$ is the response.

There are other assumptions which give a different interpretation to the best judged meaning of a probability phrase. For example, it has been suggested that subjects may think of the low, central, and high probability values for a given phrase and then perform a type of weighted average of these values to derive their best judged meaning of the phrase (Wallsten, 1986). Depending on the weights given to each of these probability values, the best judged meaning may or may not occur at the peak of the membership function for the given phrase. This will be discussed in more detail below.

Given the three assumptions initially described above, from each of the low, middle, and high probability phrase pairs, 40 (20 subjects by 2 phrases) membership functions can be generated. Ignoring the desirability factor temporarily, and using mean responses averaged over all levels of desirability, it was found that the vast majority (89%) of the derived membership functions were single-peaked. These results are shown as the last three columns in Table 7.

Next, the responses were examined at each level of desirability to determine if the desirability factor had an effect on the shape of subjects' membership functions. These results, displayed in the first nine columns of Table 7, show that there was no clear effect on the shape of subjects' membership functions due to desirability, since the total number of monotonic membership functions under

each level of desirability was about the same (12 [10%], 11 [9%], and 16 [13%]). The zero desirability condition had the fewest number of other-shaped membership functions, but this could simply mean that the desirability factor introduced more variability in subjects' responses. Individual subjects did have differently shaped functions depending on desirability level, however, which accounts for the discrepancy between this analysis and the preceding one, in which a total of only seven monotonic membership functions were found.

One finding that is the same in both analyses and consistent with other research (Wallsten, et al, in press) is that monotonic functions are more prevalent among the lower probability phrases than among the higher ones. However, the number of monotonic membership functions found in this study is much smaller than has been normally found in other research (Wallsten, et al, in press; Rapoport, Wallsten, & Cox, in press). This may indicate that mean responses of 100 do not represent the peak values of the membership functions, but rather the weighted average of low, central, and high probability values relevant to a given probability phrase. Under this interpretation, it is possible that many of the membership functions labeled as single-peaked could actually be monotonic, given the appropriate set of weights used in the averaging process. These weights are unknown parameters, however, which cannot be determined from the present data.

Based on the assumptions initially listed above, desirability seemed to have no effect on the shapes of the derived membership functions. Thus, it is hypothesized instead that IO desirability caused a shift in the location of the membership function for a given term. When the desirability was positive for the spinner and negative for the probability phrase, the membership function was shifted to the left of where it was when there was no desirability (toward lower probability values). When the desirability was negative for the spinner and positive for the probability phrase, the membership function was shifted to the right of where it was when there was no desirability (toward higher probability values). Figure 4 shows this shift for the three main types of membership functions that were encountered in this experiment.

Using this model, one can explain the effect of desirability on the probability terms, doubtful, good chance, and likely, as consisting of a significant shifting of the location of the membership functions for these terms. For the term, improbable, the shift was not completely made for the case of positive spinner desirability. A possible reason for this is that the membership function for improbable under conditions of no desirability may be located near the bottom of the range of probability. Thus, there may not have been enough room for a significant shift to the left, as predicted for the positive spinner/negative phrase desirability effect. The data show that the mean responses for improbable were

higher than those for the other low probability term, doubtful, when responses were averaged over all factors, and also when these responses were averaged over each level of desirability. These results imply that the meaning of improbable was the lowest, among all the phrases used in this study, in terms of subjective probability. This gives some credence to the notion that the membership function for improbable may be so close to the bottom of the range of probability values that it precludes a significant shift toward lower probabilities. However, the finding that improbable had a lower judged meaning than did doubtful is inconsistent with previous research (Wallsten, et al, in press).

Alternative Explanation

A different interpretation of the results of this experiment involves the use of the weighted average of the low, central, and high probability values for a given phrase (cf. Wallsten, 1986). The pattern of results for the desirability factor can be explained by simply assuming that positive spinner/negative phrase desirability causes subjects to give more weight to the low probability values, while negative spinner/positive phrase desirability causes subjects to give more weight to the high probability values. In other words, when the phrase describes a potential loss, subjects concentrate more on the low probability for that phrase, and when it describes a potential gain, they concentrate more on the high

probability for that phrase. The different desirability effects for low, middle, and high probability phrases can be explained by assuming that the weights used in the averaging process for the middle probability terms were all about equal, while for the low probability phrases they were slightly different from each other, and for the high probability phrases they were even more different from each other. Of course, further research is necessary to independently test these assumptions.

The Ellsberg Paradox

Tests of the Ellsberg paradox showed an effect opposite to Ellsberg's hypothesis under one condition of desirability. The one significant result in this analysis (See Table 5) showed that when desirability favored the phrase, subjects considered the ambiguous uncertainty (the phrase, tossup) to be more likely than the precise one (the objective probability of 0.50). This is consistent with the hypothesis that subjects' responses will be biased towards the outcome with the relatively favorable desirability.

Further, it is of interest that the results predicted from the Ellsberg paradox did not occur in any case. That is, subjects never accorded a significantly higher subjective value to the objective probability of 0.50 than to the phrases, tossup and unspecified. Possibly, Ellsberg was dealing with betting preferences, while the present study deals with judgments of relative likelihood. Thus, the Ellsberg paradox may actually represent a type of

ambiguity aversion in a gambling scenario.

If the term, unspecified, truly represented a condition of ignorance, then desirability did not exert a strong enough effect to cause subjects to consider the condition of ignorance as more likely than the condition of risk. Given the previous discussion of the fairly precise meaning of unspecified, however, it may be that this term represented ambiguity rather than ignorance.

Differences from Previous Research

There are some important differences between these results and those found by earlier studies. First, it was previously found that the effect of IO was strongest for objective probabilities of 0.50 (Irwin & Snodgrass, 1966; Morlock & Hertz, 1964; Crandall, Solomon, & Kellaway, 1955). It was therefore hypothesized that the strongest effect of IO desirability in the present study would be found in the case where the probability 0.50 was paired with the phrase, tossup. This turned out to be one of the weakest effects in the experiment, although it was in the appropriate direction. Another interesting difference was that the high terms seemed to show stronger effects of desirability than did the low terms. This result did not occur in the earlier research under conditions of risk. Both of these differences may be the result of this experiment's requiring a judgment of relative likelihood between a verbal expression of ambiguity and an objectively stated probability, whereas, the earlier studies dealt

solely with objective probabilities. One finding from the earlier research on independent outcome desirability that was replicated was that the desirability effect was maintained even though the independent outcomes were probabilistic (cf. Irwin & Metzger, 1966).

It turned out that the nature of the present task was problematic. Subjects were asked to concentrate on making their best judgment as to the relative likelihoods of the two paired events on each trial. They were told that the points were put on the screen to help them pay attention to the task and to make it more interesting. Despite the written instructions, which indicated that all outcomes were based solely on fixed, predetermined probability values, some subjects mentioned after the experiment that they tried to figure out what "system" governed the awarding of points. The questionnaire given to subjects at the end of the experiment was an attempt to determine how deeply this tendency affected subjects' responses. As pointed out above, however, the questionnaire did not seem to provide any valid information on this issue.

Summary and Conclusions

This experiment showed that independent outcome desirability is one context factor that appears to affect the vague meanings of probability phrases. Judgments of relative likelihood were found to be biased somewhat toward the outcome with the relatively favorable desirability. Given certain assumptions, independent outcome desirability

effects can be interpreted as causing shifts in the location of the membership functions for probability phrases, without changing the shape of the membership functions. A different explanation could involve the way in which desirability affects the weighted averaging process used to derive the best judged meaning for a probability phrase.

The most plausible explanation of the IO desirability effects would need to involve all the factors relevant to a judgment of relative likelihood between a probability phrase and an objective probability. Three of these factors, discussed above, are: (a) the subjective value that is accorded to the objectively stated probability, (b) the judged meaning of the probability phrase, and (c) ambiguity aversion. It would be desirable to study each of these factors in isolation within the context of examining the meanings of probability phrases.

References

- Becker, S. W., & Brownson, F. O. (1964). What price ambiguity? Or the role of ambiguity in decision making. Journal of Political Economy, 72, 62-73.
- Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.
- Budescu, D. V. & Wallsten, T. S. (in press). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Performance.
- Crandall, V. J., Solomon, D., & Kellaway, R. (1955). Expectancy statements and decision times as functions of objective probabilities and reinforcement values. Journal of Personality, 24, 192-203.
- Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. Psychological Review, 92, 433-461.
- Einhorn, H. J. & Hogarth, R. M. (in press). Decision making under ambiguity. Journal of Business.
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. Quarterly Journal of Economics, 75, 643-669.
- Fellner, W.. Distortion of subjective probabilities as a reaction to uncertainty. Quarterly Journal of Economics, 75, 670-689.

- Hakel, M. (1968). How often is often? American Psychologist, 23, 533-534.
- Irwin, F. W. (1944). The realism of expectations. Psychological Review, 51, 120-126.
- Irwin, F. W., & Graess, C. N. (1968). Tests of the discontinuity hypothesis of the effects of independent outcome values upon bets. Journal of Experimental Psychology, 76, 444-449.
- Irwin, F. W., & Metzger, M. J. (1966). Effects of probabilistic independent outcomes upon predictions. Psychonomic Science, 5, 79-80.
- Irwin, F. W., & Snodgrass, J. G. (1966). Effects of independent and dependent outcome values upon bets. Journal of Experimental Psychology, 71, 282-285.
- Knight, F. H. (1921). Risk, Uncertainty, and Profit. Boston: Houghton Mifflin Company.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). Foundations of Measurement, Vol. I. New York: Academic Press.
- Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science, 9, 563-564.

1
Marks, R. (1951). The effect of probability, desirability, and "privilege" on the stated expectations of children. Journal of Personality, 19, 332-351.

Morlock, H. C., & Hertz, K. J. (1964). Effect of the desirability of outcomes on decision making. Psychological Reports, 14, 11-17.

Pepper, S., & Prytulak, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. Journal of Research in Personality, 8, 95-101.

Pruitt, D. G., & Hoge, R. D. (1965). Strength of the relationship between the value of an event and its subjective probability as a function of method of measurement. Journal of Experimental Psychology, 69, 483-489.

Radner, R. (1972). Normative theory of individual decision. In McGuire, C. B. & Radner, R., eds.. Decision and Organization. Amsterdam: North Holland.

Rapoport, A., Wallsten, T. S., & Cox, J. A. (in press). Direct and indirect scaling of membership functions of probability phrases. Mathematical Modeling.

Savage, L. J. (1954). The Foundations of Statistics. New York: Wiley.

4

) Scheibe, K. E. (1964). The effect of value on statements of expectancy under four experimental conditions. The Psychological Record, 14, 137-144.

Simpson, R. H. (1944). The specific meanings of certain terms indicating differing degrees of frequency. The Quarterly Journal of Speech, 30, 328-330.

Simpson, R. H. (1963). Stability in meanings for quantitative terms: A comparison over 20 years. The Quarterly Journal of Speech, 49, 146-151.

Slovic, P. (1966). Value as a determiner of subjective probability. IEEE Transactions on Human Factors in Electronics, HFE-7, 22-28.

Stone, D. R., & Johnson, R. J. (1959). A study of words indicating frequency. Journal of Educational Psychology, 50, 224-227.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (in press). Measuring the vagueness of probability terms. Journal of Experimental Psychology: General.

Wallsten, T. S., Fillenbaum, S., & Cox, J. (1986). Base rate effects on the interpretations of probability and frequency expressions. Journal of Memory and Language, 25, 571-587.

)

Wallsten, T. S. (1986). National science foundation grant

proposal.

Watson, S. R., Weiss, J. J., & Donnell, M. C. (1979).

Fuzzy decision analysis. IEEE Transactions on Systems, Man, and Cybernetics, SCM-9, 1-9.

Winer, B. J. (1971). Statistical Principles in Experimental Design, 2nd ed.. New York: McGraw-Hill.

Yates, J. F., & Zukowski, L. G. (1976). Characterization of ambiguity in decision making. Behavioral Science, 21, 19-25.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-38.

Zimmer, A. C. (1983). Verbal vs. numerical processing of subjective probabilities. In R. W. Scholz (Ed.), Decision Making Under Uncertainty. Amsterdam: North Holland Publishing Co., pp. 159-182.

Table 1

Test-Retest Correlations

Subject	Simple	Pooled
1	0.69	0.69
2	0.84	0.83
3	-0.21	-0.16
4	0.68	0.69
5	0.75	0.76
6	0.86	0.86
7	0.47	0.46
8	0.69	0.69
9	0.81	0.81
10	0.93	0.92
11	0.77	0.76
12	0.71	0.72
13	0.72	0.71
14	0.42	0.42
15	0.76	0.76
16	0.62	0.63
17	0.79	0.80
18	0.54	0.53
19	0.68	0.67
20	0.24	0.22
Mean	0.61	0.61

Table 2

Desirability Effects in Terms of Mean Response Values
for Each Phrase and Combined over Similar Phrases

Spinner = Phrase =	-500 +500	0 0	+500 -500
Doubtful	108.47	112.09	118.11
Good Chance	97.77	110.68	115.56
Improbable	111.21	120.24	118.51
Likely	94.99	112.93	115.98
Tossup	96.34	101.39	103.44
Unspecified	102.68	103.32	107.36
Doubtful & Improbable	109.84	116.16	118.31
Good Chance & Likely	96.38	111.80	115.77
Tossup & Unspecified	99.51	102.36	105.40

Table 3

F-Ratios from ANOVAs on Individual Probability Phrases

Phrase	Sessions ^a	Probability ^b	Desirability ^c
Doubtful	0.86	63.51 **	5.02 *
Good Chance	0.00	73.13 **	5.67 **
Improbable	0.62	51.37 **	5.53 **
Likely	1.14	61.28 **	22.90 **
Tossup	0.01	76.60 **	1.38
Unspecified	1.08	36.11 **	0.47

a) df = (1,19)

b) df = (3,57) for Doubtful, Good Chance, Improbable, and
Likely

df = (2,38) for Tossup and Unspecified

c) df = (2,38)

* $p < 0.05$. ** $p < 0.01$.

