ELICITING SUBJECTIVE PROBABILITY UNDER UNCERTAINTY:

AN ECONOMIC EXPERIMENT

By

Zhijun Yang

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Agricultural Economics
in the Department of Agricultural Economics

Mississippi State, Mississippi

August 2006

ELICITING SUBJECTIVE PROBABILITY UNDER UNCERTAINTY:

AN ECONOMIC EXPERIMENT

By

Zhijun Yang

Approved:

Keith H. Coble
Professor of Agricultural Economics
(Director of Thesis)

Stan R. Spurlock
Professor of Agricultural Economics and Graduate Coordinator of the Department of Agricultural Economics

M. Darren Hudson
Professor of Agricultural Economics
(Committee Member)

Terrill R. Hanson
Associate Professor of Agricultural Economics
(Committee Member)

Vance H. Watson
Dean of the College of Agriculture and Life Science

Name: Zhijun Yang

Date of Degree: August 5, 2006

Institution: Mississippi State University

Major Field: Agricultural Economics

Major Professor: Dr. Keith H. Coble

Title of Study: ELICITING SUBJECTIVE PROBABILITY UNDER UNCERTAINTY: AN ECONOMIC EXPERIMENT

Pages in Study: 115

Candidate for Degree of Master of Science

This study uses a combination of statistical sampling techniques and a web-based economics experiment to evaluate hypotheses with regard to eliciting and combining subjective probability estimates. Demographic characteristics of participants including age, statistical background, risk preference, and Myers-Briggs Type Indicator have been collected.

Results show that knowledge is the core of experts' decision support system. Knowledge sharing and feedback make the subjective probability estimates more reliable. The Extroversion, Intuition, Thinking, and Judging type's people are the best to perform subjective probability assessment.

Finally, analyses show that weighting scheme and risk preference do not influence the forecast accuracy.

ACKNOWLEDGMENTS

I am heartily thankful for the support and kindness that was extended by all the people in the Department of Agricultural Economics at Mississippi State University in the last two years.

I want to express my gratitude to my Major Advisor, Dr. Keith H. Coble for his great mentoring and guidance in my study. Expressed appreciation is also going to my committee members, Dr. Darren Hudson and Dr. Terrill R. Hanson, for their invaluable aid and instruction.

Special thanks go to Dr. Stan R. Spurlock and Dr. Steven C. Turner for their support in my application for a PhD program.

Finally, I thank Mr. Vikram Venkatesh, Ms. Shalini Batra, and Mr. Amit Uppal for their great efforts to design the computer programming software and to make the economic experiment a great success.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Uncertainty is a description of the imperfect knowledge of the true value of a particular variable (Hogarth 1987). Decision makers often face a situation in which uncertainty exists. Managing uncertainty is an unavoidable challenge in a variety of decision contexts. In general, uncertainty can be reduced by supplementing the decision process with pertinent knowledge. In this case, if empirical data are available, statistics can be used to generate additional information. If objective data are absent, counselors or experts are commonly used as a source of subjective probability estimates about the variable of interest.

The following is an interesting example from history. In the year 431 B. C., the Athenian general Perikles rendered a speech to his soldiers before a battle in the war between Athens and Sparta that started in 431 B.C.

> We ourselves either ratify or even propound successful policies, finding harm not in the effect of speeches on action but in failing to get instruction by speech before proceeding to what must be done. For in that we are both especially daring and especially thorough in calculating what we attempt, we can truly be distinguished from other men, for whom ignorance is boldness but calculation brings hesitancy. Rightly would they be judged strongest in spirit who recognize

> both dangers and pleasures with utmost clarity and are on neither count deterred from risks. (Thucydides, 431 B.C.)

In modern society, rational decisions usually have been made either implicitly or explicitly with the inclusion of uncertainty. For example, in the case of a horse race, some gamblers do not have any information about the horses or the track, but they are willing to place their bets on which horse will win. It is the degree of belief, or subjective probability that really measures the perceived risk of the variable of interest. This concept has been widely employed in the research of decision theory, business management, public policy, and engineering safety. In agriculture, this concept could be seen in a variety of applications including food safety analysis, disease spread modeling, and financial management.

General Problem

Even when facing a lack of knowledge, people have to make decisions and take actions. The Athenian general Perikles had to optimize his decisions to win the battle against the enemy. His counselors would have provided their best ideas to assist him in making the best possible decisions. Several thousand years have passed, yet the difficulties decision makers experience when they struggle with uncertainty remain. A good case to illustrate this situation is space exploration. Human knowledge about space is limited. But it is the character of exploring outer space which is full of uncertainty that prompts human beings to find a way to deal with uncertainty.

> Probabilities are propagated through the logic models to determine the probability that a system will fail....Probability data may be derived from available empirical

data....If quantitative data are not available, then subjective probability estimates may be used....(NASA Reference Publication 1358: System Engineering "Toolbox" for Design-Oriented Engineers).

Specific Problem and Objectives

Three commonly used methods to deal with the elicitation and combination of expert opinions regarding uncertain future events are: the composite method, the composite with feedback method, and the consensus method. The composite method means each expert makes his judgment individually and their individual responses are aggregated into a composite in some manner. The composite with feedback means first each expert makes his judgment independently, then each opinion will be shared with others and the expert could revise his judgment accordingly. The consensus method means a group of experts attempt to reach a general statement which reflects their collective judgment.

However, there is little scientific basis to compare the accuracy of the above mentioned methods. A primary limitation has been the inability to compare results from the elicitation and combination procedure with a measurable criterion. At the core of this problem is the lack of known probabilities with which to compare with elicited predictions.

The first objective of this thesis is to compare the effectiveness of these three methods by using a controlled economic experiment. In this experiment, data are drawn from four sets of known probability distributions. Participants then assess the data and individual expert's opinions are elicited. The aggregate opinion will be achieved by applying each method. The results will be compared with the known probability distributions and the effectiveness of each method will be evaluated. Hypothetically, composite with feedback and consensus methods will perform better in terms of generating subjective probability.

The second objective of this thesis is to analyze if knowledge, personality type, and risk aversion affect the accuracy of individual expert's opinion under uncertainty. Given that, uncertainty is a property of human being's knowledge about a certain statement or event, not of the statement or event themselves. Then, subjective probability is the degree of belief that a certain statement is true or that some event will occur. Subjective probabilities are useful numbers with which to measure uncertainties. Theoretically, whether and how these two psychological factors (personality type and risk aversion) influence the accuracy of subjective probability are not clear.

Subjective probability assessments of potential events (e.g., bioterrorist attack, heart disease, Asian soybean rust, nuclear radiation) provide key elements for decision making. It is meaningful to identify if there are better elicitation and aggregation methods in terms of forecast accuracy. If so, better decisions could be derived. Agricultural economists and others interested in risk management should benefit from the work of this thesis.

CHAPTER II

LITERATURE REVIEW

As stated in Chapter I, the general populace, and decision makers in particular, often place great weight on experts' opinions. However, due to the complex, subjective nature of expert opinion, there has been no formally established methodology for treating expert judgment (Ouchi, 2004). More importantly, there is a growing body of evidence that expert opinion can be a useful source of data. The use of expert opinion is critical and often inevitable in the areas where no other source of empirical data on which to base probability estimates is available. However, the proper use of this source requires new techniques due to the heuristic errors made by experts.

The Stanford/SRI Assessment Protocol developed by the group of analysts who operated in the Department of Engineering-Economic Systems at Stanford University and at the Stanford Research Institute during the 1960s and 1970s, was widely regarded as the initial effort in establishing a systematic approach to eliciting expert opinion. A summary of this protocol can be found in a paper by Spetzler and Stael von Holstein (1975). Recently, Cooke and Goossens (2000) provide formal protocols, comprehensive procedures and guidelines on the elicitation process.

Once the elicitation procedure is completed, experts' opinions will be combined and a final probability estimate will be available to support the decision procedure.

Ouchi (2004) points out that there are extensive studies on how to reconcile multiple experts' probability assessments and many of which are extensively reviewed by Genest and Zidek (1986), Cooke (1991), Clemen and Winkler (1999), and Bedford and Cooke (2001). However, no one combining method is reportedly superior.

This chapter will examine the research on how to obtain the ideal result in eliciting expert opinion in a structured way and how to aggregate different experts' opinions.

Heuristic Procedures

People who think about and make judgments under uncertainty usually make use of a set of heuristic procedures (Morgan and Henrion, 1990). These procedures usually lead to biased outcomes or even outright errors. Kahneman et al. (1982) offer a beginning point for a more thorough exploration of this subject. Three heuristics are discussed: availability, anchoring, and representativeness.

Availability

When asked to estimate the probability of occurrence of a specific event, subjects tend to base their estimates on the ease with which they can think of or imagine previous occurrences of the event. Tversky and Kahneman (1982a) conducted the following experiment. In the experiment subjects were given the following text:

Consider the two structures A and B which are displayed below:

(A)	(B)
XXXXXXXX	XX
XXXXXXXX	XX
XXXXXXXX	XX
	XX
	XX
	XX
	XX
	XX
	XX

A path in a structure is a line that connects an element in the top row with an element in the bottom row, and passes through one and only one element in each row. In which of the two structures are there more paths? How many paths do you think there are in each structure?

In the experiment, 46 of the 54 subjects thought there were more paths in (A). In fact, there are 8^3 paths in (A) and 2^9 paths in (B); and $8^3 = 2^9 = 512$.

Why do people see more paths in (A) than in (B)? Kahneman and Tversky speculate that it is much easier to imagine paths through three points than paths through nine points, hence the paths in (A) are more easily envisioned.

Anchoring

When asked to estimate a probability, subjects usually begin with an initial value (usually the most likely value) and then make adjustments for its minimum and maximum from that first value. Frequently the adjustment is insufficient and the estimator appears to be anchored to the first estimated value.

Representativeness

A representativeness heuristic is a form of stereotyping by which people tend to emphasize some particular similar information rather than integrating information from all sources. In general, representativeness comes about from the corresponding tendency to undervalue or discard other evidence (Kahneman and Tversky, 1982a).

It is clear that human judgments about uncertainty frequently rely on a number of cognitive heuristics. These, in many cases, entail violations of normative dictates and hence produce systematic departures from rational judgments. From this standpoint, a structured protocol should be developed to deal with such errors in the eliciting procedure.

Performing Probability Assessment

In order to prevent subjective bias due to heuristics as much as possible, the need to perform subjective probability assessment in a structured approach is recognized. Two protocols have been established in an effort to formalize the procedure of eliciting expert opinion.

The Stanford/SRI Assessment Protocol

Five stages are highlighted in this protocol: motivating, structuring, conditioning, encoding, and verifying.

During the motivating phase, the analyst develops some initial rapport with the expert. The reason for the elicitation is discussed and the basic idea of probabilistic assessment is explained and justified. Then an examination of the possibility that the expert's opinion does not fully reflect his/her true beliefs should be carried out. If a significant possibility of error is found, it may be possible to overcome it, either by changing the incentive structure the expert faces, or by disaggregating the assessment task in such a way as to require judgments in which the error is less likely to occur (Morgan and Henrion, 1990).

The second phase involves structuring the uncertain quantity to be elicited. The objective is to let the expert clearly understand the definition of the quantity to be assessed and allow the expert to provide reliable judgments.

During the conditioning phase of the protocol, the objective is to get the expert "conditioned to think fundamentally about the judgment and to avoid cognitive biases" (Spetzler and Stael von Holstein, 1975).

The fourth phase of the protocol involves the actual encoding of the expert's probabilistic judgment. Spetzler and Stael von Holstein offer the following specific guidelines:

> Begin by asking the subject for what he considers to be extreme values for an uncertain quantity. Then ask for scenarios that might lead to outcomes outside of these extremes. The deliberate use of availability is designed to counteract the bias that is otherwise likely to occur. Next take a set of values and use the probability wheel (a perfectly balanced wheel divided into two segments, shaded and white, which can be used to quantify the probability of uncertain events) to encode the corresponding probability levels. Don't choose the first value in a way that may seem significant to the subject, otherwise you may cause him/her to

anchor on that value. Make the first few choices easy for the subject so that he/she will be comfortable with the task (Spetzler and Stael von Holstein, 1975).

During the final verifying phase of the protocol, the objective is to test the quantitative judgment the expert has provided to see if it, in fact, correctly reflects his/her beliefs. The results can be plotted both as a CDF and as a PDF and discussed with the expert. If disagreement between the expert's views and the elicited distributions are found, the analyst should go back through the appropriate phase of the protocol in order to make a correct elicitation.

Cooke and Goossens' Protocol

Cooke and Goossens (2000) provide a formal protocol on the elicitation process. Examples are shown from the EC/USNRC joint study on Probabilistic Accident Consequence Uncertainty Analysis. There are 15 steps in this protocol.

1. Definition of case document describing the field of interest for which expert judgments will be required.
2. Identification of target variables which are the variables whose uncertainty must be quantified through formal expert judgment.
3. Identification of query variables: in case the target variables can not be quantified by direct elicitation, the query variable derived from the target variable should be found.
4. Identification of performance variables to be assessed by the experts.
5. Identification of experts.
6. Selection of experts.

7. Definition of elicitation format document describing the exact question and format for the expert elicitations.
8. Dry run exercise describing the try out of the elicitation format document to a few experts.
9. Expert training session describing the ingredients of training experts in preparing probabilistic assessments.
10. Expert elicitation session, in which the experts' individual judgments are discussed in the presence of a normative analyst (experienced in probability issues) and a substantive analyst (experienced in the expert's field of interest).
11. Combining experts' opinions and describing the methods with which the individual expert opinion will be aggregated to one combined assessment.
12. Robustness and discrepancy analyses and describing the procedures to show the robustness of the combined results.
13. Feed back communication with the experts.
14. Post-processing analyses and describing the methods for processing the uncertainties of the combined expert assessments (from query variables to target variables).
15. Documentation of the results.

Aggregating Expert Opinion

Cooke (1991) provides three well-established mathematical modeling approaches to aggregating expert opinion: non-Bayesian axiomatic models, Bayesian models, and psychological scaling models. Excellent summaries also can be found in Genest and Zidek (1986). The following review is based on Cooke's (1991) framework.

Non-Bayesian Axiomatic Models

Suppose we have experts 1,2,....e, and each expert i gives a probability vector $P_{i1}, P_{i2},...P_{in}$ for the elements $A_1, A_2,...A_n$ of some partition of the set S of possible worlds. Let $w_1, w_2,...w_e$ be nonnegative weights that sum to unity.

Elementary r- norm

Weighted mean: $$M_r(j) = (\sum_{i=1}^{e} w_i P_{ij}^{\ r})^{1/r} \qquad (2\text{-}1)$$

r-norm probability: $$P_r(j) = \frac{M_r(j)}{\sum_{j=1}^{n} M_r(j)} \qquad (2\text{-}2)$$

where:

$r = 1 \ \Rightarrow\ P_r =$ weighted arithmetic mean of P_i,

$r = 0 \ \Rightarrow\ P_r =$ weighted geometric mean of P_i,

$r = -1 \Rightarrow\ P_r =$ weighted harmonic mean of P_i.

Bayesian Models

In this method, the decision maker uses experts' opinions as data to update his own prior belief concerning the distribution of an unknown quantity of interest, according to Bayes' Theorem. This model is very useful in repeated experiments.

Psychological Scaling Models

In these models, the decision maker asks experts to state their preference on pair-wise comparisons. The advantages of these models are:

1. They require only qualitative input from the experts.
2. They lead automatically to a sort of consensus estimate with confidence bounds if simulation is used.

The disadvantages are:

1. A large number of experts are required.
2. The models make very strong assumptions regarding the experts' psychological assessment mechanisms.

The Delphi Approach

There are many studies regarding subjective probability from the perspective of individual judgments, but there are few from the perspective of collective judgments. In reality, established methods are generally based on experience, but not validated scientifically. One widely used method is the Delphi approach. The Delphi approach regards human subjective probability estimates as legitimate and powerful elements in generating forecasts. A single expert often produces biased opinions due to heuristic procedures. But, a group of experts meeting together can suffer from the dominant figure influence and the monitor's individual intention. To overcome these deficiencies, the Delphi approach was developed in the 1950s at the RAND Corporation as an attempt to make the best forecast under a less than perfect kind of knowledge.

The procedure of the Delphi approach is as follows:

1. A group of experts is identified and each expert provides his individual opinion with regard to a set of questions.
2. The opinions of each expert are collected by the monitor. Extreme opinions are thrown away. A preliminary general opinion (consensus) is formulated by the monitor.
3. The preliminary general opinion is delivered to each expert for examination. The expert can revise his/her individual opinion.
4. The second round expert opinions are collected by the monitor and second round general opinion will be composed and delivered to each expert again for comments. This process will be repeated until a general opinion has been accepted by all or most experts in the group.

In the past century, the Delphi approach was widely applied in industry, government, and academic research. However, the mechanics underlying the Delphi approach are still largely based on intuitive consistency rather than forecast accuracy. Additionally, little research has been conducted as to how characteristics of individuals affect their forecast accuracy.

CHAPTER III

CONCEPTUAL FRAMEWORK

Three different methods (composite method, composite method with feedback, and consensus method) which will be used to elicit the expert's subjective probability have been described in Chapter I. The critical question is whether these different elicitation methods result in different predictions, and, if so, which method generates more accurate results.

Two hypotheses will be tested in a controlled economics experiment.

1. Hypothesis One: forecast accuracy of aggregated subjective estimates will depend on different methods, different underlying probability distributions, and different aggregation methods.
2. Hypothesis Two: forecast accuracy of individual subjective estimates will depend on different elicitation methods, different underlying probability distributions, individual knowledge, personality type, and risk preference.

The following sections justify these hypotheses.

Knowledge

Knowledge is defined as the range of one's information or understanding (Merriam-Webster's Collegiate Dictionary, 2001). The definition of information is knowledge obtained from investigation, study, or instruction (Merriam-Webster's Collegiate Dictionary, 2001). An expert is a subject who is knowledgeable about the field of interest and is widely recognized as a qualified candidate to make subjective assessments regarding unknown variables of interest. Without extra information, the expert under uncertainty should derive his/her conclusion from his/her experience that may be reflected in his/her subjective judgment. It is commonly assumed that a knowledgeable person will tend to make reliable decisions with regard to the problems within his/her field. This is the foundation of using experts as a means of assistance to decision making.

Either different information or different abilities to process the same information can cause subjective probabilities to vary among individuals (Pindyck and Rubinfeld, 2001). So, combining a group of experts' opinions by some aggregation scheme makes it clear that the information possessed by the group of experts has been exploited as much as possible. Given that, knowledge is inherently an important attribute in eliciting expert opinion. The hypothesis is knowledge has a positive relationship to the reliability of subjective probability judgment.

In the framework of composite method (Treatment 1), the expert will exercise his/her own individual judgment in the economic experiment. No further information will be provided other than his/her own knowledge domain.

In the framework of composite with feedback method (Treatment 2), the expert initially will make his/her individual judgment in the first round. Then starting from the second round, each expert can access the other group members' responses and he/she has the chance to revise his/her judgment.

In the framework of consensus method (Treatment 3), all procedures are the same as those in composite with feedback method except that there will be an opportunity for experts to communicate with each other and an incentive encouraging them to reach a consensus.

From the perspective of knowledge upon which an individual expert bases his/her subjective probability, the scope of knowledge has gradually expanded from Treatment One to Treatment Two and Treatment Three. It is hypothesized that the reliability of subjective estimates will improve accordingly.

Risk Preference

It is widely accepted that people's preferences toward risk affect decision making. Under uncertainty, which means people are facing many possible outcomes with unknown likelihoods, most people find risk undesirable, but some people find it more undesirable than others. Not everyone displays risk aversion behavior; some indicate risk seeking behavior. The risk seeker might be eager to enter into a gamble while an individual who is risk averse may be willing to buy insurance with the expectation that the potential risk could be taken by someone else. Thus, risk preference is a fundamental psychological element in standard theories of decision science, asset valuation, contracts, and insurance.

Probabilities can be used to describe the likelihoods that possible outcomes will occur. Some probabilities can be deducted through observation; e.g., if you flip a coin, the probability that it ends up with tail is 0.5. This can be verified through a large number of repeated flips. However, in many cases, likelihoods of possible outcomes can not be deduced in this way.

For example, Asian soybean rust is a plant disease that has reduced yields and raised production costs for soybeans and other legumes in every major production region of the world – except the United States before 2005. However, on April 27, 2005 Asian soybean rust was confirmed in Seminole County, GA. An outbreak could pose economic risks for producers and consumers, and affect agricultural programs, such as crop insurance, commodity programs, research and extension, and pesticide regulations. Now, what is the likelihood Asian soybean rust will occur in Mississippi in 2006? Since no other source is available, a panel of experts could be the only source to provide a reliable estimate. In this case, subjective probability should follow the same two important probability rules as objective probability does:

1. The probability of each potential outcome is between 0 and 1.
2. The sum of the probabilities of all potential outcomes is equal to 1.

Generally, experts cannot be assumed to have the same risk preference. So the relationship between risk preference and the accuracy of the subjective estimate should be investigated.

The widely used approach to modeling behavior under uncertainty is the expected utility approach which was first advocated by Von Neumann and Morgenstern in 1944.