Table 4

F-Ratios from ANOVAS on Combined Probability Phrases

Phrase Pair	Phrases ^a	Probability ^b	Desirability ^c	Probability ^d x Desirability
Doubtful & Improbable	3.39	66.94 **	6.81 **	0.78
Good Chance & Likely	0.00	78.46 **	13.71 **	1.89
Tossup & Unspecified	3.84	65.69 **	1.24	2.08

a) df = (1,19)

b) df = (3,57) for Doubtful & Improbable, Good Chance &
Likely

df = (2,38) for Tossup & Unspecified

c) df = (2,38)

d) df = (6,114) for Doubtful & Improbable, Good Chance &
Likely

df = (4,76) for Tossup & Unspecified

* $p < 0.05$. ** $p < 0.01$.

Table 5

Tests of Whether Mean Responses are Significantly
Below 100 for the Neutral Phrases vs. 0.50

Spinner =	-500	0	+500
Phrase =	+500	0	-500
Tossup			
Mean Response	94.06	97.58	103.93
<u>T</u> -Value	-2.01 *	-0.79	1.26
Unspecified			
Mean Response	98.00	99.61	105.06
<u>T</u> -Value	-0.48	-0.11	1.37

* $p < 0.05$.

Table 6

F-Ratios from ANOVAs for Subjects who
Understood/Did not Understand the Instructions

Phrase Pair	Phrase	Probability	Desirability	Probability x Desirability
Subjects who Understood the Instructions				
Doubtful & Improbable (df)=	0.79 (1,10)	47.62 ** (3,30)	3.25 (2,20)	1.23 (6,60)
Good Chance & Likely (df)=	0.81 (1,10)	153.03 ** (3,30)	7.41 ** (2,20)	1.40 (6,60)
Tossup & Unspecified (df)=	5.86 * (1,10)	49.60 ** (2,20)	2.48 (2,20)	0.31 (4,40)
Subjects who did not Understand the Instructions				
Doubtful & Improbable (df)=	4.75 (1,8)	24.34 ** (3,24)	4.21 * (2,16)	0.52 (6,48)
Good Chance & Likely (df)=	2.51 (1,8)	14.67 ** (3,24)	5.95 * (2,16)	1.14 (6,48)
Tossup & Unspecified (df)=	0.02 (1,8)	19.50 ** (2,16)	2.29 (2,16)	0.93 (4,32)

* $p < 0.05$. ** $p < 0.01$.

Table 7

Monotonic (M), Single-Peaked (SP), and Other (O)
 Membership Functions for Each Phrase and Desirability Level,
 and Averaged over All Desirability Levels

Spinner = Phrase =	-500 +500	0 0	+500 -500	Overall
	M SP O	M SP O	M SP O	M SP O
Low phrases				
Doubtful	3 14 3	2 14 4	1 13 6	2 17 1
Improbable	3 11 6	5 12 3	6 12 2	3 16 1
Sub-Total	6 25 9	7 26 7	7 25 8	5 33 2
Middle phrases				
Tossup	2 15 3	1 17 2	2 15 3	0 20 0
Unspecified	2 14 4	2 17 1	6 14 0	2 18 0
Sub-Total	4 29 7	3 34 3	8 29 3	2 38 0
High phrases				
Good Chance	2 16 2	1 18 1	0 14 6	0 18 2
Likely	0 16 4	0 18 2	1 17 2	0 18 2
Sub-Total	2 32 6	1 36 3	1 31 8	0 36 4
Totals	12 86 22	11 96 13	16 85 19	7 107 6

Figure 1

Graph of Mean Responses as a Function of Probability for Individual Phrases, Ignoring Desirability

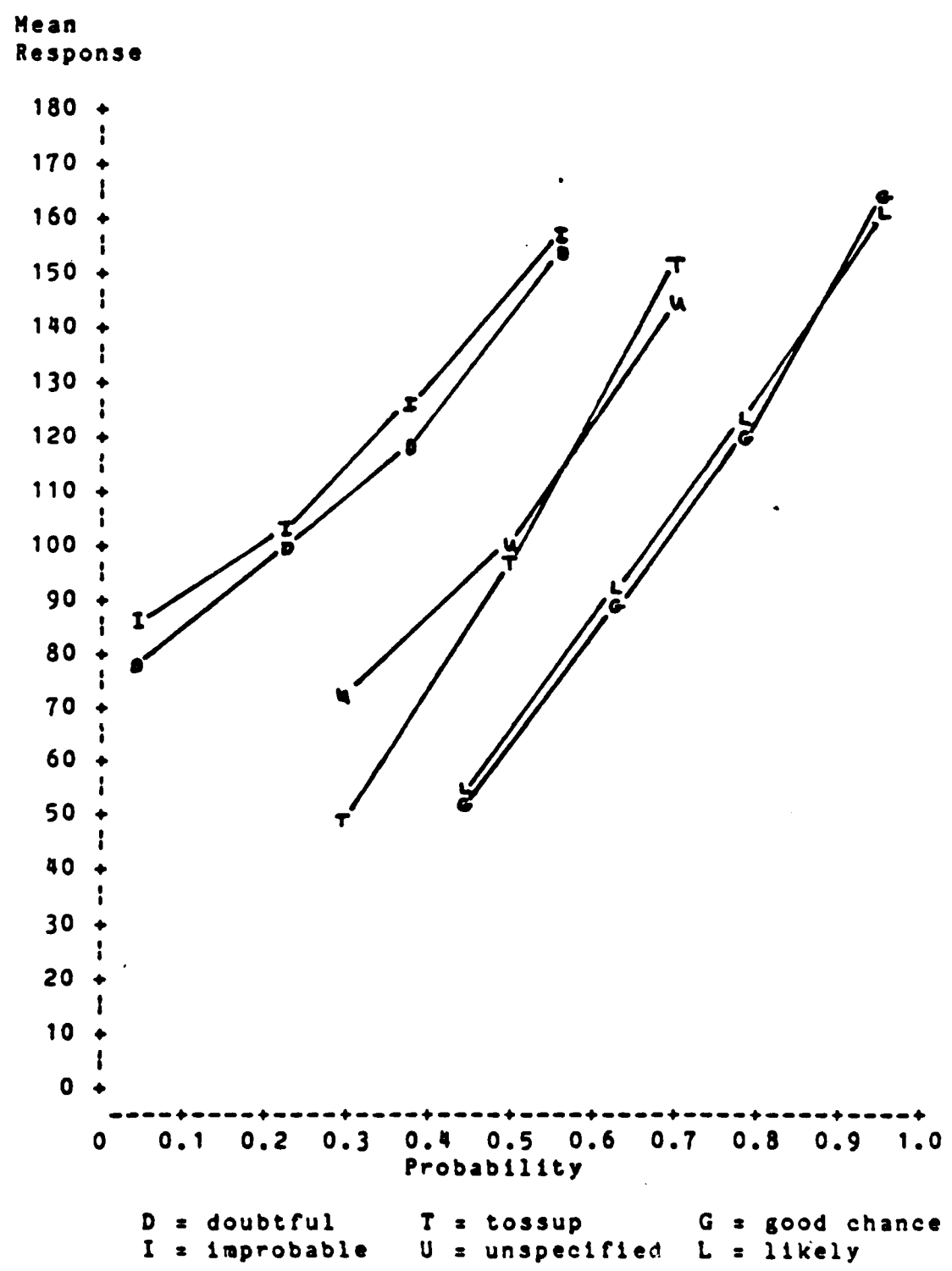
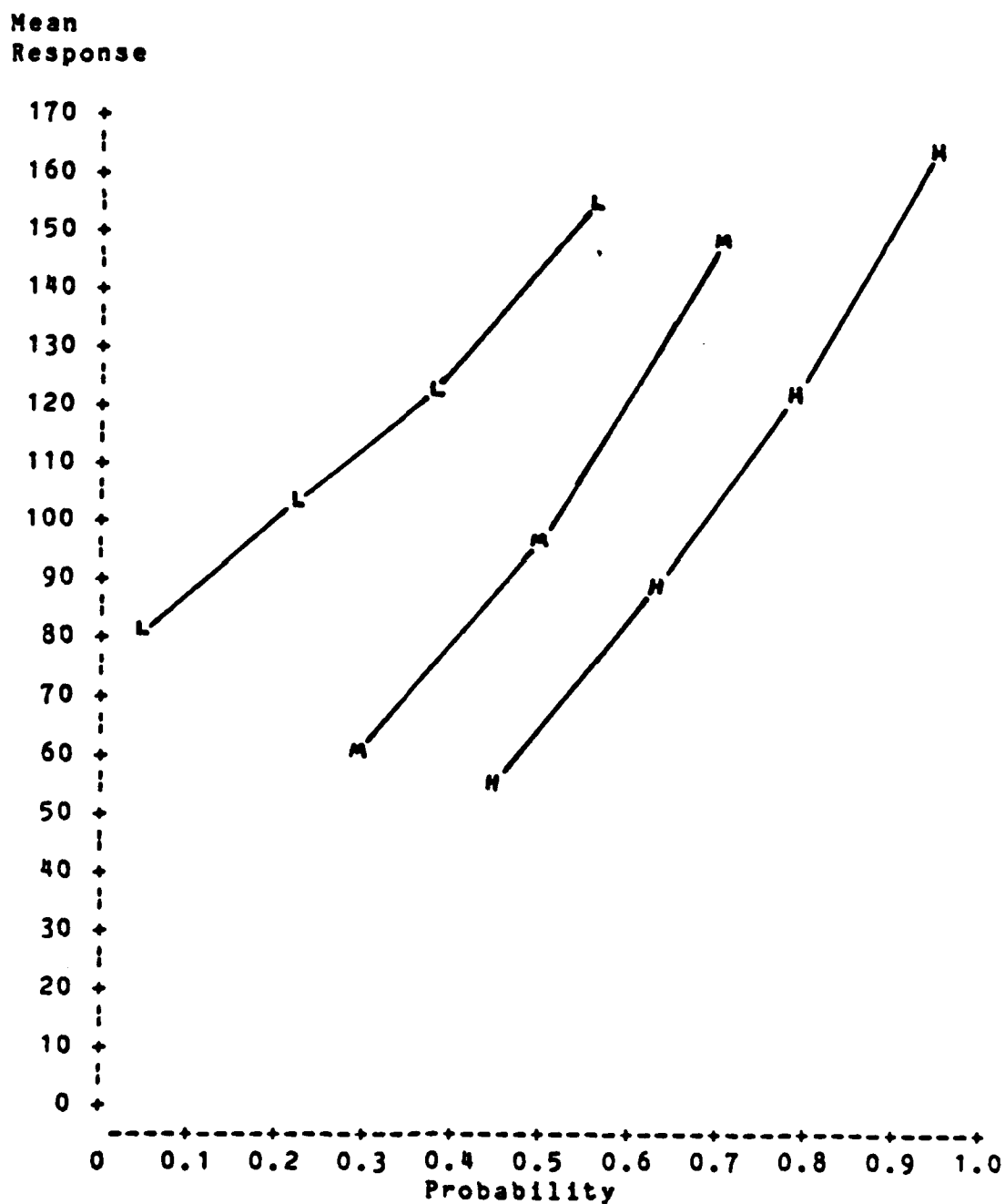


Figure 2

Graph of Mean Responses as a Function of Probability for
Phrase Pairs, Ignoring Desirability

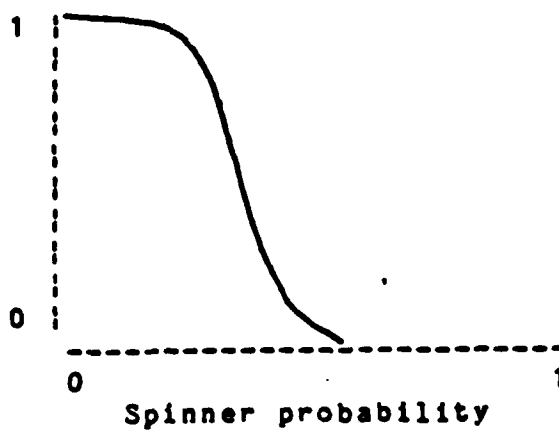


L = Low probability phrases (doubtful & improbable)
M = Middle probability phrases (tossup & unspecified)
H = High probability phrases (likely & good chance)