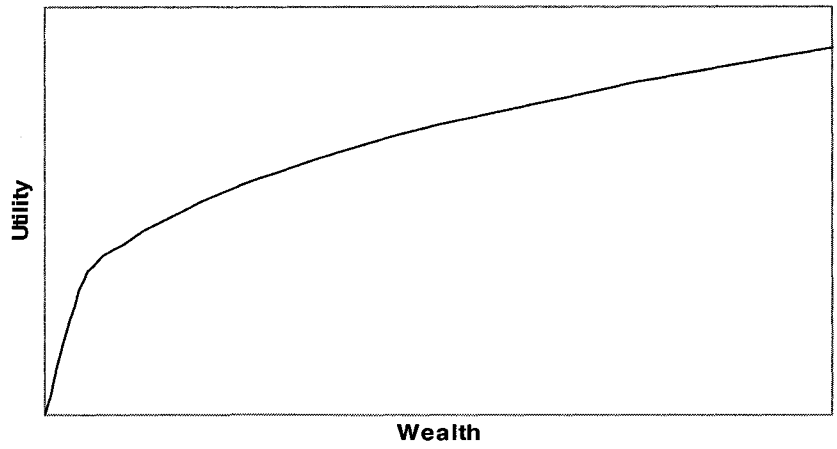

Figure 3.1 Utility Function for a Risk Averse Decision Maker

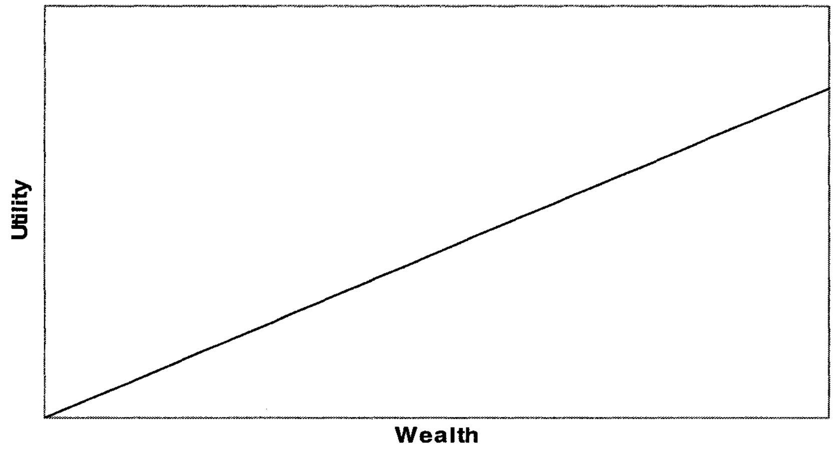

Figure 3.2 Utility Function for a Risk Neutral Decision Maker

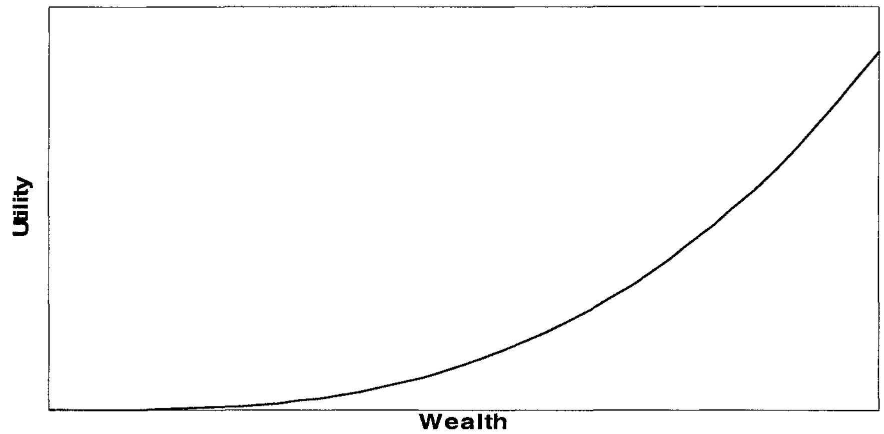

Figure 3.3 Utility Function for a Risk Loving Decision Maker

Figure 3.1 shows that the risk averse decision maker prefers a certain wealth to a probability weighted expected utility with the same expected wealth value. Figure 3.2 shows that the risk neutral decision maker is indifferent between a certain wealth and a probability weighted expected utility with the same expected wealth value. So as long as the expected value is equal, people will ignore the presence of uncertainty. Figure 3.3 shows that the risk loving decision maker prefers a probability weighted expected utility to a certain wealth with the same expected wealth value.

Arrow (1971) and Pratt (1964) defined the quantitative measures of risk aversion as follows:

Absolute risk aversion coefficient:

$$AR = -\frac{U''(W)}{U'(W)} \tag{3-1}$$

Relative risk aversion coefficient:

$$RR = -\frac{U''(W)}{U'(W)} \times W \tag{3-2}$$

where:

W = wealth,

$U(W)$ = the Von Neumann Morgenstern utility function,

$U'(W)$ = the marginal utility of wealth,

$U''(W)$ = the rate of change of marginal utility with respect to wealth.

A Von Neumann Morgenstern utility function such as

$$U(W) = \frac{W^{1-r}}{1-r} \tag{3-3}$$

exhibits constant relative risk aversion (CRRA) because $RR = r$. Practically, CRRA is convenient since the initial wealth W does not influence the risk preference.

There are several approaches to determine risk aversion. However, one of the most widely accepted approach to estimate the risk aversion coefficient has been done by Holt and Laury in 2002. In a paper published in the *American Economics Review*, they conducted an economic experiment where the decision makers were presented with different levels of money rewards and the probabilities of different level of money rewards were specified. Their findings provide a useful approach to identify each expert's risk preference in an economic experiment.

Myers-Briggs Type Indicator

Another psychological element assumed to have some kind of influence on human subjective judgment is personality type. The idea of personality type was initially advocated by Swiss psychologist Carl G. Jung in 1920s. Isabel B. Myers who established the famous Myers-Briggs Type Indicator instrument made this idea widely acceptable to the world. "Each of us is born with different gifts, with unique imprints of how we prefer to use our minds and values and feelings in the business of living everyday" (Myers, 1980). The latest 1998 publication of Form M has been proven very useful in identifying different personality types and has been warmly welcomed around the world.

When people take the MBTI assessment, they are evaluated on four dichotomies, each of which is made up of two opposite personality characteristics:

1.	Where you focus your attention	Extroversion (E) vs. Introversion (I)
2.	The way you take in information	Sensing (S) vs. Intuition (N)
3.	The way you make decisions	Thinking (T) vs. Feeling (F)
4.	How you deal with the outer world	Judging (J) vs. Perceiving (P)

In total there are 16 possible types of personality. Each type has a different behavior indication. For example, type ENFP means people tend to relate easily to the outer world, tend to use their imagination to see new possibilities and insights, tend to base decisions on values and people-centered concerns, and tend to not want to miss anything; life is likely to be spontaneous and flexible.

Hypothetically, knowledge will be positive related to the forecast accuracy of subjective probability estimates. However, how people's risk preference and personality type affect the forecast accuracy of subjective probability estimates are not known. In a controlled economic experiment, their relationship will be examined.

CHAPTER IV

METHODS

In Chapter III, two hypotheses were put forward to shed light on the mechanics underlying elicitation and combination of subjective probabilities. In a controlled economic experiment, a web-based software was developed and incentive-compatible reward was applied to participants in order to induce rational decision-making. This chapter describes the experimental designs and the methods used to analyze the collected data.

Participants

The minimum qualification for a student to participate was junior college standing with three credit hours of statistics. A total of 105 participants were recruited from various departments at Mississippi State University. Each of them took part in one of three economic experiments: (1) Treatment 1 (T1, composite method), (2) Treatment 2 (T2, composite with feedback), and (3) Treatment 3 (T3, consensus method). Among them, 35 students participated in Treatment 1, 25 students participated in Treatment 2, and 45 students participated in Treatment 3. The core task in the economic experiment is to conduct a subjective analysis regarding a known population distribution, where a randomly selected sample dataset with a fixed number of observations was available for review. Many studies involving laboratory behavior of both students and relevant professionals show that performances across these two groups do not vary substantially. For example, commodity traders tend to produce price bubbles due to over speculation, as do college students (Smith, Schanek, and Williams, 1988). Some environmentalists were found to free ride in a manner not much different from college students (Mestelman, Stuart, and Feeny, 1988).

Four Datasets and Four Known Population Distributions

In subjective probability analysis, a primary limitation to validating alternative subjective elicitation and aggregation procedures has been the inability to compare results from the elicitation and aggregation procedure with a measurable criterion. At the core of this problem is the lack of known probabilities with which to compare with elicited predictions.

The salient characteristic of this economic experiment is using datasets drawn from known population distributions. Each participant has to estimate the known population distribution based on a sample dataset. The estimate will be used to compare with the known population distribution and the accuracy of the subjective estimate will be examined.

Four different datasets were randomly generated. Each of them was randomly drawn from one known population distribution (see Table 4.1 through 4.8).

Table 4.1 Dataset One (30 observations)

Number of observations		Data
1		95.49652
2		80.83475
3		103.6639
4		119.1471
5		117.9753
6		125.997
7		67.24619
8		96.48728
9		116.4253
10		83.69949
11		89.64694
12		74.64352
13		72.29634
14		85.33556
15		88.39739
16		68.23103
17		91.48113
18		93.93929
19		102.0228
20		94.51761
21		95.09514
22		94.44639
23		120.1396
24		98.72073
25		97.20764
26		92.30189
27		129.5832
28		112.9851
29		135.6348
30		90.1764

Table 4.2 Known Population Distribution underlying Dataset One
Normal Population Distribution N (100, 15)

Scope	$(-\infty, 80)$	[80, 90)	[90, 100)	[100, 110)	[110, 120)	$[120,+\infty)$
Probability	0.0912	0.1613	0.2475	0.2475	0.1613	0.0912

Table 4.3 Dataset Two (50 observations)

Number of observations		Data	
1		103.2595	109.1557
2		91.14425	90.85654
3		160.2977	84.23164
4		79.70133	94.85669
5		128.8921	117.345
6		100.1444	104.0985
7		99.83075	100.0144
8		69.02864	107.9974
9		89.77836	132.9955
10		93.19357	82.88541
11		109.0556	114.1397
12		79.64545	81.69324
13		100.5307	84.36209
14		86.60144	127.1054
15		86.71169	69.48088
16		93.75552	92.41214
17		110.4863	108.3728
18		145.9224	124.87
19		128.3762	94.84305
20		143.3972	87.764
21		51.73448	
22		72.38638	
23		98.71329	
24		71.19346	
25		106.0513	
26		71.72557	
27		112.5951	
28		178.1583	
29		70.4105	
30		59.8507	

Table 4.4 Known Population Distribution underlying Dataset Two
Normal Population Distribution N (100, 25)

Scope	(-∞, 80)	[80, 90)	[90, 100)	[100, 110)	[110, 120)	[120,+∞)
Probability	0.2119	0.1327	0.1554	0.1554	0.1327	0.2119

Table 4.5 Dataset Three (100 observations)

Number of observations		Data			
1		107.6057	97.07202	106.3382	95.28591
2		99.24693	92.70128	99.57109	103.3386
3		82.08696	115.6485	89.28499	123.4948
4		108.9669	103.9824	102.2287	93.28413
5		94.35936	105.5851	88.96719	93.75972
6		92.52946	94.8845	111.3243	111.9039
7		100.3546	104.243	115.0832	97.13982
8		94.70228	97.02328	92.41164	94.19717
9		102.3316	105.112	100.8859	99.24846
10		100.9773	93.70666	105.5271	100.6503
11		97.74773	108.5273	97.76577	
12		101.5643	103.9551	99.48761	
13		102.4134	102.2578	105.4684	
14		90.43132	90.11081	96.65414	
15		100.5967	94.32165	114.1611	
16		98.48452	118.2707	98.32802	
17		103.2731	101.2829	101.2922	
18		88.72495	84.29139	93.12997	
19		94.60522	108.9121	99.75632	
20		104.6646	81.32489	110.6473	
21		105.8645	99.64306	105.5556	
22		71.09408	103.2207	82.56117	
23		99.03885	105.6102	103.0231	
24		88.21797	96.44385	115.9061	
25		105.44	112.8254	130.6216	
26		106.7264	105.5253	102.8148	
27		93.24477	110.7804	95.8897	
28		92.83343	102.4907	93.95649	
29		91.35662	101.1812	89.98276	
30		114.1361	100.3868	108.9978	

Table 4.6 Known Population Distribution underlying Dataset Three
Normal Population Distribution N (100, 10)

Scope	$(-\infty, 80)$	[80, 90)	[90, 100)	[100, 110)	[110, 120)	$[120, +\infty)$
Probability	0.0228	0.1359	0.3413	0.3413	0.1359	0.0228

Table 4.7 Dataset Four (30 observations)

Number of observations		Data
1		94.7
2		124.193
3		109.687
4		109.233
5		125.832
6		52.688
7		101.464
8		80.546
9		109.123
10		129.598
11		54.721
12		124.257
13		84.482
14		98.753
15		102.397
16		103.663
17		91.25
18		94.199
19		96.611
20		78.672
21		58.359
22		82.414
23		81.527
24		83.515
25		74.61
26		123.381
27		74.945
28		122.768
29		110.273
30		95.421

Table 4.8 Known Population Distribution underlying Dataset Four
Beta Population Distribution B (6, 2) with range of (0, 133.33)

Scope	(-∞, 80)	[80, 90)	[90, 100)	[100, 110)	[110, 120)	[120,+∞)
Probability	0.1624	0.1203	0.1653	0.199	0.2043	0.1487

Experiment Design

The economic experiment consists of four parts (see Appendix A for details of the experiment procedure).

Part One: The Myers-Briggs Type Indicator Survey

This part was done on paper. Each participant was given a copy of MBTI Booklet Form M (copyrighted by CPP, Inc., CA) and an answer sheet. Participants were told there are no standard answers in this survey and their responses did not account for their final payments. This information was used to identify individual's personality type.

Part Two: The Statistical Knowledge Survey

This survey was done on the computer. Each participant was allowed a maximum of 10 minutes to complete 10 multiple-choice questions. There was a time clock informing the participant how much time was left for him/her to complete the task. When 10 minutes elapsed, the computer automatically submitted the responses and proceeded to the next part. The purpose of this part was to test the quality of the participant's statistical knowledge and to provide a means to weight his/her performance. The participant earned 50 cents for each question answered correctly. The maximum payoff in this part is $5.

Part Three: The Risk Preference Survey

This part was done on the computer. Each participant was allowed a maximum of 10 minutes to complete 10 paired-choice questions. There was a time clock informing the participant how much time was left for him/her to complete the job. When 10 minutes elapsed, the computer automatically submitted the responses and proceeded to the next part. Participant's payoff in this part was one of four values: $1, $8, $10, and $19 depending on the participant's decision choices and chance. The purpose of this part was to identify the participant's risk preference.

Table 4.9 Risk Preference Decision Sheet

Question	Option A	Option B	Which Option is preferred?
1	10% chance of $10.00, 90% chance of $8.00	10% chance of $19.00, 90% chance of $1.00	
2	20% chance of $10.00, 80% chance of $8.00	20% chance of $19.00, 80% chance of $1.00	
3	30% chance of $10.00, 70% chance of $8.00	30% chance of $19.00, 70% chance of $1.00	
4	40% chance of $10.00, 60% chance of $8.00	40% chance of $19.00, 60% chance of $1.00	
5	50% chance of $10.00, 50% chance of $8.00	50% chance of $19.00, 50% chance of $1.00	
6	60% chance of $10.00, 40% chance of $8.00	60% chance of $19.00, 40% chance of $1.00	
7	70% chance of $10.00, 30% chance of $8.00	70% chance of $19.00, 30% chance of $1.00	
8	80% chance of $10.00, 20% chance of $8.00	80% chance of $19.00, 20% chance of $1.00	
9	90% chance of $10.00, 10% chance of $8.00	90% chance of $19.00, 10% chance of $1.00	
10	100% chance of $10.00, 0% chance of $8.00	100% chance of $19.00, 0% chance of $1.00	

The payoffs for "safe" Option A, $10.00 or $8.00, are less variable than the potential payoffs of $19.00 or $1.00 in the "risky" Option B.

Table 4.10 Expected Payoff

Question	Option A	Option B	Difference
1	$8.20	$2.80	$5.40
2	$8.40	$4.60	$3.80
3	$8.60	$6.40	$2.20
4	$8.80	$8.20	$0.60
5	$9.00	$10.00	-$1.00
6	$9.20	$11.80	-$2.60
7	$9.40	$13.60	-$4.20
8	$9.60	$15.40	-$5.80
9	$9.80	$17.20	-$7.40
10	$10.00	$19.00	-$9.00

NOTE:

This expected payoff table was not available to the participants.

In the first question, the probability of the high payoff ($10.00 or $19.00) was 10%, so only the extreme risk lover would choose Option B. As indicated by the difference of expected payoff, the incentive to choose Option A is $5.40. When the probability of the high payoff increased enough (moving down the questions), a rational person should shift from Option A to Option B. For example, a risk neutral person would choose A four times before switching to Option B. Even the most risk averse person should switch over by question 10 since Option B provided a sure payoff of $19.00 in that case. Therefore, the total number of "safe" Option A choices for each of the ten questions would be used as an indicator of risk aversion.

Table 4.11 Risk Aversion Coefficient

Number of safe choices	Range of relative risk aversion for $U(W) = \frac{W^{1-r}}{1-r}$	Middle point of relative risk aversion	Risk preference classification
0-1	-1.76[a] < rr < -0.93	-1.365	highly risk loving
2	-0.97 < rr < -0.49	-0.73	very risk loving
3	-0.49 < rr < -0.13	-0.31	risk loving
4	-0.13< rr < 0.19	0.03	near risk neutral
5	0.19 < rr < 0.48	0.335	slightly risk averse
6	0.48 < rr < 0.78	0.63	risk averse
7	0.78 < rr < 1.13	0.955	very risk averse
8	1.13 < rr < 1.6	1.365	highly risk averse
9-10	1.6 < rr < 2.2[a]	1.9	stay in bed

[a] these two lower and upper bound are subjectively determined

When participant completed his/her decisions, the number of "safe" Option A choices would be calculated. The corresponding middle point of relative risk aversion would be used as indicator of his/her risk preference.

Part Four: The Subjective Probability Elicitation Section

This part was done on the computer. Each participant was requested to provide answers to the following six questions for each population.

1. What is the chance the variable will fall below 80 next period?
2. What is the chance the variable will fall at or above 80 and below 90 next period?
3. What is the chance the variable will fall at or above 90 and below 100 next period?
4. What is the chance the variable will fall at or above 100 and below 110 next period?

5. What is the chance the variable will fall at or above 110 and below 120 next period?

6. What is the chance the variable will fall at or above 120 next period?

At the beginning, an Excel spreadsheet containing a sample dataset with 50 observations. Participants were told this sample dataset was randomly drawn from a uniformly distributed population with a range between 75 and 125. Then they were allowed to analyze this sample dataset and provide answers to the above-mentioned six questions. This would acquaint the participants with what they were anticipating to do in the ensuing formal eliciting process.

After that, there were four datasets to be estimated and the known population distribution underlying each dataset would not be announced to participants. The payoff was based on the accuracy of the subjective estimate. The more accurate the estimate, the more payoff the participant earned.

$$Payoff = Maximum(\$0.00, \$10.00 - 0.025 \times \sum_{i=1}^{6} (A_i - B_i)^2) \quad (4\text{-}1)$$

According to (4-1), the maximum payoff for each dataset was \$10.00 if the total squared deviation is zero. The minimum payoff is \$0.00 if the total squared deviation is greater than 400. In any case, participants did not lose money.

Treatment One: Composite Method

Participants were recruited to the economics experimental lab and a \$5 show up fee was paid for their presence. The moderator gave each participant a test ID and assigned a laptop to each person. By inserting the test ID, participants accessed the experiment website. When they logged onto the experiment website, each one was assigned a user ID by the computer.

Two basic principles were announced by the moderator. The first principle stated that the participation was completely voluntary. The second principle stated that all information provided by the participants would be kept strictly confidential and would be used only for the purposes of this research.

Two rules were observed during the experiment. One was for participants to sit some distance from any of the other participants. The other was no talking. Any kind of direct communication between participants was strictly prohibited. However, participants could raise their hand if they had questions at any time during the experiment procedure. If participants did not have any questions, they would proceed to the next interface. There they were required to answer several basic questions regarding age, gender, academic classification, and statistics background.

The next step was Part One of the experiment – The Myers-Briggs Type Indicator (MBTI) Survey. The moderator gave each participant one copy of the MBTI Booklet and an answer sheet. Also a pencil and a blank sheet of paper was provided. Since this part was done on paper, test ID and user ID were required to be written on the answer sheet in order to ensure information consistency and confidentiality. Participants were told there are no standard answers in this part and their responses will not account for their final payments. The time limit to complete this part was 20 minutes.

When part one was completed, the moderator collected the MBTI Booklet and the answer sheet. Participants were then told the following three parts would be done on the computer.

The first section conducted on the computer was Part Two of the experiment – The Statistical Knowledge Survey. The time constraint to compete 10 multiple-choice questions was 10 minutes. At the beginning of Part Three of the experiment – The Risk Preference Survey, the moderator read the instructions. The time constraint to complete this part was 10 minutes.

In part three there were ten questions to be answered. In each question there were two choices, "Option A" or "Option B". Option A had two paired events (Event A or Event B). Option B had two paired events (Event C or Event D). Each event had been assigned a different probability. Participants were asked to choose either Option A or Option B. After the participants submitted responses, the computer would make two random selections to decide their payment. The first random selection the computer would make was to select which of the ten questions would be used. So among the ten choices only one would be used to determine the payoff. However which one would be used was not known in advance. The computer program was insured that each question as well as each choice had an equal chance of being used.

The second random selection the computer made was to select which event of the choice decided by the first random selection would be used to determine payment. The computer program was insured that each event would be selected based on its assigned probability.