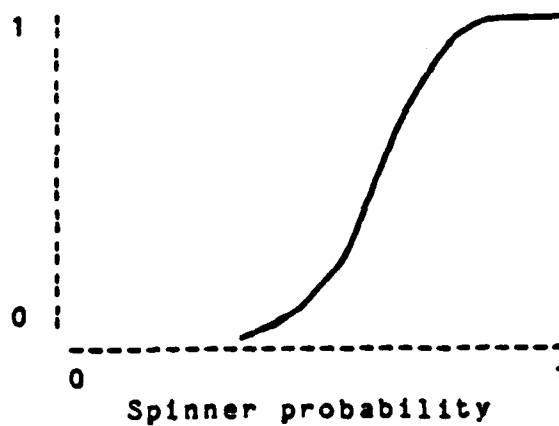
Figure 3

Types of Membership Functions

1. Monotonic decreasing



2. Monotonically increasing



3. Single-peaked

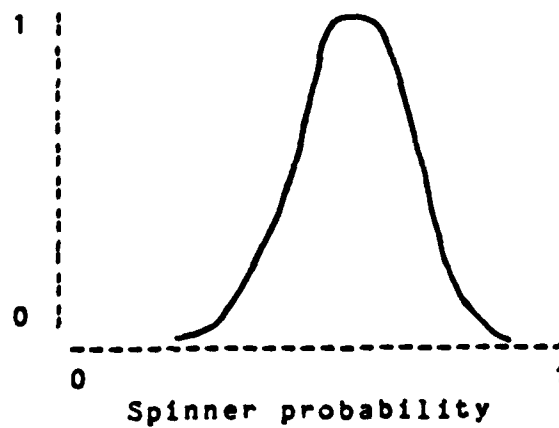
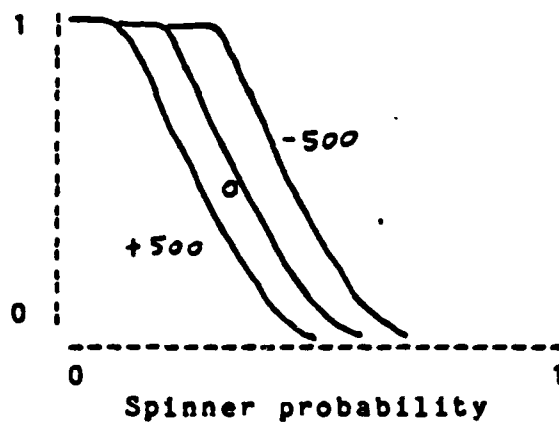


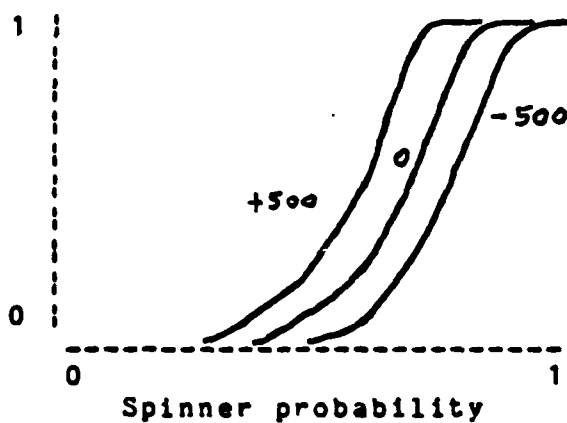
Figure 4

Hypothesized Effect of Independent Outcome Desirability
on Membership Functions

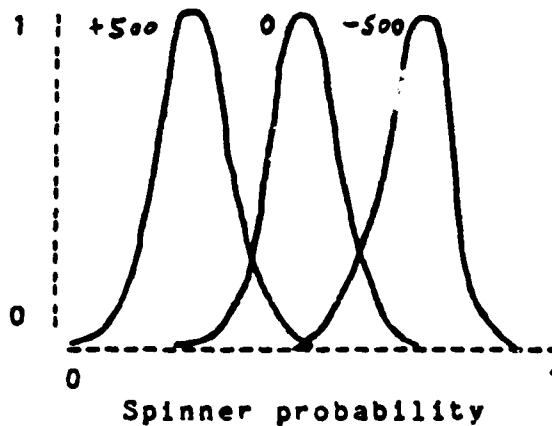
1. Monotonic decreasing



2. Monotonic increasing



3. Single-peaked



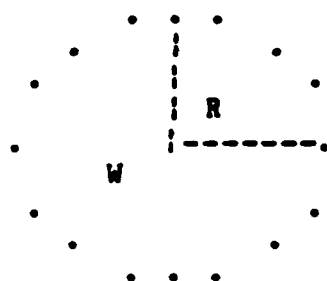
Appendix

1)

WHITE: -500
RED: 0

WINNINGS:
1000

WHITE: 500
RED: 0



LIKELY
to land on
white.

✓

In the screen drawn above, the visible spinner is on the left and the word, Likely, is on the right. Inside the spinner, the letters W and R stand for the colors white and red, respectively. Suppose that you viewed this screen during the experiment and you positioned the arrow approximately where it is in the drawing above.

Given that you have the arrow positioned on the right side of the line segment, what effect does this have on the likelihood of the visible spinner's landing on white?

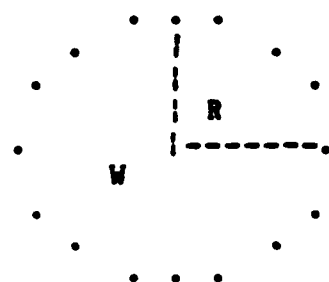
- A) It increases the likelihood of the visible spinner's landing on white.
- B) It decreases the likelihood of the visible spinner's landing on white.
- C) It has no effect on the likelihood of the visible spinner's landing on white.

2)

WHITE: -500
RED: 0

WINNINGS:
1000

WHITE: 500
RED: 0



LIKELY
to land on
white.

∨

----->

In the screen drawn above, the visible spinner is on the left and the word, Likely, is on the right. Inside the spinner, the letters W and R stand for the colors white and red, respectively. Suppose that you viewed this screen during the experiment and you positioned the arrow approximately where it is in the drawing above.

Now, it turned out that both the visible spinner and the invisible spinner landed on white. In this case, describe what happens to your cumulative total of points:

- A) It increases by 500 points.
- B) It decreases by 500 points.
- C) It increases by 1000 points.
- D) It decreases by 1000 points.
- E) Nothing happens to your cumulative total of points.

3) a) In general, did you think that where you placed the arrow influenced the outcome of the spinners?

b) Did you think that your placement of the arrow had an effect on the amount of points that you could win or lose on a given trial?

c) What guided your placement of the arrow when you were presented the word, "unspecified?"

THE EFFECT OF OUTCOME DESIRABILITY
ON COMPARISONS OF NUMERICAL AND LINGUISTIC PROBABILITIES

by

Brent L. Cohen

A Thesis submitted to the faculty of The University
of North Carolina at Chapel Hill in partial
fulfillment of the requirements for the degree of
Master of Arts in the Department of Psychology.

Chapel Hill

1986

Approved by:

Thomas S. Wallsten
Advisor

Harvey Lousman
Reader

Samuel T. Little
Reader

BRENT L. COHEN. The Effect of Outcome Desirability on
Comparisons of Numerical and Linguistic Probabilities
(Under the direction of THOMAS S. WALLSTEN.)

ABSTRACT

When subjects have complete knowledge of the probability distribution governing a set of outcomes, their subjective probability estimates have been found to be biased towards the outcome with the relatively favorable desirability. With the use of linguistic probability expressions, this study attempted to extend this finding to situations in which subjects have only vague knowledge or no knowledge of the probability distribution relevant to a set of outcomes. It was found that subjects' comparisons were biased, in varying degrees depending on the particular phrase under consideration, towards the relatively more desirable outcome. This finding indicates that a complete model of judgment under uncertainty must include a factor for context effects such as outcome desirability.

ACKNOWLEDGEMENTS

This research was supported by contract MDA903-83-K-0347 from the U.S. Army Research Institute. The views, opinions, and findings contained in this paper are those of the author and should not be construed as official Department of the Army position, policy, or decision.

I would like to thank my advisor, Dr. Thomas S. Wallsten, for all of his advice and encouragement to me as I worked on this project. His comments and constructive criticisms on the earlier drafts of this paper have greatly improved the quality of the final product. Additionally, the financial support from his research grant made it possible to pay subjects and perform the computer analyses. Also, the other members of my committee, Dr. Samuel Fillenbaum and Dr. Marcy Lansman, gave many helpful suggestions and aided my thinking on this project. I also want to thank my fellow students in the Psychometric Lab who helped me with the chore of data management.

A great debt of thanks is due my parents who constantly encouraged me to finish. Finally, I must thank Marcy Freund, who was a great inspiration and support to me.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	111
LIST OF TABLES	v
LIST OF FIGURES	vi
INTRODUCTION	1
LITERATURE REVIEW	3
PURPOSE AND HYPOTHESES	14
METHOD	16
RESULTS	20
DISCUSSION	27
SUMMARY AND CONCLUSIONS	40
REFERENCES	42
APPENDIX	58

LIST OF TABLES

Table	Page
1. Test-retest correlations	47
2. Desirability effects in terms of mean response values for each phrase and combined over similar phrases	48
3. F-ratios from ANOVAs on individual probability phrases	49
4. F-ratios from ANOVAs on combined probability phrases	50
5. Tests of whether mean responses are significantly below 100 for the neutral phrases vs. 0.50	51
6. F-ratios from ANOVAs for subjects who understood/ did not understand the instructions	52
7. Monotonic (M), single-peaked (SP), and other (O) membership functions for each phrase and desirability level, and averaged over all desirability levels	53

LIST OF FIGURES

Figure	Page
1. Graph of mean responses as a function of probability for individual phrases, ignoring desirability	54
2. Graph of mean responses as a function of probability for phrase pairs, ignoring desirability	55
3. Types of membership functions	56
4. Hypothesized effect of independent outcome desirability on membership functions	57

Introduction

Earlier theories of judgment under uncertainty (e.g. Knight, 1921; Savage, 1954) have assumed that the decision-maker is a rational being, conforming to basic axioms of probability. This study examines an aspect of irrational decision-making; namely, the ways in which judgments about uncertain outcomes are affected by the independent desirability of those outcomes. The independent desirability of an outcome refers to the value of the outcome itself, independently of one's prediction. This study extends the finding (Marks, 1951; Irwin & Snodgrass, 1966; Irwin & Metzger, 1966; Slovic, 1966) that independent outcome desirability affects subjective probability, in direct opposition to a fundamental axiom of rationality.

The work of Irwin, et al., on independent outcome desirability was motivated in part from Irwin's interest in the types of factors affecting people's expectations of outcomes. He held that there were cognitive factors, which would lead to realistic expectations, as well as affective factors, which would lead to unrealistic expectations (Irwin, 1944). Underlying this research may have been Irwin's intent to develop a more complete model of judgment than the subjectively expected utility (SEU) model. The results found by Irwin, et al. indicate that, at the least,

it may be necessary to control for independent outcome desirability in SEU models.

Uncertainty is a broad concept, including all states of knowledge other than perfect about the occurrence of events. As an attempt to segment the construct of uncertainty, Ellsberg (1961) described three states of knowledge regarding the probabilities of future events. He labeled these conditions "risk", "ambiguity", and "complete ignorance", representing precise knowledge of probabilities, vague or imprecise knowledge of probabilities, and absolutely no information about probabilities, respectively. If the probabilities are known to be either exactly 0 or exactly 1, however, this corresponds to a situation of certainty.

As will be documented below, it is already known that under conditions of risk, subjective probabilities are positively related to independent outcome desirability. However, people commonly make decisions based on information that is ambiguous or imprecise, or even on the basis of no information at all, which corresponds to Ellsberg's states of ambiguity and complete ignorance, respectively. This study examines the effect of independent outcome desirability on judgments under ambiguity and complete ignorance.

There are at least two major ways to express uncertainty in an ambiguous, vague, or imprecise manner. One way is to use a range of numerical probability values to describe the probability of an outcome, while the other

is to use verbal expressions of uncertainty. The latter was used in the present study, in order to further investigate the factors that affect the meanings of probability terms.

Literature Review

Rationality, Uncertainty, and the Ellsberg Paradox

For people to be rational decision-makers, they must conform to certain basic axioms of rationality. One of these axioms holds that choices are completely ordered according to the judged values of the consequences of those choices. Another important axiom is the sure-thing principle (Savage, 1954), according to which, people's ordering of alternative choices should be unaffected by adding a constant to the value of the consequences of each choice. A third principle requires people to be free of "wishful thinking" (Radner, 1972). That is, one's ranking of probabilities of events should be independent of the consequences of those events.

These axioms of rationality are not predictive of actual behavior, however. Ellsberg (1961) has shown that certain commonly obtained choice patterns are in violation of the first two of these axioms. The present study is designed to examine violations of all three of these axioms, but the primary hypothesis of this experiment is that decision-makers will consistently violate the third principle listed above.

According to Ellsberg, a condition of risk exists when a decision-maker knows with certainty the probability distribution over a set of outcomes. A state of complete ignorance exists when the decision-maker has absolutely no information about the probability distribution over outcomes. Ellsberg showed that a person's ordering of gambles depends upon whether that person is operating under a condition of risk or a condition of ignorance.