Table 4.12 Example Used to Explain the Payoff Determination in Part Three

Question	Option A	Option B	Decision
1	10% chance of $10.00, 90% chance of $8.00 which means: Event A: $10; Probability: 10% Event B: $8; Probability: 90%	10% chance of $19.00, 90% chance of $1.00 which means: Event C: $19; Probability: 10% Event D: $1; Probability: 90%	Option B
10	100% chance of $10.00, 0% chance of $8.00 which means: Event A: $10; Probability: 100% Event B: $8; Probability: 0%	100% chance of $19.00, 0% chance of $1.00 which means: Event C: $19; Probability: 100% Event D: $1; Probability: 0%	Option A

Suppose in the first random selection, Question 1 was selected by the computer. If the decision was Option B, in the second random selection, the computer would select either "Event C: 10% chance of $19.00" or "Event D: 90% chance of $1.00" in Option B as final payment. If the participant was "lucky", he/she earned $19. Here, Event C had a 10% chance to be selected. If the participant was "unlucky", he/she earned $1 since Event D had a 90% chance to be selected.

If the decision was Option A, in the second random selection, the participant might earn either $10 (if Event A has been selected) or $8 (if Event B has been selected). In this case, Event A had a probability of 10% while Event B had a probability of 90% to be chosen.

Suppose in the first random selection, Question 10 was selected by the computer. If the decision was Option A, $10 would be paid definitely since only Event A would occur. If the decision was Option B, $19 would be paid definitely since Event C was the only choice.

Generally speaking, each question composed of four potential events ($10, $8, $19, $1). After the first random selection, there would be only two potential events left ($10/$8 or $19/$1). After the second random selection, the final event would be chosen and the participant's payment would be determined. The participant's choice and the chance would jointly determine the final payment. In the experiment design, participants would have access to their payoff in part three immediately after they submitted their responses.

Part Four of the experiment – The Subjective Probability Elicitation Section was the last part of this treatment. At the beginning, a sample dataset would be downloaded to the participants' laptops. Even though participants were told the underlying population was a uniformly distributed within a range between 75 and 125, most of them would spend time analyzing the sample dataset. This exercise acquainted them with the subjective estimation task in the following steps. Then six questions required answers and the purpose for them would be explained. At this time all participants should clearly understand how to provide valid subjective probability estimates. If this was the case, the moderator would explain the payoff rule until everyone knew how his/her payoff would be affected by his/her subjective estimates.

Table 4.13 Payoff Calculation in Part Four

	$(-\infty, 80)$	$[80, 90)$	$[90, 100)$	$[100, 110)$	$[110, 120)$	$[120,+\infty)$
KPD	A_1	A_2	A_3	A_4	A_5	A_6
SE	B_1	B_2	B_3	B_4	B_5	B_6
SD	$(A_1 - B_1)^2$	$(A_2 - B_2)^2$	$(A_3 - B_3)^2$	$(A_4 - B_4)^2$	$(A_5 - B_5)^2$	$(A_6 - B_6)^2$

where:

KPD denotes known population distribution, *SE* denotes subjective estimate, *SD* denotes squared deviation.

NOTE:

$$\sum_{i=1}^{6} B_i = 1 \tag{4-2}$$

$0 \le B_i \le 1$, i=1, 2 … 6

If no questions existed, formal estimation would begin with dataset one. The time constraint was 10 minutes. There was a time clock informing the participant how much time was left to complete his/her job. When 10 minutes elapsed, the computer automatically submitted the responses and proceeded to the next dataset. The same elicitation process was repeated in dataset two, dataset three, and dataset four. When the subjective estimates of dataset four had been submitted, the experiment was completed. At this time, total payout was revealed and each participant was paid in cash.

Treatment Two: Composite with Feedback

The only difference between Treatment Two and Treatment One was in Part Four. In Treatment Two, there were three rounds in each dataset. The first round remained the same as in Treatment One. Participants conduct their individual estimates within the ten minute limit. By using the same test ID (this is controlled by the moderator), five people were randomly assigned to a group. After these five people submitted their first round responses, they were allowed to proceed to round two. At this time, there would be a "show others" button on the computer screen. By clicking on it, participants could access other group members' responses in round one. Within a four minute limit, a participant could rethink the questions and modify his/her subjective estimates. There was a time clock informing the participant of how much time was left to complete his/her job. After four minutes elapsed, the computer automatically submitted the responses and proceeded to round three. Again, in round three, by clicking the "show others" button, a participant could access other group members' responses in round one and round two. He/She would have four minutes to rethink the questions and modify his/her subjective estimates. Only the responses in round three would be used to account for payment and forecast accuracy in each dataset.

Treatment Three: Consensus Method

The only difference between Treatment Two and Treatment Three was in Part Four. In Treatment Three, in round two and round three of each dataset, there would be one more button named "join chat." By clicking on it, participants could access a chat room. In this chat room, participants would know there would be a $2 bonus for each group member if they could reach a consensus in round three of each dataset. They could write anything they want to share with other group members in the chat room. Whether they reach a consensus or not, their responses in round three would be used to determine their payoffs and forecast accuracies in each dataset.

Data Analysis

The purposes of this economic experiment are two fold: one is to find whether the forecast accuracy or error of aggregated subjective probabilities will depend on different treatment, different datasets, and different aggregation methods. The other is to find whether personality type, risk preference, and knowledge could attribute to the forecast accuracy or error of individual subjective probabilities. Forecast accuracy or error is defined as the squared deviation between aggregated/individual subjective probabilities and the true probabilities from the known population distribution.

Aggregated Subjective Estimates

Table 4.14 Error of Aggregated Subjective Estimates Calculation

	SS	(-∞, 80)	[80, 90)	[90, 100)	[100, 110)	[110, 120)	[120,+∞)
$SE\,1$	SS_1	B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}
$SE\,2$	SS_2	B_{21}	B_{22}	B_{23}	B_{24}	B_{25}	B_{26}
$SE\,3$	SS_3	B_{31}	B_{32}	B_{33}	B_{34}	B_{35}	B_{36}
$SE\,4$	SS_4	B_{41}	B_{42}	B_{43}	B_{44}	B_{45}	B_{46}
$SE\,5$	SS_5	B_{51}	B_{52}	B_{53}	B_{54}	B_{55}	B_{56}
KPD		A_1	A_2	A_3	A_4	A_5	A_6

NOTE:

$$\sum_{j=1}^{6} B_{ij} = 1 \tag{4-3}$$

$0 \le B_{ij} \le 1 \quad i = 1, 2, 3, 4, 5$ and $j = 1, 2, 3, 4, 5, 6$

$$AAEAE = \sum_{j=1}^{6} \left(A_j - \frac{\sum_{i=1}^{5} B_{ij}}{5} \right)^2 \tag{4-4}$$

$$AAEAEL = (A_1 - \sum_{i=1}^{5} B_{i1})^2 \tag{4-5}$$

$$AAEAER = (A_6 - \sum_{i=1}^{5} B_{i6})^2 \tag{4-6}$$

$$AAEAET = (A_1 - \sum_{i=1}^{5} B_{i1})^2 + (A_6 - \sum_{i=1}^{5} B_{i6})^2 \tag{4-7}$$

$$WAEAE = \sum_{j=1}^{6} \left(A_j - \sum_{i=1}^{5} \left(\frac{SS_i}{\sum_{i=1}^{5} SS_i} \times B_{ij} \right) \right)^2 \tag{4-8}$$

$$WAEAEL = (A_1 - \sum_{i=1}^{5} (\frac{SS_i}{\sum_{i=1}^{5} SS_i} \times B_{i1}))^2 \tag{4-9}$$

$$WAEAER = (A_6 - \sum_{i=1}^{5} (\frac{SS_i}{\sum_{i=1}^{5} SS_i} \times B_{i6}))^2 \tag{4-10}$$

$$WAEAET = (A_1 - \sum_{i=1}^{5} (\frac{SS_i}{\sum_{i=1}^{5} SS_i} \times B_{i1}))^2 + (A_6 - \sum_{i=1}^{5} (\frac{SS_i}{\sum_{i=1}^{5} SS_i} \times B_{i6}))^2 \tag{4-11}$$

where:

KPD denotes known population distribution, SS denotes individual score in part two, SE denotes individual subjective estimate, $AAEAE$ denotes arithmetic average of error of aggregated estimates across entire distribution, $AAEAEL$ denotes arithmetic average of error of aggregated estimates across left tail, $AAEAER$ denotes arithmetic average of error of aggregated estimates across right tail, $AAEAET$ denotes arithmetic average of error of aggregated estimates across two tails, $WAEAE$

denotes weighted average of error of aggregated estimates across entire distribution,

$WAEAEL$ denotes weighted average of error of aggregated estimates across left tail,

$WAEAER$ denotes weighted average of error of aggregated estimates across right tail,

$WAEAET$ denotes weighted average of error of aggregated estimates across two tails.

Consider the following regression model:

$$EAE = \beta_1 + \beta_2 TD_2 + \beta_3 TD_3 + \beta_4 DD_2 + \beta_5 DD_3 + \beta_6 DD_4 + \beta_7 WD + \varepsilon \qquad (4\text{-}12)$$

where:

EAE = error of aggregated estimates

TD_2 = 1 if the error of aggregated estimates is obtained from treatment two

= 0 otherwise

TD_3 = 1 if the error of aggregated estimates is obtained from treatment three

= 0 otherwise

DD_2 = 1 if the error of aggregated estimates is obtained from dataset two

= 0 otherwise

DD_3 = 1 if the error of aggregated estimates is obtained from dataset three

= 0 otherwise

DD_4 = 1 if the error of aggregated estimates is obtained from dataset four

= 0 otherwise

WD = 1 if the error of aggregated estimates is of simple arithmetic average

= 0 otherwise (i.e., weighted average)

In this model, independent variables are all exclusively qualitative variables. Such models are called Analysis of Variance (ANOVA) models. ANOVA models are used to test if there is a statistically significant relationship between a quantitative dependent variable and qualitative independent variables. It is a convenient measure to compare the differences in the mean values of two or more different treatments. Regression (4-12) was used for the first purpose of this economic experiment.

Individual Subjective Estimates

Table 4.15 Error of Individual Subjective Estimates Calculation

	$(-\infty, 80)$	$[80, 90)$	$[90, 100)$	$[100, 110)$	$[110, 120)$	$[120, +\infty)$
SE	B_1	B_2	B_3	B_4	B_5	B_6
KPD	A_1	A_2	A_3	A_4	A_5	A_6

NOTE:

$$\sum_{i=1}^{6} B_i = 1 \tag{4-13}$$

$$0 \le B_i \le 1 \quad i = 1, 2, 3, 4, 5, 6$$

$$EIEE = \sum_{i=1}^{6} (A_i - B_i)^2 \tag{4-14}$$

$$EIEL = (A_1 - B_1)^2 \tag{4-15}$$

$$EIER = (A_6 - B_6)^2 \tag{4-16}$$

$$EIET = (A_1 - B_1)^2 + (A_6 - B_6)^2 \tag{4-17}$$

where:

KPD denotes known population distribution, SE denotes individual subjective estimate, $EIEE$ denotes error of individual subjective estimates across entire distribution, $EIEL$ denotes error of individual subjective estimates across left tail, $EIER$ denotes error of individual subjective estimates across right tail, $EIET$ denotes error of individual subjective estimates across two tails.