Between the extremes of risk and complete ignorance lies the state of ambiguity, in which the decision-maker can rule out at least one probability distribution. However, under ambiguity, more than one probability distribution remains under consideration, and the decision-maker does not have enough information to determine which probability distribution is operative. One of the purposes of this study is to extend Ellsberg's results to situations of ambiguity represented by probability phrases such as likely, doubtful, good chance, and so forth.

The importance of distinguishing the three types of uncertainty becomes apparent in light of the Ellsberg paradox. This paradox can be described by considering a situation in which there exist two urns containing 100 red and black balls each. The proportion of red and black balls in Urn I is completely unknown, whereas, Urn II contains exactly 50 red balls and 50 black balls. Thus, Urn I corresponds to a state of nearly complete ignorance, while Urn II corresponds to a state of risk. Let RedI, BlackI, RedII, and BlackII represent the event of drawing a

red ball from Urn I, a black ball from Urn I, a red ball from Urn II, and a black ball from Urn II, respectively. Then, the paradox lies in typical answers to the following four questions:

1. Which do you prefer to bet on, RedI or BlackI, or are you indifferent?
2. Which do you prefer to bet on, RedII or BlackII, or are you indifferent?
3. Which do you prefer to bet on, RedI or RedII, or are you indifferent?
4. Which do you prefer to bet on, BlackI or BlackII, or are you indifferent?

Most people answer the first two questions with indifference responses, implying that they feel that $\text{Pr}(\text{RedI}) = \text{Pr}(\text{BlackI})$ and $\text{Pr}(\text{RedII}) = \text{Pr}(\text{BlackII})$. Since $\text{Pr}(\text{RedI}) + \text{Pr}(\text{BlackI}) = 1$ and $\text{Pr}(\text{RedII}) + \text{Pr}(\text{BlackII}) = 1$, these indifference responses further imply that $\text{Pr}(\text{RedI}) = \text{Pr}(\text{RedII}) = \text{Pr}(\text{BlackI}) = \text{Pr}(\text{BlackII}) = 0.50$. A paradox occurs because most people will then respond to questions (3) and (4) with RedII and BlackII, respectively. The response of RedII in question (3) implies that $\text{Pr}(\text{RedII}) > \text{Pr}(\text{RedI})$, from which it logically follows that $\text{Pr}(\text{BlackII}) < \text{Pr}(\text{BlackI})$, due to the fact that $\text{Pr}(\text{RedI}) + \text{Pr}(\text{BlackI}) = 1$. However, this is contradicted by the choice of BlackII in question (4), which implies that $\text{Pr}(\text{BlackII}) > \text{Pr}(\text{BlackI})$. This pattern of responses violates the first two axioms of rationality described above.

From the above discussion, it appears that people prefer to bet on events with known rather than unknown probabilities. Such betting preferences may depend on how probabilities of events are judged under ambiguity, but they may also depend on an aversion to gambling under conditions of ignorance. For research purposes, it is useful to employ a technique that distinguishes judgments under uncertainty from avoidance of ambiguity in gambling scenarios. In the present study, subjects responded by indicating their degree of confidence in their judgment that one event is more likely to occur than is another event. Other studies have examined aspects of the Ellsberg paradox (Fellner, 1961; Becker & Brownson, 1964; Yates & Zukowski, 1976; Einhorn & Hogarth, 1985; Einhorn & Hogarth, in press). In these studies, however, the values of outcomes were dependent upon subjects' responses, as opposed to the present study's use of independent outcome values.

Ambiguity and Probability Terms

One can describe the uncertainty of an outcome in an ambiguous fashion by defining a range of numerical probability values relating to the event in question. An alternative approach toward ambiguity can be found in the use of probability words or phrases. Probability terms, such as likely, doubtful, good chance, etc., are generally used by people to describe uncertain events.

There are many possible reasons that people frequently prefer to use non-numerical estimates of probabilities.

One reason suggested by Wallsten, Budescu, Rapoport, Zwick, and Forsyth (in press) is that people prefer to express imprecise opinions in imprecise terms, such as probability phrases. Secondly, Zimmer (1983) held that most people deal more effectively with verbal than with numerical expressions of uncertainty. Note that children learn how to talk long before they learn how to perform arithmetic and calculate percentages. Additionally, Zimmer noted that a mathematical treatment of probability was not conceived until 1660, while verbal expressions of uncertainty existed in most languages for a long time before that. Another reason for the common use of verbal expressions of uncertainty may stem from people's desire to use a verbal rather than numerical style of written and spoken communication. There are many real-world events for which a numerical probability estimate would seem less appropriate than a verbal probability expression. For example, it seems much more sensible to say "it was possible but not likely", rather than "there was a 10-15% chance" with reference to the event that Shakespeare was thinking of Ann Hathaway when he wrote his twelfth sonnet.

Beyth-Marom (1982) suggested other reasons for the use of non-numerical probability estimates. First, they prevent one's estimate from being judged. Secondly, they allow one to avoid making explicit derivations to defend the estimate. Thirdly, many people believe that a probability phrase is no more ambiguous than is a range of

numerical probability values. Because most people use probability phrases to express degrees of uncertainty, it is important to study the factors that affect the meanings of these probability terms.

Measurement of Vagueness

A problem with studying ambiguity in verbal rather than numerical expressions is that the measurement of ambiguity is not as straightforward. It involves the formulation of a reasonable means of transforming verbal expressions of uncertainty into measurable quantities. Although it will not be necessary to actually measure the meanings of terms used in this study, it is still appropriate to discuss measurement techniques for two reasons. First, prior measurement of each of the terms used in this study (except for the term, unspecified) was required to determine important parts of the experimental design. Secondly, the present research may provide knowledge about factors that should be considered when measuring the meanings of vague terms.

One approach toward the measurement of the meanings of probability words involves gathering numerical probability estimates for probability phrases from a large number of subjects. From these data, information about the central probability value and the variability in probability values for each term can be determined. This technique has been used often in previous research on the meanings of probability or frequency terms (Simpson, 1944, 1963; Stone & Johnson, 1959; Lichsteinstein & Newman, 1967; Hake, 1968;

Beyth-Marom, 1982). All of these studies found considerable variability over subjects in the assignment of numbers to probability phrases. However, the probability phrases did have fairly consistent numerical meanings to individual subjects over time (Beyth-Marom, 1982). A problem with this technique is that it does not address the issue of whether the probability phrases, themselves, have vague meanings to individuals, or whether people merely differ in their interpretation of probability values (Budescu & Wallsten, 1985).

In other research, it has been suggested that the meaning of a probability term can be represented by a membership function on the $[0,1]$ interval (Zadeh, 1978; Watson, Weiss & Donnell, 1979). A membership function indicates the degree to which an event, object, or value "fits" into a vague description or category. In the case of probability terms, a membership function would show how well each probability value is described by a particular term. These membership functions can take on any shape whatsoever, but continuous single-peaked or monotonic functions are the easiest to analyze and make the most sense conceptually.

The specific numerical probability value (or values) that are best described by a particular phrase would take on the maximum value of the membership function (usually one). Those probability values that are not at all described by a probability phrase would take on the minimum

value of the function (usually zero). Other probabilities with intermediate degrees of membership in the concept represented by the phrase would generally take on intermediate values between zero and one in this function.

Wallsten, et al, (in press) utilized a modified pair-comparison procedure to measure the vagueness of probability phrases. In this procedure, subjects judged pairs of probability values with respect to how much better one rather than the other was described by each of various probability terms. If the properties of the resulting data satisfy the axioms of an additive difference structure (Krantz, Suppes, Luce, & Tversky, 1971), then it is legitimate to use certain scaling procedures to derive the membership function for each probability term. These scaling techniques allow one to make predictions about subjects' responses in the pair-comparison task. It was found that observed and predicted responses were correlated in the range of 0.75 to 0.77. Independent judgments in a related modified pair-comparison task were also found to be predicted by the responses on the first task. Further, responses were found to be consistent for subjects over time (average correlation of 0.88 between sessions). The shapes and locations of the membership functions gathered in that study revealed large individual differences, however.

Factors Affecting the Meanings of Probability Terms

Environmental factors may play a large role in people's interpretations of probability phrases. The

research problem could be expressed in terms of determining the individual difference and context factors that affect the judged meanings of probability terms, regardless of the way in which the meanings are derived. Factors that might affect the meaning of a probability phrase to an individual include, among other things: (1) the reliability of the information source(s); (2) the number of sources of information; (3) the number of words in one's vocabulary of probability phrases; (4) general semantic context; (5) the importance of the event being analyzed; (6) the base rate of occurrence of the event. The effects of some of these context factors on probability or frequency phrases have been the focus of previous research (Pepper & Prytulak, 1974; Beyth-Marom, 1982; Wallsten, Fillenbaum & Cox, 1986).

This study is primarily concerned with a specific context effect -- the desirability of the outcomes of the event being analyzed by the decision-maker. Desirability is somewhat related to the importance of the event, although outcomes can have differing desirabilities while importance is held constant. It seems reasonable to assume that a continuum of desirability exists at each level of importance of events. For example, both flunking a test and nuclear war could be rated by a subject as highly undesirable, but certainly the latter event is more important. Importance was held constant in this study, while desirability was manipulated by the winning or losing

of points on each trial.

Outcome Desirability

The effects of outcome desirability on subjective probabilities under conditions of risk have been studied in depth (Marks, 1951; Scheibe, 1964; Morlock & Hertz, 1964; Pruitt & Hoge, 1965; Irwin & Snodgrass, 1966; Irwin & Metzger, 1966; Slovic, 1966; Irwin & Graess, 1968). In these studies, subjects' choices or estimates of the likelihoods of events were positively related to the independent values of the events. The independent outcome values (IOs) are to be distinguished from the dependent outcome values (DOs), which depend on the accuracy of the person's prediction.

Irwin & Snodgrass (1966) gauged subjective probabilities by means of a betting procedure, in which the frequency of subjects' positive bets was the dependent variable. Subjects made bets within a prescribed range on whether or not a marked card would be drawn from a deck with a known number of marked and unmarked cards. In addition to winning or losing the bet, subjects won or lost money if a marked card was drawn, regardless of their prediction. The bet thus determined the DO and the additional winnings or losses attributable to the drawing of a marked card became the IO. It was found that positive IOs led to a higher frequency of positive bets than did negative IOs. Further, this effect was strongest for objective probabilities (proportions of marked cards) closest to 0.50. In a related study, Irwin and Metzger

(1966) found that the positive relationship between IO and estimated likelihood was maintained even when IO was probabilistic. In this study, the IOs had well-defined positive or negative expected values, but on any given trial subjects did not know with certainty whether they would win or lose money for drawing a marked card.

A slightly different paradigm was utilized by Slovic (1966) to test the effect of desirability on judged likelihood. In his study, subjects gave direct numerical estimates of subjective probabilities under conditions of varying IO. Again, it was found that estimated likelihoods of events were positively related to desirabilities of IOs. Slovic termed this effect "partial optimism" because it related only to the sign of the IO, not to its magnitude. Also, Irwin & Graess (1968) were unable to find a consistent significant relationship between IO and choice probabilities for paired events when the IOs for the two events under consideration had the same sign, but different magnitudes. There was, however, a significant effect of IO magnitude at some levels of objective probability. Crandall, Solomon, & Kellaway (1955) found a significant overall interaction between IO and choice probabilities when both IOs had the same sign. When this effect was analyzed at each level of objective probability, however, it was not significant.

In all cases described above, the events for which subjects estimated probabilities were precisely defined.

) That is, the information on which subjects based their subjective probabilities was either full knowledge of the probability distribution of events or sample information relating to a specific probability distribution of events. Thus, subjects were responding under conditions of risk, as described by Ellsberg. If subjects had had only vague or ambiguous information on the distribution of outcomes, or perhaps no information at all (creating an Ellsberg paradox situation), their subjective probabilities might have been affected differently by IO desirability.

Purpose and Hypotheses

The main purpose of this experiment was to examine the effect of independent outcome desirability on judgments made under conditions of ambiguity and ignorance, as represented by verbal expressions. A secondary goal of the experiment was to determine if different probability phrases show different effects due to IO desirability.

) On every trial, subjects were given information about the likelihood of occurrence of two outcomes. Each outcome had a certain IO associated with it. For one outcome, subjects were given objective probability information, and for the other outcome, they were provided either ambiguous information (a probability term) or no information at all. The latter case represented a condition similar to the Ellsberg paradox. Subjects then indicated their degree of confidence that one outcome was more likely to occur than the other outcome. It was hypothesized that subjects'

1

responses in this experiment would be positively biased toward the outcome with the relatively favorable desirability factor. Further, this bias should be strongest in cases where an objective probability of 0.50 was matched with the probability terms, tossup and unspecified (the Ellsberg paradox condition), according to the results of Irwin & Snodgrass (1966). Thus, when the objective probability of 0.50 has associated with it a negative desirability and the probability phrase, unspecified, has associated with it a positive desirability, the verbal expression should be interpreted with a subjective value greater than that accorded the objective probability. This would be a contradiction of Ellsberg's contention that people's intuition usually leads them to accord a higher subjective value to an objective probability of 0.50 (risk) than to an event for which people were completely ignorant of probabilities. This bias due to desirability could be construed as: (1) a change in the judged meaning of a given probability phrase (perhaps a change in the shape or location of the membership function for that phrase), (2) a change in the subjective value accorded the objective value of 0.50, (3) a type of ambiguity aversion, or (4) some combination of the above.