Consider the following regression model:

$$\begin{aligned} EIE &= \beta_1 + \beta_2 TD_2 + \beta_3 TD_3 + \beta_4 DD_1 + \beta_5 DD_2 + \beta_6 DD_3 + \beta_7 MD_1 \\ &+ \beta_8 MD_2 + \beta_9 MD_3 + \beta_{10} MD_4 + STAT + RISK + \varepsilon \end{aligned} \tag{4-18}$$

where:

EIE = error of individual subjective estimates

TD_2 = 1 if the error of aggregated estimates is obtained from treatment two

= 0 otherwise

TD_3 = 1 if the error of aggregated estimates is obtained from treatment three

= 0 otherwise

DD_2 = 1 if the error of aggregated estimates is obtained from dataset two

= 0 otherwise

DD_3 = 1 if the error of aggregated estimates is obtained from dataset three

= 0 otherwise

DD_4 = 1 if the error of aggregated estimates is obtained from dataset four

= 0 otherwise

MD_1 = 1 if the individual MBTI is in category Extraversion

= 0 if the individual MBTI is in category Introversion

MD_2 = 1 if the individual MBTI is in category Sensing

= 0 if the individual MBTI is in category Intuition

MD_3 = 1 if the individual MBTI is in category Thinking

= 0 if the individual MBTI is in category Feeling

MD_4 = 1 if the individual MBTI is in category Judging

= 0 if the individual MBTI is in category Perceiving

STAT = individual's score of statistical knowledge

RISK = individual's coefficient of relative risk aversion

This model contains both qualitative and quantitative independent variables. A regression model containing a mixture of quantitative and qualitative independent variables is called an analysis of covariance (ANCOVA) model. This is an extension of the ANOVA model. Regression (4-18) was used for the second purpose of this economic experiment.

CHAPTER V

RESULTS

Previous chapters discussed the problems and prior studies with regard to elicitation and aggregation of subjective probability estimates. The conceptual framework and methods employed in this study have been analyzed. A web-based economic experiment has been executed. This chapter discusses the empirical results of this research.

Demographic Characteristics of Participants

One hundred five students at Mississippi State University were recruited to take part in the experiment. Among them, thirty-five attended the Treatment One, twenty-five attended the Treatment Two in five groups of five participants per group (randomly grouped), and forty-five attended the Treatment Three in nine groups of five participants per group (randomly grouped). The time for Treatment One was about one hour forty minutes. The time for Treatment Two and Three was about two hour ten minutes. Average payoff was about $40 per participant.

In the experiment design, there was a time clock informing the participant how much time was left for him/her to carry out the subjective probability estimate in part four. When the time limitation was reached, the computer would submit the responses and proceed to the next step automatically. Some participants failed to type valid probability estimates by the end of the time limit (e.g., $\sum P_i \neq 1$); those observations were dropped. In the end, twenty-eight valid probability estimates were obtained from Treatment One; twenty-five valid probability estimates were obtained from Treatment Two. In Treatment Three, if one participant failed to provide valid probability estimates, since he/she had participated in the chat room discussion, the data from the entire group had to be discarded. Finally, twenty-five valid probability estimates were obtained from Treatment Three.

In order to compare with the data observed from Treatment 2 and Treatment 3, twenty-five observations were randomly selected and grouped (five per group) from twenty-eight valid probability estimates obtained from Treatment One. Table 5.1 presents a summary of those seventy-five participants.

Table 5.1 Demographic Characteristics of Participants

Treatment	**Age**		**Gender (%)**		**Classification(%)**			
	Average	S.D.[a]	Male	Female	Graduate	Senior	Junior	Other
1	26.920	4.813	23 (92%)	2 (8%)	20 (80%)	4 (16%)	0	1 (4%)
2	25.640	4.600	17 (68%)	8 (32%)	17 (68%)	4 (16%)	2 (8%)	2 (8%)
3	28.480	8.466	17 (68%)	8 (32%)	14 (56%)	9 (36%)	1 (4%)	1 (4%)
Total	27.013	6.244	57 (76%)	18 (24%)	51 (68%)	17 (23%)	3 (4%)	4 (5%)

[a] Denotes standard deviation

Table 5.1 Demographic Characteristics of Participants (Continued)

Treatment	Statistical Class		Statistical Score		Risk Preference # of safe choice[1]		Myers-Briggs Type Indicator E		Myers-Briggs Type Indicator S		Myers-Briggs Type Indicator T		Myers-Briggs Type Indicator J	
	Average	S.D.[a]	Average	S.D.[a]	Average	S.D.[a]	Average	S.D.[a]	Average	S.D.[a]	Average	S.D.[a]	Average	S.D.[a]
1	1.6000	1.4142	8.1200	1.2689	5.2500	1.6819	0.60	0.50	0.36	0.49	0.48	0.51	0.60	0.50
2	1.6000	2.1602	8.4400	1.4457	4.6842	1.4550	0.44	0.51	0.36	0.49	0.52	0.51	0.52	0.51
3	1.2000	0.9129	7.1200	1.4236	5.0588	1.9834	0.60	0.50	0.48	0.51	0.6	0.5	0.52	0.51
Total	1.467	1.5711	7.893	1.4757	5.0000	1.6949	0.55	0.50	0.40	0.49	0.53	0.50	0.55	0.50

[a] Denotes standard deviation

[1] We drop those subjects who switched from B back to A. So the valid observations are 20, 19, and 17 in Treatment 1, 2, and 3, respectively

Table 5.1 shows that the average age of the selected twenty-five participants in Treatment One was 26.92. The average age of the participants in Treatment Two was 25.64. The average age of the participants in Treatment Three was 28.48. Overall, the average age of those seventy five participants in the economic experiment was 27.0133. This number indicates that most of the participants were graduate students.

The majority of the participants in the three treatments were male. Only 2, 8, and 8 female students attended in Treatment 1, 2, and 3, respectively.

Table 5.1 shows that the average number of statistical classes the participants had taken before attending this experiment was 1.4667. The mean statistical score 7.8933 out of 10 indicates the majority participants did well on the statistics knowledge survey.

As to risk preference, those who had switched from B back to A (an indicator of non-rational decision) were deleted. So the valid observations with regard to risk preference were 20, 19, and 17 in Treatment 1, 2, and 3, respectively. The mean number of safe choice was 5.25, 4.6842, and 5.0588 in Treatment 1, 2, and 3, respectively. The mean number among the total 56 valid observations was 5 demonstrating that the majority of participants exhibited slightly risk averse behavior.

As to personality type, in Treatment 1 60% of the participants belonged to the Extroversion group, 36% belonged to the Sensing group, 48% belonged to the Thinking group, and 60% belonged to the Judging group. In Treatment 2, those numbers were 44%, 36% 52%, and 52%, respectively; and in Treatment 3, those numbers were 60%, 48%, 60%, and 52%, respectively. Overall, among those seventy-five participants, 55%

belonged to the Extroversion group, 40% belonged to the Sensing group, 53% belonged to the Thinking group, and 55% belonged to the Judging group.

Results of the Error of Aggregated Estimate Model

There were five groups in each treatment. Each group consisted of five individuals. Two aggregation techniques were employed to combine the individual's probability estimates. One was a simple arithmetic average, which had been widely used due to its straightforwardness and simplicity. The other was a weighted average by assigning differential weights according to an individual's statistical score. The logic underlying these two aggregation techniques was that in the first method, the quality of experts' expertise cannot be differentiated. In the second method, the quality of experts' expertise can be differentiated according to their performance in the statistical knowledge survey.

Table 5.2 shows the regression output for the error of aggregated estimate model. The baseline category was the mean error of aggregated estimate obtained from Treatment 1 by Dataset 1. The left tail, right tail, and two tails situations were helpful in revealing the estimates of extreme probability events.

Table 5.2 Regression output for the Error of Aggregated Estimate Model

	Error of Aggregated Estimate across entire distribution (-∞,+∞)		Error of Aggregated Estimate across left tail (-∞, 80)		Error of Aggregated Estimate across right tail (120,+∞)		Error of Aggregated Estimate across two tails (-∞, 80) and (120,+∞)	
Variables	coefficient	p - value	coefficient	p - value	coefficient	p - value	coefficient	p - value
Intercept	428.828	**< .0001*** **	15.2945	**< .0001*** **	22.7093	**< .0001*** **	38.0038	**< .0001*** **
TD2	-17.3643	0.3286	0.1011	0.9425	-6.0287	**0.0394** **	-5.9275	**0.0746***
TD3	-28.0505	0.1158	-1.8683	0.1843	-4.2548	0.1442	-6.1231	**0.0656***
DD2	-333.2251	**< .0001*** **	-9.4043	**< .0001*** **	-3.6121	0.2819	-13.0164	**0.0009*** **
DD3	-345.9699	**< .0001*** **	-13.5723	**< .0001*** **	-15.5028	**< .0001*** **	-29.0752	**< .0001*** **
DD4	-77.9412	**0.0002*** **	-2.844	**0.0809***	5.9547	**0.0774***	3.1106	0.4151
WD	-1.8242	0.8998	0.1673	0.8838	-0.1396	0.953	0.0277	0.9918
# of Obs	120		120		120		120	
R^2	0.7994		0.4434		0.3007		0.4531	

***, **, * Denote significance at the 1%, 5%, and 10% levels, respectively.

Treatment Effect

Participants in Treatment 1 did not have a chance to interact with each other. Those in Treatment 2 had two chances to review the responses from group members. Those in Treatment 3 not only had two chances to access group members' responses but also had two chances to communicate with each other through a chat room before they reached final estimates. If these kinds of interactions help, the error of aggregated estimate should be shrinking through treatments. In other words, the coefficients of TD2 and TD3 should be negative, and absolute value of the coefficient of TD3 should be greater than that of TD2.

Table 5.2 shows that in column four (across two tails), the signs of these two coefficients were negative, and the absolute value of TD3 was greater than that of TD2. However, the difference was not statistically significant. The other significant result of the treatment effect was the negative sign of TD2 in column three (across right tail) which means holding other variables constant, Treatment 2 produces more accurate subjective probability estimate than Treatment 1 does in right tail situation.

Compared to the baseline category Treatment 1, the negative sign of TD2 in three situations (except left tail) means that Treatment 2 works better than Treatment 1 with regard to the accuracy of aggregated subjective probability estimate. This finding is statistically significant in right tail and two tail situation.

Compared to the baseline category Treatment 1, the negative sign of TD3 in four situations indicates that as far as the accuracy of aggregated subjective probability is concerned, Treatment 3 resulted in better forecast than Treatment 1. Holding other variables constant, the mean error of aggregated estimate in Treatment 3 was 28.0505, 1.8683, 4.2548, and 6.1231 less than the mean error in Treatment 1. In the two tails situation which usually is commonly used to predict the odds of a low probability event, this result is statistically significant.

Dataset Effect

Dataset 1 was composed of 30 observations drawn from a normal distribution. Dataset 2 was composed of 50 observations drawn from a normal distribution. Dataset 3 was composed of 100 observations drawn from a normal distribution. Dataset 4 was composed of 30 observations drawn from a left skewed beta distribution.

Given the same normal population distribution, the more observations the dataset has, the more accurate the subjective estimate is expected. This hypothesis means that not only the signs of DD2 and DD3 should be negative but the absolute value of the coefficient of DD3 should be greater than that of DD2.

Table 5.2 shows that in columns one (across entire distribution), two (across left tail), and four the signs of these two coefficients were negative, and the absolute value of DD3 was greater than that of DD2. The difference was statistically significant in column two and four but not significant in column one.

The other significant result of dataset effect was the negative sign of DD3 in column three which means holding other variables constant, Dataset 3 produces more accurate subjective probability estimate than Dataset 1 does for the right tail situation.

Compared to baseline category Dataset 1, the negative sign of DD3 in four situations indicates that as far as the accuracy of aggregated subjective probability is concerned, Dataset 3 was easier to forecast than Dataset 1. Holding other variables constant, the mean error of aggregated estimate in Dataset 3 was 345.97, 13.5723, 15.5028, and 29.0752 less than those numbers in Dataset 1. In all situations, this result was statistically significant.

Compared to baseline category Dataset 1, the negative sign of DD2 in four situations indicates that as far as the accuracy of aggregated subjective probability is concerned, Dataset 2 was easier to predict than Dataset 1. Holding other variables constant, the mean error of aggregated estimate in Dataset 2 was 333.225, 9.4043, 3.6121, and 13.0164 less than those numbers in Dataset 1. In all except for right tail situations, this result was statistically significant.

Given the same number of observations, datasets drawn from different population distributions may produce different levels of accuracy of subjective probability estimates. Table 5.2 shows the coefficient of DD4 was negative in column one and two, but positive in column three. These statistically significant results show that given the same 30 observations, the dataset drawn from a normal population distribution produced more accurate subjective probability estimates across the entire distribution and across the left tail, but less accurate estimates across the right tail than the dataset drawn from a left skewed beta population distribution.

Aggregation Technique Effect

Table 5.2 shows that the coefficient of WD in all situations was not significantly different from zero. This means no matter which approach to aggregate experts' subjective probability estimate, whether by means of simple arithmetic average or weighted average based on statistical knowledge, the final mean accuracy of aggregated estimate does not change.

After examining the data, by using the WD dummy variable there may be correlation between the residuals calculated by arithmetic average and those calculated by weighted average. This potentially violates the Gauss-Markov assumption of OLS regression.

Paired T-Test was carried out to see if there was serious problem caused by WD dummy variable. Table 5.3 shows the results of Paired T-Test. Among 48 paired comparisons between arithmetic average of error of aggregated estimate and weighted average of error of aggregated estimate, only one result (from Treatment 1, Dataset 2) was significant at 10% level. In short, the results from Paired T-Test were consistent with the regression result.

Additionally, the bias and Mean-Squared Error of aggregated estimate had been calculated to see if there was difference between these two aggregated methods from the perspective of accuracy of forecast prediction. Table 5.4 shows the results.

Table 5.5 shows the output of regression by using bias as dependent variable instead of error of aggregated estimate in the error of aggregated estimate model. Table 5.6 shows the output of regression by using Mean-Squared Error instead of error of aggregated estimate in the error of aggregated estimate model. These two regression outputs again indicate that the WD dummy variable does not make any difference.

Table 5.3 Paired T-Test for the Arithmetic Average and Weighted Average of Error of Aggregated Estimate

Treatment	Dataset	Range	E(AAEAE)[1]	E(WAEAE)[2]	p - value
		(-∞, +∞)	435.0568	438.6119	0.6972
1	1	(-∞, 80)	13.0304	13.0671	0.9481
		(120, +∞)	19.0071	18.7491	0.6481
		(-∞, 80) or (120, +∞)	32.0375	31.8162	0.7313
		(-∞, +∞)	91.1212	86.4900	0.2991
1	2	(-∞, 80)	10.1481	8.8322	**0.0614**
		(120, +∞)	18.3785	17.0119	0.2562
		(-∞, 80) or (120, +∞)	28.5266	25.8441	0.1392
		(-∞, +∞)	60.7900	57.9760	0.3876
1	3	(-∞, 80)	1.2912	1.0518	0.3201
		(120, +∞)	11.6544	10.0117	0.3261
		(-∞, 80) or (120, +∞)	12.9456	11.0635	0.3250
		(-∞, +∞)	365.7920	373.2170	0.2106
1	4	(-∞, 80)	12.2455	11.7172	0.2070
		(120, +∞)	29.0701	30.9128	0.5095
		(-∞, 80) or (120, +∞)	41.3156	42.6300	0.6491
		(-∞, +∞)	436.0420	437.3305	0.6767
2	1	(-∞, 80)	20.0035	19.5703	0.3042
		(120, +∞)	18.3392	17.7928	0.3438
		(-∞, 80) or (120, +∞)	38.3428	37.3631	0.2061
		(-∞, +∞)	61.5148	61.5760	0.9447
2	2	(-∞, 80)	2.3944	2.3222	0.4354
		(120, +∞)	11.2145	11.0259	0.6783
		(-∞, 80) or (120, +∞)	13.6089	13.3481	0.5890
		(-∞, +∞)	59.8724	60.5960	0.3751
2	3	(-∞, 80)	1.1408	1.1407	0.9815
		(120, +∞)	0.2269	0.1805	0.3639
		(-∞, 80) or (120, +∞)	1.3677	1.3212	0.3929
		(-∞, +∞)	324.2735	328.9358	0.1870
2	4	(-∞, 80)	12.7409	12.8796	0.2489
		(120, +∞)	23.7915	23.9951	0.4477
		(-∞, 80) or (120, +∞)	36.5324	36.8747	0.3712
		(-∞, +∞)	361.0657	368.5592	0.4071
3	1	(-∞, 80)	11.3441	11.7189	0.1959
		(120, +∞)	19.3095	22.0725	0.3267
		(-∞, 80) or (120, +∞)	30.6536	33.7914	0.2775
		(-∞, +∞)	87.3396	89.2739	0.2871
3	2	(-∞, 80)	4.0681	4.5433	0.6377
		(120, +∞)	17.9985	17.9685	0.9470
		(-∞, 80) or (120, +∞)	22.0666	22.5118	0.7580
		(-∞, +∞)	80.7380	80.8742	0.3739
3	3	(-∞, 80)	1.3648	1.3110	0.3739
		(120, +∞)	0.1168	0.0630	0.3739
		(-∞, 80) or (120, +∞)	1.4816	1.3740	0.3739
		(-∞, +∞)	307.3722	309.4284	0.8107
3	4	(-∞, 80)	11.2384	10.8485	0.2430
		(120, +∞)	21.1153	22.1135	0.4932
		(-∞, 80) or (120, +∞)	32.3537	32.9620	0.6537

[1] denotes mean arithmetic average of error of aggregated estimates;[2] denotes mean weighted average of error of aggregated estimates

Table 5.4 Bias and Mean-Squared Error of Aggregated Estimate

		Aggregated Estimate across left tail (-∞, 80)				Aggregated Estimate across Right Tail (120, +∞)				Aggregated Estimate across two tails (-∞, 80) and (120, +∞)			
		Bias		MSE		Bias		MSE		Bias		MSE	
Treatment		AAAE[1]	WAAE[2]	AAAE	WAAE	AAAE[1]	WAAE[2]	AAAE	WAAE	AAAE[1]	WAAE[2]	AAAE	WAAE
1	D1	3.3045	3.38	13.558	13.5	4.2285	4.21	19.2888	19.0113	7.5331	7.59	61.5191	61.6966
	D2	-2.59	-2.8775	11.0081	14.55	-3.87	-3.734	19.2289	17.7948	-6.46	-6.6115	57.0036	59.8102
	D3	-0.36	-0.3417	1.5816	0.944	2.32	2.158	13.2224	11.3409	1.96	1.8163	18.2736	13.8408
	D4	2.2125	2.154	14.0832	13.47	4.5732	4.768	31.1091	32.9662	6.7857	6.922	73.3161	73.1846
2	D1	4.4496	4.408	20.0547	19.62	4.1804	4.144	18.5551	17.9539	8.63	8.552	75.18	73.5782
	D2	-1.5	-1.48	2.4305	2.352	-3.27	-3.264	11.3449	11.1229	-4.77	-4.744	23.1284	22.7602
	D3	-0.92	-0.954	1.2144	1.2	-0.03	-0.054	0.2834	0.2262	-0.95	-1.008	2.178	2.0198
	D4	3.55	3.572	12.7755	12.9	4.8568	4.88	23.8422	24.0108	8.4068	8.452	71.5	72.0933
3	D1	2.8912	3	12.0904	12.42	4.2524	4.442	19.6161	22.6503	7.1436	7.442	53.9104	57.6764
	D2	-1.79	-1.788	4.2841	4.883	-4.07	-4.074	18.3569	18.289	-5.86	-5.862	37.8116	37.9435
	D3	-0.92	-1.032	1.4944	1.373	-0.12	-0.232	0.1424	0.0653	-1.04	-1.264	2.4336	2.0358
	D4	3.2	3.08	11.488	11.2	4.57	4.688	21.1729	22.1704	7.77	7.768	62.1649	63.1388

[1] denotes arithmetic average of aggregated estimates; [2] denotes weighted average of aggregated estimates

Table 5.5 Regression output for the Bias of Aggregated Estimate Model

Variables	Bias of Aggregated Estimate across left tail (-∞, 80)		Bias of Aggregated Estimate across right tail (120,+∞)		Bias of Aggregated Estimate across two tails (-∞, 80) and (120,+∞)	
	coefficient	p - value	coefficient	p - value	coefficient	p - value
Intercept	3.2218	**< .0001*** **	4.6062	**< .0001*** **	7.828	**< .0001*** **
TD2	0.7805	**0.005*** **	-0.4013	0.2114	0.3792	0.4555
TD3	0.2199	0.3771	-0.6497	**0.0507***	-0.4297	0.3987
DD2	-5.5765	**< .0001*** **	-7.9566	**< .0001*** **	-13.533	**< .0001*** **
DD3	-4.3268	**< .0001*** **	-3.5692	**< .0001*** **	-7.8961	**< .0001*** **
DD4	-0.6108	**0.0435***	0.4798	0.1964	-0.131	0.8219
WD	0.0339	0.866	-0.0259	0.9195	0.008	0.9844
# of Obs	24		24		24	
R^2	0.9718		0.977		0.9789	