In this experiment, one could have asked subjects to bet on either the risky event or on the event with the vaguely specified probability, while manipulating the desirability factor. The drawback to that design, however,

is that any effects due to desirability might be confounded with subjects' desire to avoid ambiguity in a gambling scenario. The present design controls for this factor by asking subjects to give their best judgments of the relative likelihoods of the two paired events, the consequences of which are independent of their judgments.

Method

Subjects

Twenty graduate students recruited with notices distributed in the social science departments, the Business School, and the School of Journalism at the University of North Carolina at Chapel Hill served as paid subjects. At the end of the experiment, a cash bonus of \$10 was paid to those ten people whose point totals were the highest.

Procedure

Subjects performed the experiment individually while seated in a small cubicle. Stimuli were presented on the color monitor of an IBM PC, and responses were made by means of a joystick connected to the computer.

The computer screen was, in effect, split in half from top to bottom. One side of the screen showed a probability spinner with a radial portion of the spinner colored white and the remainder colored red. Subjects were informed that the displayed proportion of white in the spinner represented the true probability of that spinner's landing on white.

On the other side of the screen, subjects received information relating to the likelihood of another spinner's landing on white, but were not able to actually view this other spinner. Instead, they saw one of six phrases. There were five probability terms (doubtful, improbable, tossup, likely, or good chance) which were each followed by the phrase, "... to land on white." These phrases described the likelihood of that spinner's landing on white. (Hereafter, this will be referred to as the invisible spinner.) In one condition, subjects saw the phrase, "Unspecified to land on white," which they understood to mean that absolutely no information was available about the invisible spinner. Except for the case involving the word, unspecified, the actual probability distribution that governed the likelihood of an invisible spinner's landing on white was represented by a single probability value that was based on previous research on the measurement of the meanings of probability terms (Wallsten, et al, in press). The probability terms used in this experiment were assigned probability values as follows: doubtful: 0.18; improbable: 0.32; tossup: 0.50; good chance: 0.64; and likely: 0.78. Unspecified was assigned a probability value of 0.50, based on the reasonable assumption that complete ignorance of probability information should lead one to assume an attitude of indifference between any two mutually exclusive and exhaustive outcomes.

On a single trial, a probability spinner appeared on a

1

randomly selected side of the screen, and one of the six phrases appeared on the other side. The proportion of white in the visible spinner was varied among four values uniquely predetermined for each of the six phrases with which it was matched. The phrases "likely to land on white" and "good chance to land on white" were matched with visible spinners that had proportions of white of 0.45, 0.62, 0.79, and 0.95. Phrases "doubtful to land on white" and "improbable to land on white" were matched with visible spinners that had proportions of white of 0.05, 0.22, 0.39, and 0.55. The phrases "tossup to land on white" and "unspecified to land on white" were matched with visible spinners with proportions of white of 0.30, 0.50, 0.50, and 0.70.

To incorporate the concept of outcome desirability in the experiment, each spinner's landing on white had associated with it an independent point value. These values were either +500, -500, or 0 points. In all cases, the value of a spinner's landing on red was 0 points. There were three different conditions of desirability presented. One condition involved a desirability factor of +500 points associated with the visible spinner and -500 points associated with the probability phrase. In the second condition the signs of these desirability factors were reversed. The third condition had desirability factors of 0 points associated with both the visible spinner and the probability phrase.

At the end of the experiment, subjects answered a few questions concerning their response strategies and their understanding of the instructions. This manipulation check consisted of a brief written multiple choice questionnaire.

The sequence of events on a single trial was as follows:

1. A randomly selected screen was displayed to the subject. On one side of the screen was a spinner and the desirability value associated with that spinner's landing on white. On the other side of the screen was one of the six probability phrases referring to the likelihood of an invisible spinner's landing on white and the desirability value associated with that spinner's landing on white. Traversing the width of the screen at the bottom was a line segment with an arrow positioned in the middle.

2. The subject indicated which spinner (visible or invisible) he or she judged to be more likely to land on white by using a joystick to position the arrow along the line segment. Moving the arrow toward one or the other end of the line segment indicated stronger confidence in the judgment that the spinner on that side of the screen would land on white. The arrow could be positioned at any one of 200 points along the line segment, but there were only three hash marks visible to the subject--one at each end and one in the middle of the line segment.

3. After the arrow was positioned, the subject pushed a button on the joystick assembly that caused the spinners to "spin". The subject did not actually see the spinners

spinning and was not told the outcome (white or red) of an individual spinner. However, a noise effect simulated spinning while the word "spinning" appeared at the top of each side of the screen. The outcome of each spinner was based on the probability distributions described above.

4. If either or both spinners landed on white, the corresponding desirability values were added to or subtracted from the subject's cumulative total of points. This total was displayed and updated on every trial in a window near the top of the screen.

5. The subject pushed a button on the joystick assembly to initiate the next trial.

Design

There were six probability phrases matched with four different probability spinners each. Also, each spinner-phrase pair had associated with it one of three conditions of desirability. This created a total of 72 trials which constituted a block. Subjects performed one block of trials in random order on each of two days. Sessions were separated by 24 to 48 hours. There was no practice, but subjects read detailed instructions while viewing the first randomly selected screen of their session.

Results

Reliability

The first step was to assess the reliability of the data by calculating the linear correlation between each subject's responses over the two sessions. Table 1 shows

both a simple Pearson correlation of each subject's responses on the two days, as well as a pooled correlation. Pooling was achieved by standardizing the responses so that they had a mean of zero for each of the six probability phrases. As can be seen in the table, this adjustment did not significantly change the correlations. The mean correlation over all subjects between responses for session one and session two was 0.61 for both the simple and pooled cases. The simple correlations were above 0.80 for four subjects and above 0.70 for ten subjects.

Analyses on Individual Phrases

The three main factors in the experimental design for each of the six probability phrases were (a) the probability of the visible spinner landing on white (four levels), (b) the outcome desirability combination (three levels), and (c) session (two levels). Figure 1 shows mean responses for each phrase as a function of probability, ignoring desirability level. Responses are scaled so that 0 means certain that the invisible spinner is more likely to land on white, 200 means certain that the visible spinner is more likely to land on white, and intermediate values designate intermediate degrees of certainty. As expected, responses increase with visible spinner probability. Mean responses at each desirability condition for each phrase, ignoring probability level, are shown as the first six rows of Table 2.

The main analyses sought to determine the effect, if

any, of the desirability values on subjects' responses. The initial analyses involved six distinct 4 x 3 x 2 (Probability x Desirability x Session) repeated measures ANOVAs, one for each probability phrase. The ANOVAs were performed in the manner recommended by Winer (1971), in which the error term for each effect was the interaction of that effect with the factor of subjects.

Subjects' responses, scaled such that a higher value of the response indicated a stronger confidence that the visible spinner would land on white, was the dependent variable. It was hypothesized that mean responses would increase as the level of spinner probability increased, as they do (see Figure 1). No other response pattern would be appropriate if the subjects believed that the spinners were fair. The main hypothesis of the experiment, however, was that mean response values would be lowest for the negative spinner/positive phrase desirabilities, and greatest for the positive spinner/negative phrase desirabilities.

The results of these analyses are shown in Table 3. Session was not a significant factor for any of the six probability phrases. As expected, the probability of the visible spinner landing on white was significant in the appropriate direction for all six probability phrases. From Tables 2 and 3, it can be seen that level of desirability was significant in the hypothesized direction for the probability phrases, doubtful, good chance, and likely. It was significant, but with somewhat unexpected mean response values, for the probability phrase,

improbable. Desirability was not significant, but had mean response values in the hypothesized direction for the probability phrases, tossup and unspecified.

Analyses on Phrase Pairs

In order to increase power, the phrases were combined into three pairs--high probability (good chance and likely), middle probability (tossup and unspecified), and low probability (doubtful and improbable)--and a 2 x 4 x 3 (Phrase x Probability x Desirability) ANOVA was performed on each pair. There was no hypothesized pattern of mean responses for the probability phrase factor, so it was not expected to be significant. The mean responses for each phrase pair at each level of desirability, ignoring probability level, are shown as the last three rows of Table 2. Mean responses for each phrase pair as a function of probability, ignoring desirability level, are plotted in Figure 2.

The results of these ANOVAs, shown in Table 4, reveal that probability phrase was not significant in any of the three ANOVAs, while the probability of the visible spinner landing on white was significant in the hypothesized direction in every case. Desirability was significant with the predicted response pattern for both the high and the low probability phrases, but not for the middle probability phrases. The interaction of probability with desirability was not significant in any of these analyses.

As can be seen in Figures 1 and 2, the probability

factor had a generally large effect. In contrast, however, when the IO desirability factor was significant, the effect was rather small, as shown in Table 2. In general, the high probability terms, good chance and likely, showed much larger effects of both desirability and objective probability than did the low probability terms, doubtful and improbable. The middle probability terms obviously showed small effects of desirability, since they were non-significant, but showed rather large effects of objective probability.

The Ellsberg Paradox

Ellsberg (1961) claimed that people commonly violate a fundamental axiom of rationality by preferring to bet on outcomes stated under conditions of risk rather than on outcomes stated under conditions of ignorance or high levels of ambiguity, even though the best estimate of the probabilities of all the outcomes under consideration was 0.50. This finding, known as the Ellsberg paradox, would be obtained in the present study if subjects accorded a higher subjective probability to the objective probability of 0.50 than to the phrases, tossup and unspecified, or in other words if subjects gave mean responses significantly above the neutral response of 100 when these two phrases were each matched with a visible spinner with a 0.50 probability of landing on white. In contrast, a major hypothesis of this study was that independent outcome desirability biases the subjective probability toward the outcome with the relatively favorable desirability, thereby

overwhelming any effects due to ambiguity avoidance. Simple t tests, comparing the appropriate mean responses to 100, are suitable for testing this hypothesis. The results of this analysis, shown in Table 5, reveal that only with the phrase, tossup, under negative spinner/positive phrase desirability, was the mean response significantly below 100. Thus, in this instance, IO desirability appeared to produce a bias toward the outcome with the relatively favorable desirability and the results predicted from the Ellsberg paradox failed to hold. Notably, however, in no case were the results expected on the basis of those from Ellsberg obtained at a significant level.

Analysis of the Questionnaire Data

The results of the questionnaire given to subjects at the end of the experiment showed that eighteen out of twenty subjects understood that the spinners were fair and that their responses had no effect on the probability of either the visible or invisible spinner landing on white. Six questions, phrased like the first one in the Appendix, but with a slightly different situation in each case, were used to assess subjects' understanding of this point. Four questions, phrased like the second one in the Appendix, revealed that nine subjects out of the twenty thought that their responses had some effect on whether or not they won points. Additionally, a final question, which was identical to the third question in the Appendix, verified subjects' understanding of the two main ideas of the

instructions, and explicitly asked subjects to explain their response strategy when presented with the probability term, unspecified.

Given the above results, it was decided to perform separate analyses on the eleven subjects who seemed to understand the instructions perfectly and on the nine who seemed not to. The $2 \times 4 \times 3$ ANOVAs, as described above for the three pairs of probability phrases, were performed separately for the two groups of subjects, for a total of six analyses, with results shown in Table 6.

Probability phrase was significant only for the phrase pair, tossup and unspecified, for those who understood the instructions, while the probability of the visible spinner landing on white was significant in the predicted direction in every case. Desirability was significant in the hypothesized direction for the high probability phrase pair, good chance and likely, for both groups of subjects. For the low probability phrase pair, doubtful and improbable, desirability was significant in an inappropriate pattern for those subjects who did not understand the instructions, and was not significant, but in the hypothesized direction for those subjects who did understand. Desirability was not significant for either group of subjects for the middle probability phrase pair, tossup and unspecified. The subjects who did not understand the instructions had mean responses in the appropriate pattern, however, while those subjects who did understand the instructions had an inappropriate pattern of

mean responses. The interaction of probability and desirability was not significant in any of these cases.

Reanalyses on Most Reliable Subjects

Analyses, similar to the ones described above, were performed on the ten subjects with between session correlations of responses at or above the median level of 0.70. Individual phrase analyses showed that, for every phrase, the probability of the visible spinner landing on white was significant in the hypothesized direction. However, only for the phrase, likely, was the spinner desirability effect significant, and this effect was not precisely in the hypothesized direction. No other effects or interactions were significant.

For the analyses on phrase pairs, the probability of the visible spinner landing on white was significant in the hypothesized direction in every case. For the low probability phrases, spinner desirability was significant in the hypothesized direction, while for the high probability phrases, it was significant, but not exactly in the hypothesized pattern. Spinner desirability was not significant for the middle probability phrases. In short, these analyses did not produce any better results than had been obtained previously on all twenty subjects.

Discussion

In general, the results show that IO desirability affects judgments of relative probability when the

uncertainty is represented by verbal expressions. The probability phrases, doubtful, good chance, and likely, showed this effect clearly when analyzed separately. The phrase, improbable, was also significantly affected by desirability, but did not show precisely the hypothesized pattern of responses. The middle phrases, tossup and unspecified, showed the expected pattern of mean responses, but the results were not statistically significant.

The results of the analyses on pairs of phrases showed that desirability was significant in the hypothesized direction for the high pair, good chance and likely, as expected. The increased power of this analysis also produced a significant desirability effect in the appropriate direction for the low pair, doubtful and improbable. The middle pair, tossup and unspecified, still failed to achieve significance for the desirability effect, although the pattern of responses was in the hypothesized direction.

Differential Effects

The results of this experiment indicate a clear difference in the effects of desirability on the low, middle, and high probability phrases. Any explanation of the overall effect of desirability on judgments under ambiguity and ignorance must take this finding into account.

Judged meanings of phrases. The low phrases, doubtful and improbable, showed smaller effects of both objective probability and desirability than did the high

phrases, good chance and likely. Other research (Wallsten, Fillenbaum, & Cox, 1986) has also shown that low probability phrases are less affected by context than are neutral and high probability phrases. Context in that case referred to the base rate of occurrence of events. A possible reason that context has less effect on low than on high phrases might be that the meanings of low phrases are further from the center of the probability interval than are the meanings of high phrases. This would give the meanings of high probability phrases more freedom to undergo shifts when influenced by factors such as desirability or objective probability value.

The data from this experiment are not suitable for precisely determining the best judged meanings of the phrases. By examining the data for each subject averaged over all levels of desirability, however, one can note at what approximate point the phrase and probability were judged equal (i.e. where the mean response was 100). Out of 40 such determinations for the low terms, 26 were located below the probability value of 0.22. Eleven of these 26 determinations were for doubtful and the remaining 15 were for improbable. In contrast, the high terms had only 3 out of 40 determinations above the probability value of 0.79, all three occurring for the phrase, good chance. This indicates that the best judged meanings for high probability phrases are less extreme, on average, than they are for low probability phrases.

Precision of meaning. A reason for the lack of significant results for the middle probability terms may stem from the relative precision of their meanings. Tossup was the most specific of all the terms used in this study, with a relatively precise meaning of 0.50. In contrast, unspecified would seem to be the least specific of all the terms used in this experiment. However, the results of the questionnaire (see Appendix, question 3) indicated that unspecified was also treated as having a relatively precise meaning of 0.50 probability. Thirteen of the twenty subjects explicitly mentioned the 0.50 probability in the visible spinner as the cutoff point for deciding whether the visible or the invisible spinner was more likely to land on white. The meanings of both tossup and unspecified may therefore have been more impervious to the effects of independent outcome desirability than the meanings of the other phrases used in this study, resulting in the small, non-significant results reported here.

Unfortunately, the above result and interpretation also imply that the phrase, unspecified, may not truly represent Ellsberg's condition of "complete ignorance". It may be more closely related to a condition of ambiguity.

Shifts in subjective value. Previous research (Irwin & Graee, 1968; Morlock & Hertz, 1964; Crandall, Solomon, & Kellaway, 1955) showed that IO desirability caused a significant shift in the subjective values accorded to objective probabilities, with the largest shifts found near the middle of the probability range. In the present study,

however, this pattern of results did not occur. Note that it cannot be determined whether the significant effects of desirability found at the low and high ends of the continuum are due to shifts in the subjective values accorded the visible probabilities or to changes in the meanings of the phrases. However, the lack of any significant effects due to desirability in the middle range of probability would seem to indicate that only very small shifts, if any, occurred in the subjective probability accorded to the visible spinner in that range.

That the shifts in subjective value accorded the visible spinner probability were not significant (for the middle range of probability, at least) may relate to the fact that this experiment is fundamentally different from the earlier studies of independent outcome desirability. That is, subjects were never before required to compare the relative likelihoods of probabilities and probability phrases. The task of making this type of comparison may interact with desirability or produce effects that add to those of desirability. One such effect, discussed further below, is ambiguity aversion. For example, since there is some evidence that the middle probability phrases were the most precise in meaning, and thus the least ambiguous, it should not be surprising if ambiguity aversion effects were smallest for these terms.

Asymmetry of effects. One interesting facet relating to this interpretation of the results is that the IO

desirability effect was generally not symmetric. That is, for four of the six phrases (all except doubtful and unspecified), the effect was greater for negative spinner/positive phrase desirability than it was for the opposite type of desirability. This implies that either the judged meaning of the phrase increased more, or that the subjective value accorded the objective probability decreased more for this condition than for the opposite one. One interpretation of this result is that people normally have a bias against ambiguity, and towards firm probabilities, as previous research has shown (Ellsberg, 1961; Becker & Brownson, 1964; Yates & Zukowski, 1976; Einhorn & Hogarth, 1986). Further evidence for this point is given in Table 2, where the mean responses at zero desirability are all greater than 100, indicating an avoidance of ambiguity. Therefore, attempts to make the ambiguity-avoidance bias stronger will show fairly small effects. In contrast, attempts to induce the opposite type of bias may show larger effects.

Membership functions. Another explanation for the results of this experiment incorporates the concept of membership functions. There are two commonly encountered shapes of membership functions for probability terms: (a) single-peaked, and (b) monotonic (Wallsten, et al, in press; Rapoport, et al, in press). Figure 3 shows some examples of the different types of membership functions one might encounter in this experiment.

In order to link the results from this experiment to

the concept of membership functions one must posit some fairly strong assumptions. First, it must be assumed that unique membership functions exist for each probability phrase and that these functions are generally of a monotonic or single-peaked shape. Secondly, it must be assumed that these membership functions have a maximum, which may occur at either a single point or over a range of values. The third assumption, which is the most tenuous, is that in judging how much more likely one spinner rather than the other is to land on white, the subject focuses only on how well the displayed objective probability is described by the phrase associated with the invisible spinner. In other words, the response depends only on the value of that objective probability in that phrase's membership function. Thus, if an objective probability with the value 1.0 in a phrase's membership function were paired with that phrase, the subject would consider both the visible and invisible spinners equally likely to land on white, and would leave the arrow in the middle of the response line.

Recall that the response line is scaled 0 (definite that the invisible spinner is more likely to land on white) to 200 (definite that the visible spinner is more likely to land on white). Given the above assumptions, responses can be transformed to membership function values by the following formula:

$$m(p) = 1 - [(100 - x) / 100] ,$$

where $x \in [0, 200]$ is the response.

There are other assumptions which give a different interpretation to the best judged meaning of a probability phrase. For example, it has been suggested that subjects may think of the low, central, and high probability values for a given phrase and then perform a type of weighted average of these values to derive their best judged meaning of the phrase (Wallsten, 1986). Depending on the weights given to each of these probability values, the best judged meaning may or may not occur at the peak of the membership function for the given phrase. This will be discussed in more detail below.

Given the three assumptions initially described above, from each of the low, middle, and high probability phrase pairs, 40 (20 subjects by 2 phrases) membership functions can be generated. Ignoring the desirability factor temporarily, and using mean responses averaged over all levels of desirability, it was found that the vast majority (89%) of the derived membership functions were single-peaked. These results are shown as the last three columns in Table 7.

Next, the responses were examined at each level of desirability to determine if the desirability factor had an effect on the shape of subjects' membership functions. These results, displayed in the first nine columns of Table 7, show that there was no clear effect on the shape of subjects' membership functions due to desirability, since the total number of monotonic membership functions under

each level of desirability was about the same (12 [10%], 11 [9%], and 16 [13%]). The zero desirability condition had the fewest number of other-shaped membership functions, but this could simply mean that the desirability factor introduced more variability in subjects' responses. Individual subjects did have differently shaped functions depending on desirability level, however, which accounts for the discrepancy between this analysis and the preceding one, in which a total of only seven monotonic membership functions were found.

One finding that is the same in both analyses and consistent with other research (Wallsten, et al, in press) is that monotonic functions are more prevalent among the lower probability phrases than among the higher ones. However, the number of monotonic membership functions found in this study is much smaller than has been normally found in other research (Wallsten, et al, in press; Rapoport, Wallsten, & Cox, in press). This may indicate that mean responses of 100 do not represent the peak values of the membership functions, but rather the weighted average of low, central, and high probability values relevant to a given probability phrase. Under this interpretation, it is possible that many of the membership functions labeled as single-peaked could actually be monotonic, given the appropriate set of weights used in the averaging process. These weights are unknown parameters, however, which cannot be determined from the present data.

Based on the assumptions initially listed above, desirability seemed to have no effect on the shapes of the derived membership functions. Thus, it is hypothesized instead that IO desirability caused a shift in the location of the membership function for a given term. When the desirability was positive for the spinner and negative for the probability phrase, the membership function was shifted to the left of where it was when there was no desirability (toward lower probability values). When the desirability was negative for the spinner and positive for the probability phrase, the membership function was shifted to the right of where it was when there was no desirability (toward higher probability values). Figure 4 shows this shift for the three main types of membership functions that were encountered in this experiment.

Using this model, one can explain the effect of desirability on the probability terms, doubtful, good chance, and likely, as consisting of a significant shifting of the location of the membership functions for these terms. For the term, improbable, the shift was not completely made for the case of positive spinner desirability. A possible reason for this is that the membership function for improbable under conditions of no desirability may be located near the bottom of the range of probability. Thus, there may not have been enough room for a significant shift to the left, as predicted for the positive spinner/negative phrase desirability effect. The data show that the mean responses for improbable were

higher than those for the other low probability term, doubtful, when responses were averaged over all factors, and also when these responses were averaged over each level of desirability. These results imply that the meaning of improbable was the lowest, among all the phrases used in this study, in terms of subjective probability. This gives some credence to the notion that the membership function for improbable may be so close to the bottom of the range of probability values that it precludes a significant shift toward lower probabilities. However, the finding that improbable had a lower judged meaning than did doubtful is inconsistent with previous research (Wallsten, et al, in press).

Alternative Explanation

A different interpretation of the results of this experiment involves the use of the weighted average of the low, central, and high probability values for a given phrase (cf. Wallsten, 1986). The pattern of results for the desirability factor can be explained by simply assuming that positive spinner/negative phrase desirability causes subjects to give more weight to the low probability values, while negative spinner/positive phrase desirability causes subjects to give more weight to the high probability values. In other words, when the phrase describes a potential loss, subjects concentrate more on the low probability for that phrase, and when it describes a potential gain, they concentrate more on the high

probability for that phrase. The different desirability effects for low, middle, and high probability phrases can be explained by assuming that the weights used in the averaging process for the middle probability terms were all about equal, while for the low probability phrases they were slightly different from each other, and for the high probability phrases they were even more different from each other. Of course, further research is necessary to independently test these assumptions.

The Ellsberg Paradox

Tests of the Ellsberg paradox showed an effect opposite to Ellsberg's hypothesis under one condition of desirability. The one significant result in this analysis (See Table 5) showed that when desirability favored the phrase, subjects considered the ambiguous uncertainty (the phrase, tossup) to be more likely than the precise one (the objective probability of 0.50). This is consistent with the hypothesis that subjects' responses will be biased towards the outcome with the relatively favorable desirability.

Further, it is of interest that the results predicted from the Ellsberg paradox did not occur in any case. That is, subjects never accorded a significantly higher subjective value to the objective probability of 0.50 than to the phrases, tossup and unspecified. Possibly, Ellsberg was dealing with betting preferences, while the present study deals with judgments of relative likelihood. Thus, the Ellsberg paradox may actually represent a type of

ambiguity aversion in a gambling scenario.

If the term, unspecified, truly represented a condition of ignorance, then desirability did not exert a strong enough effect to cause subjects to consider the condition of ignorance as more likely than the condition of risk. Given the previous discussion of the fairly precise meaning of unspecified, however, it may be that this term represented ambiguity rather than ignorance.

Differences from Previous Research

There are some important differences between these results and those found by earlier studies. First, it was previously found that the effect of IO was strongest for objective probabilities of 0.50 (Irwin & Snodgrass, 1966; Morlock & Hertz, 1964; Crandall, Solomon, & Kellaway, 1955). It was therefore hypothesized that the strongest effect of IO desirability in the present study would be found in the case where the probability 0.50 was paired with the phrase, tossup. This turned out to be one of the weakest effects in the experiment, although it was in the appropriate direction. Another interesting difference was that the high terms seemed to show stronger effects of desirability than did the low terms. This result did not occur in the earlier research under conditions of risk. Both of these differences may be the result of this experiment's requiring a judgment of relative likelihood between a verbal expression of ambiguity and an objectively stated probability, whereas, the earlier studies dealt

solely with objective probabilities. One finding from the earlier research on independent outcome desirability that was replicated was that the desirability effect was maintained even though the independent outcomes were probabilistic (cf. Irwin & Metzger, 1966).

It turned out that the nature of the present task was problematic. Subjects were asked to concentrate on making their best judgment as to the relative likelihoods of the two paired events on each trial. They were told that the points were put on the screen to help them pay attention to the task and to make it more interesting. Despite the written instructions, which indicated that all outcomes were based solely on fixed, predetermined probability values, some subjects mentioned after the experiment that they tried to figure out what "system" governed the awarding of points. The questionnaire given to subjects at the end of the experiment was an attempt to determine how deeply this tendency affected subjects' responses. As pointed out above, however, the questionnaire did not seem to provide any valid information on this issue.

Summary and Conclusions

This experiment showed that independent outcome desirability is one context factor that appears to affect the vague meanings of probability phrases. Judgments of relative likelihood were found to be biased somewhat toward the outcome with the relatively favorable desirability. Given certain assumptions, independent outcome desirability

4

effects can be interpreted as causing shifts in the location of the membership functions for probability phrases, without changing the shape of the membership functions. A different explanation could involve the way in which desirability affects the weighted averaging process used to derive the best judged meaning for a probability phrase.

The most plausible explanation of the IO desirability effects would need to involve all the factors relevant to a judgment of relative likelihood between a probability phrase and an objective probability. Three of these factors, discussed above, are: (a) the subjective value that is accorded to the objectively stated probability, (b) the judged meaning of the probability phrase, and (c) ambiguity aversion. It would be desirable to study each of these factors in isolation within the context of examining the meanings of probability phrases.

References

- Becker, S. W., & Brownson, F. O. (1964). What price ambiguity? Or the role of ambiguity in decision making. Journal of Political Economy, 72, 62-73.
- Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.
- Budescu, D. V. & Wallsten, T. S. (in press). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Performance.
- Crandall, V. J., Solomon, D., & Kellaway, R. (1955). Expectancy statements and decision times as functions of objective probabilities and reinforcement values. Journal of Personality, 24, 192-203.
- Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. Psychological Review, 92, 433-461.
- Einhorn, H. J. & Hogarth, R. M. (in press). Decision making under ambiguity. Journal of Business.
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. Quarterly Journal of Economics, 75, 643-669.
- Fellner, W.. Distortion of subjective probabilities as a reaction to uncertainty. Quarterly Journal of Economics, 75, 670-689.

Hakel, M. (1968). How often is often? American Psychologist, 23, 533-534.

Irwin, F. W. (1944). The realism of expectations. Psychological Review, 51, 120-126.

Irwin, F. W., & Graess, C. M. (1968). Tests of the discontinuity hypothesis of the effects of independent outcome values upon bets. Journal of Experimental Psychology, 76, 444-449.

Irwin, F. W., & Metzger, M. J. (1966). Effects of probabilistic independent outcomes upon predictions. Psychonomic Science, 5, 79-80.

Irwin, F. W., & Snodgrass, J. G. (1966). Effects of independent and dependent outcome values upon bets. Journal of Experimental Psychology, 71, 282-285.

Knight, F. H. (1921). Risk, Uncertainty, and Profit. Boston: Houghton Mifflin Company.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). Foundations of Measurement, Vol. I. New York: Academic Press.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science, 9, 563-564.

Marks, R. (1951). The effect of probability, desirability, and "privilege" on the stated expectations of children. Journal of Personality, 19, 332-351.

Morlock, H. C., & Hertz, K. J. (1964). Effect of the desirability of outcomes on decision making. Psychological Reports, 14, 11-17.

Pepper, S., & Prytulak, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. Journal of Research in Personality, 8, 95-101.

Pruitt, D. G., & Hoge, R. D. (1965). Strength of the relationship between the value of an event and its subjective probability as a function of method of measurement. Journal of Experimental Psychology, 69, 483-489.

Radner, R. (1972). Normative theory of individual decision. in McGuire, C. B. & Radner, R., eds.. Decision and Organization. Amsterdam: North Holland.

Rapoport, A., Wallsten, T. S., & Cox, J. A. (in press). Direct and indirect scaling of membership functions of probability phrases. Mathematical Modeling.

Savage, L. J. (1954). The Foundations of Statistics. New York: Wiley.

) Scheibe, K. E. (1964). The effect of value on statements of expectancy under four experimental conditions.

The Psychological Record, 14, 137-144.

Simpson, R. H. (1944). The specific meanings of certain terms indicating differing degrees of frequency. The Quarterly Journal of Speech, 30, 328-330.

Simpson, R. H. (1963). Stability in meanings for quantitative terms: A comparison over 20 years. The Quarterly Journal of Speech, 49, 146-151.

Slovic, P. (1966). Value as a determiner of subjective probability. IEEE Transactions on Human Factors in Electronics, HFE-7, 22-28.

Stone, D. R., & Johnson, R. J. (1959). A study of words indicating frequency. Journal of Educational Psychology, 50, 224-227.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (in press). Measuring the vagueness of probability terms. Journal of Experimental Psychology: General.

Wallsten, T. S., Fillenbaum, S., & Cox, J. (1986). Base rate effects on the interpretations of probability and frequency expressions. Journal of Memory and Language, 25, 571-587.

) Wallsten, T. S. (1986). National science foundation grant

proposal.

Watson, S. R., Weiss, J. J., & Donnell, M. C. (1979).

Fuzzy decision analysis. IEEE Transactions on Systems, Man, and Cybernetics, SCM-9, 1-9.

Winer, B. J. (1971). Statistical Principles in Experimental Design, 2nd ed.. New York: McGraw-Hill.

Yates, J. F., & Zukowski, L. G. (1976). Characterization of ambiguity in decision making. Behavioral Science, 21, 19-25.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-38.

Zimmer, A. C. (1983). Verbal vs. numerical processing of subjective probabilities. In R. W. Scholz (Ed.), Decision Making Under Uncertainty. Amsterdam: North Holland Publishing Co., pp. 159-182.

Table 1

Test-Retest Correlations

Subject	Simple	Pooled
1	0.69	0.69
2	0.84	0.83
3	-0.21	-0.16
4	0.68	0.69
5	0.75	0.76
6	0.86	0.86
7	0.47	0.46
8	0.69	0.69
9	0.81	0.81
10	0.93	0.92
11	0.77	0.76
12	0.71	0.72
13	0.72	0.71
14	0.42	0.42
15	0.76	0.76
16	0.62	0.63
17	0.79	0.80
18	0.54	0.53
19	0.68	0.67
20	0.24	0.22
Mean	0.61	0.61

Table 2

Desirability Effects in Terms of Mean Response Values
for Each Phrase and Combined over Similar Phrases

Spinner = Phrase =	-500 +500	0 0	+500 -500
Doubtful	108.47	112.09	118.11
Good Chance	97.77	110.68	115.56
Improbable	111.21	120.24	118.51
Likely	94.99	112.93	115.98
Tossup	96.34	101.39	103.44
Unspecified	102.68	103.32	107.36
Doubtful & Improbable	109.84	116.16	118.31
Good Chance & Likely	96.38	111.80	115.77
Tossup & Unspecified	99.51	102.36	105.40

Table 3

F-Ratios from ANOVAs on Individual Probability Phrases

Phrase	Sessions ^a	Probability ^b	Desirability ^c
Doubtful	0.86	63.51 **	5.02 *
Good Chance	0.00	73.13 **, .	5.67 **
Improbable	0.62	51.37 **	5.53 **
Likely	1.14	61.28 **	22.90 **
Tossup	0.01	76.60 **	1.38
Unspecified	1.08	36.11 **	0.47

a) df = (1,19)

b) df = (3,57) for Doubtful, Good Chance, Improbable, and
Likely

df = (2,38) for Tossup and Unspecified

c) df = (2,38)

* $p < 0.05$. ** $p < 0.01$.

Table 4

F-Ratios from ANOVAS on Combined Probability Phrases

Phrase Pair	Phrases ^a	Probability ^b	Desirability ^c	Probability ^d x Desirability
Doubtful & Improbable	3.39	66.94 **	6.81 **	0.78
Good Chance & Likely	0.00	78.46 **	13.71 **	1.89
Tossup & Unspecified	3.84	65.69 **	1.24	2.08

a) df = (1,19)

b) df = (3,57) for Doubtful & Improbable, Good Chance &
Likely

df = (2,38) for Tossup & Unspecified

c) df = (2,38)

d) df = (6,114) for Doubtful & Improbable, Good Chance &
Likely

df = (4,76) for Tossup & Unspecified

* $p < 0.05$. ** $p < 0.01$.

Table 5

Tests of Whether Mean Responses are Significantly
Below 100 for the Neutral Phrases vs. 0.50

Spinner =	-500	0	+500
Phrase =	+500	0	-500
Tossup			
Mean Response	94.06	97.58	103.93
<u>T</u> -Value	-2.01 *	-0.79	1.26
Unspecified			
Mean Response	98.00	99.61	105.06
<u>T</u> -Value	-0.48	-0.11	1.37

* $p < 0.05$.

Table 6

F-Ratios from ANOVAs for Subjects who
Understood/Did not Understand the Instructions

Phrase Pair	Phrase	Probability	Desirability	Probability x Desirability
Subjects who Understood the Instructions				
Doubtful & Improbable (df)=	0.79 (1,10)	47.62 ** (3,30)	3.25 (2,20)	1.23 (6,60)
Good Chance & Likely (df)=	0.81 (1,10)	153.03 ** (3,30)	7.41 ** (2,20)	1.40 (6,60)
Tossup & Unspecified (df)=	5.86 * (1,10)	49.60 ** (2,20)	2.48 (2,20)	0.31 (4,40)
Subjects who did Not Understand the Instructions				
Doubtful & Improbable (df)=	4.75 (1,8)	24.34 ** (3,24)	4.21 * (2,16)	0.52 (6,48)
Good Chance & Likely (df)=	2.51 (1,8)	14.67 ** (3,24)	5.95 * (2,16)	1.14 (6,48)
Tossup & Unspecified (df)=	0.02 (1,8)	19.50 ** (2,16)	2.29 (2,16)	0.93 (4,32)

* $p < 0.05$. ** $p < 0.01$.

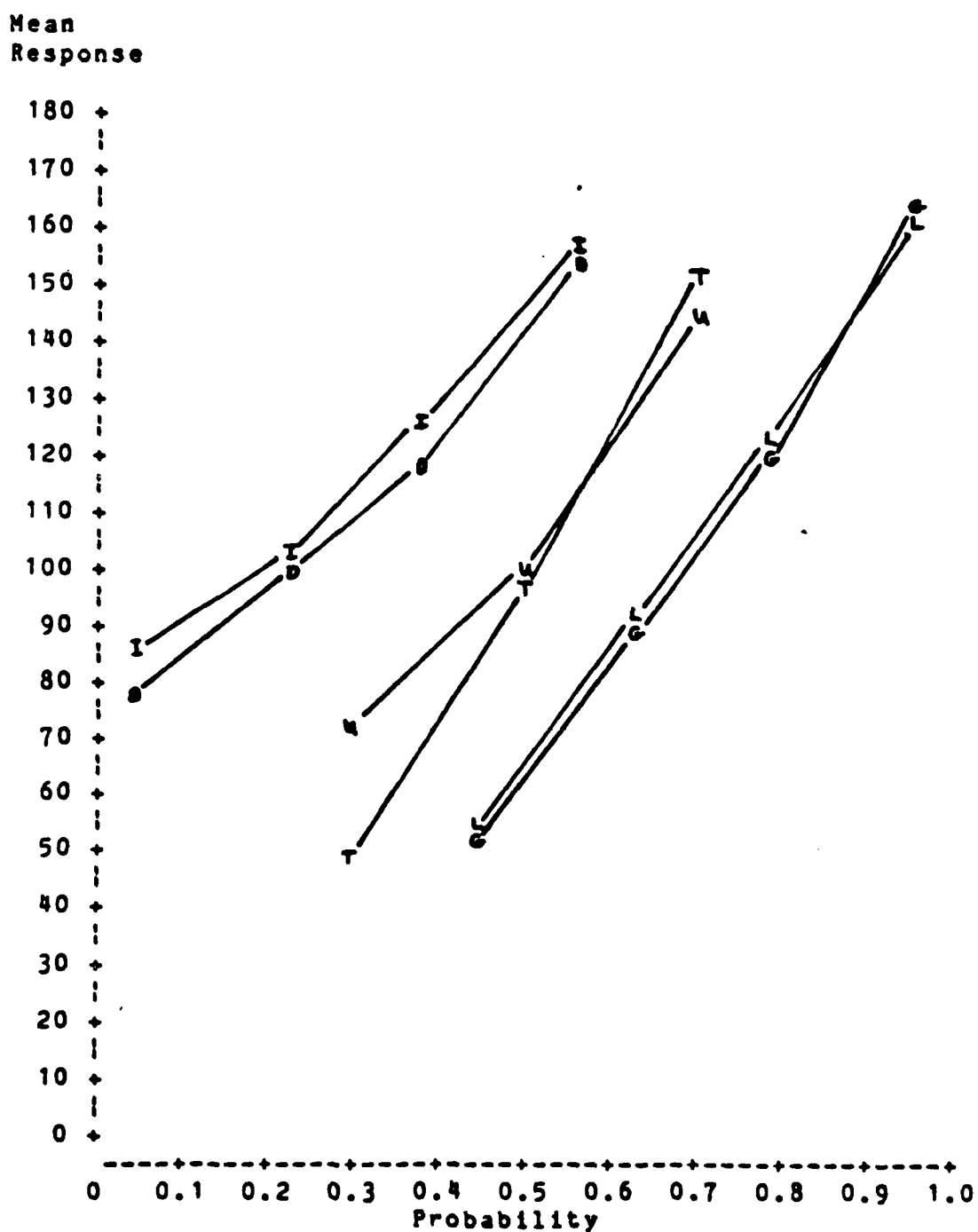
Table 7

Monotonic (M), Single-Peaked (SP), and Other (O)
 Membership Functions for Each Phrase and Desirability Level,
 and Averaged over All Desirability Levels

Spinner =	-500			0			+500			Overall		
Phrase =	+500			0			-500					
	M	SP	O	M	SP	O	M	SP	O	M	SP	O
Low phrases												
Doubtful	3	14	3	2	14	4	1	13	6	2	17	1
Improbable	3	11	6	5	12	3	6	12	2	3	16	1
Sub-Total	6	25	9	7	26	7	7	25	8	5	33	2
Middle phrases												
Tossup	2	15	3	1	17	2	2	15	3	0	20	0
Unspecified	2	14	4	2	17	1	6	14	0	2	18	0
Sub-Total	4	29	7	3	34	3	8	29	3	2	38	0
High phrases												
Good Chance	2	16	2	1	18	1	0	14	6	0	18	2
Likely	0	16	4	0	18	2	1	17	2	0	18	2
Sub-Total	2	32	6	1	36	3	1	31	8	0	36	4
Totals	12	86	22	11	96	13	16	85	19	7	107	6

Figure 1

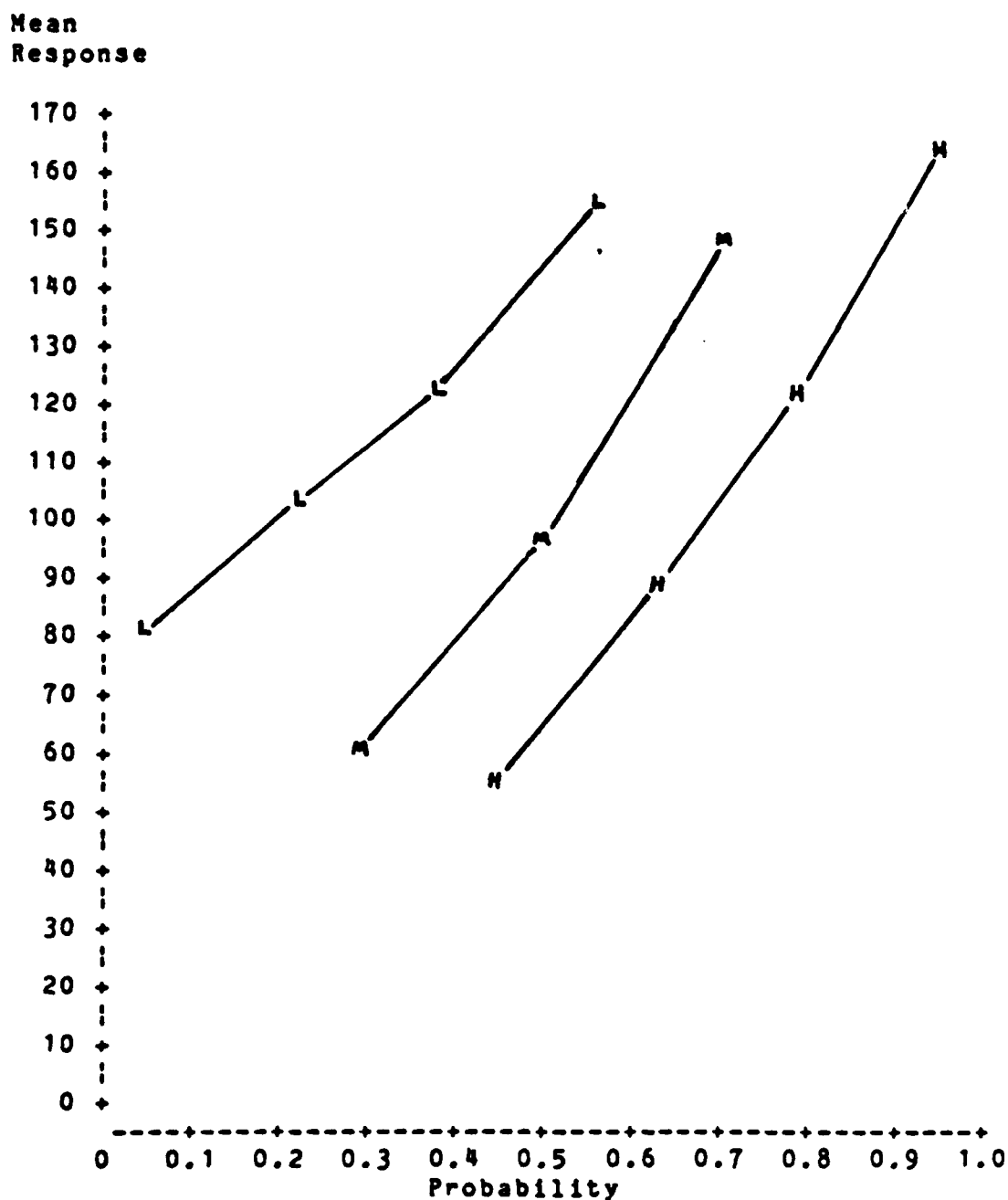
Graph of Mean Responses as a Function of Probability for
Individual Phrases, Ignoring Desirability



D = doubtful T = tossup G = good chance
I = improbable U = unspecified L = likely

Figure 2

Graph of Mean Responses as a Function of Probability for
Phrase Pairs, Ignoring Desirability

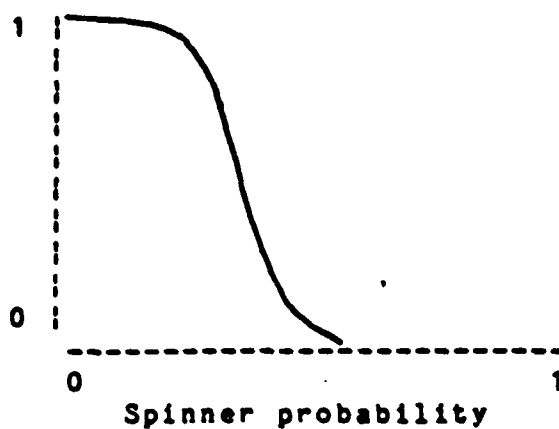


L = Low probability phrases (doubtful & improbable)
M = Middle probability phrases (tossup & unspecified)
H = High probability phrases (likely & good chance)

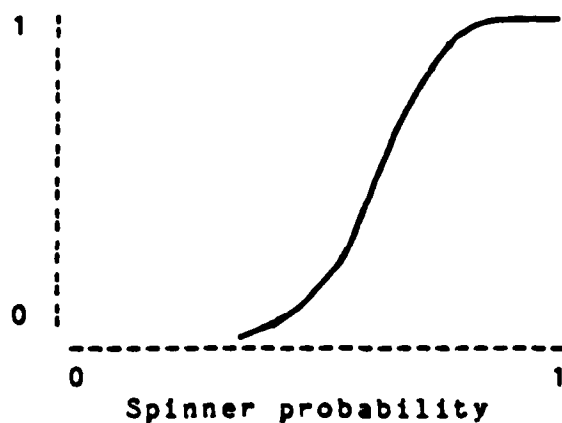
Figure 3

Types of Membership Functions

1. Monotonic decreasing



2. Monotonic increasing



3. Single-peaked

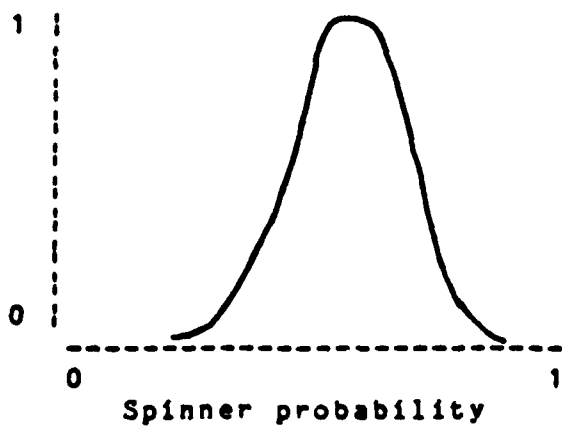
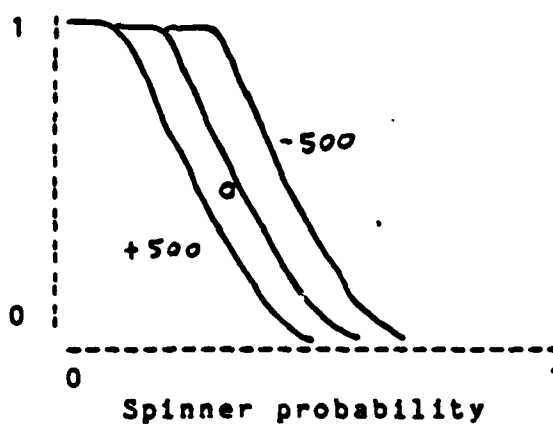


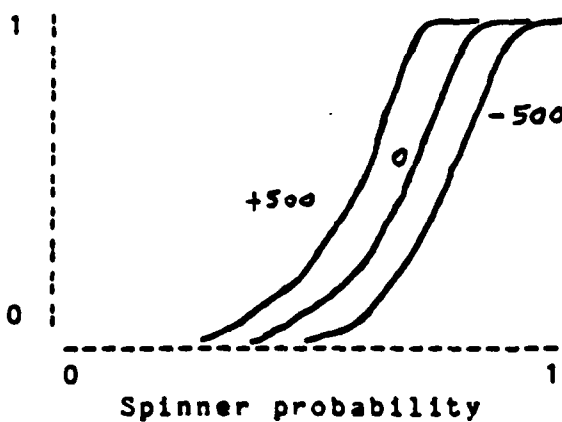
Figure 4

Hypothesized Effect of Independent Outcome Desirability
on Membership Functions

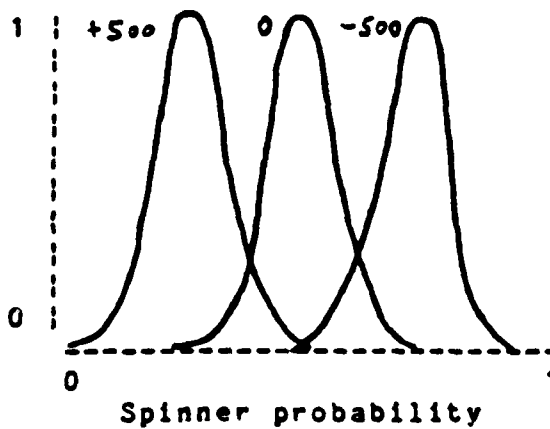
1. Monotonic decreasing



2. Monotonic increasing



3. Single-peaked



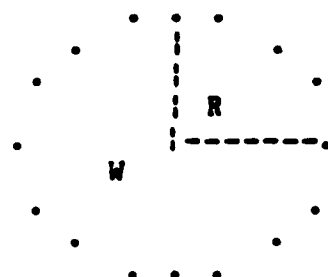
Appendix

1)

WHITE: -500
RED: 0

WINNINGS:
1000

WHITE: 500
RED: 0



LIKELY
to land on
white.

✓

----->

In the screen drawn above, the visible spinner is on the left and the word, Likely, is on the right. Inside the spinner, the letters W and R stand for the colors white and red, respectively. Suppose that you viewed this screen during the experiment and you positioned the arrow approximately where it is in the drawing above.

Given that you have the arrow positioned on the right side of the line segment, what effect does this have on the likelihood of the visible spinner's landing on white?

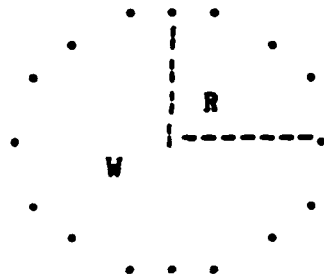
- A) It increases the likelihood of the visible spinner's landing on white.
- B) It decreases the likelihood of the visible spinner's landing on white.
- C) It has no effect on the likelihood of the visible spinner's landing on white.

2)

WHITE: -500
RED: 0

WINNINGS:
1000

WHITE: 500
RED: 0



LIKELY
to land on
white.

\\

----->

In the screen drawn above, the visible spinner is on the left and the word, Likely, is on the right. Inside the spinner, the letters W and R stand for the colors white and red, respectively. Suppose that you viewed this screen during the experiment and you positioned the arrow approximately where it is in the drawing above.

Now, it turned out that both the visible spinner and the invisible spinner landed on white. In this case, describe what happens to your cumulative total of points:

- A) It increases by 500 points.
- B) It decreases by 500 points.
- C) It increases by 1000 points.
- D) It decreases by 1000 points.
- E) Nothing happens to your cumulative total of points.

i 3) a) In general, did you think that where you placed the arrow influenced the outcome of the spinners?

b) Did you think that your placement of the arrow had an effect on the amount of points that you could win or lose on a given trial?

c) What guided your placement of the arrow when you were presented the word, "unspecified?"

)