***, **, * Denote significance at the 1%, 5%, and 10% levels, respectively.

Table 5.6 Regression output for the Mean-Squared Error of Aggregated Estimate Model

	MSE of Aggregated Estimate across left tail (-∞, 80)		MSE of Aggregated Estimate across right tail (120,+∞)		MSE of Aggregated Estimate across two tails (-∞, 80) and (120,+∞)	
Variables	coefficient	p - value	coefficient	p - value	coefficient	p - value
Intercept	16.7026	**< .0001*** **	23.661	**< .0001*** **	71.389	**< .0001*** **
TD2	-1.2681	0.4256	-7.0779	**0.0005*** **	-9.5258	**0.0469** **
TD3	-2.9321	**0.0763***	-5.1874	**0.0061*** **	-12.6912	**0.011** **
DD2	-8.621	**0.0002*** **	-3.4897	**0.0858***	-24.1839	**0.0002*** **
DD3	-13.9041	**< .0001*** **	-15.2992	**< .0001*** **	-57.1299	**< .0001*** **
DD4	-2.5515	0.173	6.366	**0.004*** **	5.3062	0.3157
WD	-0.1949	0.8797	-0.1199	0.9304	-0.1132	0.9755
# of Obs	24		24		24	
R^2	0.8178		0.9013		0.9192	

***, **, * Denote significance at the 1%, 5%, and 10% levels, respectively.

Table 5.4 shows that the bias was positive in Dataset 1 and 4 in all three treatments across the left tail, the right tail, and both tails evaluations. These suggested that the participants had overestimated the distributions in tails of the underlying populations with small (30) observations. The bias was negative in Dataset 2 in all three treatments across the left tail, the right tail, and both tails evaluations which mean the participants had underestimated the distributions in tails of the underlying population with 50 observations. The bias was also negative in Dataset 3 in Treatment 1, 2, and 3 across the left tail. In the right tail and both tails evaluations, it was positive in Treatment 1 but negative in Treatment 2 and 3. Table 5.4 also shows that the smallest Mean-Squared Error was associated with Dataset 3 which has the most observations (100). This suggests greater subjective probability accuracy when given more information on which decisions can be formed.

Table 5.5 reports a regression output for the bias on the same independent variables as Table 5.2. It shows that compared to the baseline category Treatment 1, the mean bias of aggregated estimate was 0.7805 greater in Treatment 2 across the left tail and 0.6497 less in Treatment 3 across the right tail holding other variables constant. As to the dataset effect, compared to the baseline category Dataset 1, the mean bias of aggregated estimate was 5.5765, 4.3268, and 0.6108 less across left tail in Dataset 2, 3, and 4, respectively. It was 7.9566 and 3.5692 less in Dataset 2 and 3 across the right tail. It was 13.533 and 7.8961 less in Dataset 2 and 3 in both tails evaluation.

Table 5.6 reports a regression output for the Mean-Squared Error on the same independent variables as Table 5.2. It shows that compared to the baseline category

Treatment 1, the mean Mean-Squared Error of aggregated estimate was 2.9321 less in Treatment 3 across the left tail, 7.0779 and 5.1874 less in Treatment 2 and 3 across the right tail, 9.5258 and 12.6912 less in Treatment 2 and 3 across two tails holding other variables constant. As to the dataset effect, the coefficients of DD2 and DD3 are all statistically negative across the left tail, the right tail, and both tails. The coefficient of DD4 was statistically positive across the right tail.

Overall, with regard to the treatment effect, Treatment 2 and 3 do not reduce the bias but they do reduce the Mean-Squared error of aggregated estimate. As to the dataset effect, Dataset 2 and 3 were estimated with less bias and Mean-Squared Error of aggregated estimate. For the left skewed beta distribution Dataset 4, there was less bias across the left tail and more Mean-Squared Error across the right tail.

Results of the Error of Individual Estimate Model

Table 5.7 shows the regression output for the error of individual estimate model. The baseline category is the mean error of individual estimate obtained from Treatment 1 by Dataset 1.

Table 5.7 Regression output for the Error of Individual Estimate Model

	Error of Individual Estimate across entire distribution (-∞,+∞)		Error of Individual Estimate across left tail (-∞, 80)		Error of Individual Estimate across right tail (120,+∞)		Error of Individual Estimate across two tails (-∞, 80) and (120,+∞)	
Variables	coefficient	p - value	coefficient	p - value	coefficient	p - value	coefficient	p - value
Intercept	861.6293	**<.0001*** **	65.9054	**<.0001*** **	141.2804	**0.0072*** **	207.1858	**0.0005*** **
TD2	-97.0366	**0.0091*** **	-8.2458	**0.0826***	-25.3732	0.1171	-33.619	**0.0673***
TD3	-114.5273	**0.0105** **	-10.5297	**0.0657***	-26.1358	0.1796	-36.6655	**0.0971***
DD2	-437.1964	**<.0001*** **	-0.1559	0.9772	-0.1195	0.9949	-0.2755	0.9896
DD3	-435.9869	**<.0001*** **	-14.6464	**0.0077*** **	-2.8551	0.8779	-17.5015	0.4066
DD4	-195.6332	**<.0001*** **	-1.5941	0.7804	24.1884	0.2156	22.5944	0.3073
MD1	-15.3605	0.6323	-1.2013	0.7704	-27.4404	**0.0516***	-28.6417	**0.0729***
MD2	8.5139	0.8005	8.7103	**0.0448** **	39.2367	**0.0083*** **	47.947	**0.0045*** **
MD3	42.5189	0.223	-3.1618	0.4793	-29.5273	**0.0536***	-32.6891	**0.0594***
MD4	-87.014	**0.0078*** **	-5.8462	0.1608	-24.2048	**0.089***	-30.051	**0.0627***
STAT	-25.7053	**0.0498** **	-4.5868	**0.0066*** **	-8.9195	0.1189	-13.5063	**0.0377** **
RISK	9.7996	0.7322	-3.0151	0.412	-2.2178	0.8594	-5.2329	0.7125
# of Obs	224		224		224		224	
R^2	0.4502		0.1249		0.106		0.1266	

***, **, * Denote significance at the 1%, 5%, and 10% levels, respectively.

Treatment Effect

Participants in Treatment 1 did not have a chance to interact with each others. Those in Treatment 2 had two chances to review the responses from group members. Those in Treatment 3 not only had two chances to access group members' responses but also had two chances to communicate with each other through a chat room before they reached final estimates. If these kinds of interactions help, the error of aggregated estimate should be shrinking through treatments. In other words, the coefficients of TD2 and TD3 should be negative, and absolute value of the coefficient of TD3 should be greater than that of TD2.

Table 5.7 shows that in columns one, two, and four the signs of these two coefficients were negative, and the absolute value of the coefficient of TD3 was greater than that of TD2. However, the differences were not statistically significant.

Compared to baseline category Treatment 1, the negative sign of TD2 in the four situations means Treatment 2 results in more accurate estimates than Treatment 1. Holding other variables constant, the mean error of individual estimate in Treatment 2 was 97.0366, 8.2458, 25.3732, and 33.619 less than those numbers in Treatment 1. This finding is statistically significant in all but the right tail situation.

Compared to baseline category Treatment 1, the negative sign of TD3 in the four situations indicate that Treatment 3 was more accurate than Treatment 1. Holding other variables constant, the mean error of individual estimate in Treatment 3 was 114.5273, 10.5297, 26.1358, and 36.6655 less than those numbers in Treatment 1. This finding was statistically significant in all but the right tail situation.

Dataset Effect

Dataset 1 was composed of 30 observations drawn from a normal distribution. Dataset 2 was composed of 50 observations drawn from a normal distribution. Dataset 3 was composed of 100 observations drawn from a normal distribution. Dataset 4 was composed of 30 observations drawn from a skewed beta distribution.

Given the same normal population distribution, the more observations the dataset has, the more accurate the subjective estimate is expected. This hypothesis means that not only the sign of the coefficients of DD2 and DD3 should be negative but the absolute value of the coefficient of DD3 should be greater than that of DD2. Table 5.7 shows that this hypothesis did not hold. In column 2, 3, and 4 the signs of these two coefficients are negative and the absolute value of DD3 is greater than that of DD2. However, only the coefficient of DD3 in column 2 is significant different from zero.

The other significant result of dataset effect was the negative sign of DD2 and DD3 in column one.

Compared to baseline category Dataset 1, the negative sign of DD2 in situation one indicates that Dataset 2 produced more accurate forecast than Dataset 1. Holding other variables constant, the mean error of individual estimate in Dataset 2 was 437.1964 less than the mean error in Dataset 1. And this is statistically significant.

Compared to baseline category Dataset 1, the negative sign of DD3 in situations one and two indicates that the accuracy of individual subjective probability were more accurate for Dataset 3 than Dataset 1. Holding other variables constant, the mean error of individual estimate for Dataset 3 was 435.9869 and 14.6464 less than for Dataset 1. Both findings were statistically significant.

Given the same number of observations, datasets drawn from different population distributions may produce subjective probability estimates with different levels of accuracy. Table 5.7 shows the coefficient of DD4 was negative and significant when the entire distribution is evaluated. This shows that given the same number of observations, the dataset drawn from a normal population distribution was less accurately estimated than the dataset drawn from a skewed beta population distribution.

Knowledge Effect

It is assumed that knowledgeable people tend to make sound, reliable judgment in their domain of expertise. The expert decision elicitation paradigm is built upon such belief. In our experiment, the null hypothesis with regard to knowledge is confirmed by the negative sign of coefficient of STAT variable. Table 5.7 shows that the negative sign has been exhibited in all four evaluations of estimation error. The result is statistically significant in three when evaluating the full distribution, the left tail, and the two tails. This finding suggests the more knowledge the expert has, the less subjective probability estimate error which is consistent with the null hypothesis.

Personality Type Effect

Table 5.7 shows the coefficient of dummy variable MD1 (1 = Extroversion, 0 = Introversion) was negative and statistically significant when evaluating the right tail or both tails of the distribution. Extroverts focus attention on outer world of people and things. Introverts prefer to focus on the inner world of ideas and impressions. In our experiment, participants were presented with some datasets (facts) and they had chances to interact with group members in Treatment 2 and 3. Those in category Extroversion were more efficient in utilizing or analyzing these contexts than those categorized as introverts.

The coefficient of dummy variable MD2 (1 = Sensing, 0 = Intuition) had a positive sign in all four error measurement scenarios and was statistically significant in three. People categorized as Sensing trust information gained from their five senses. "What comes from other people indirectly through the spoken or written word is less trustworthy" (Myers, 1980). People categorized as intuitive listen to their unconsciousness instead of experience gained from senses. The fact those categorized as Intuitive performed better than those who are in the Sensing category in the experiment indicates the importance of intuition in subjective probability estimation.

The coefficient of dummy variable MD3 (1 = Thinking, 0 = Feeling) had a significant negative sign when evaluating performances for the right tail, or both tails. People in the Thinking category tend to base decisions primarily on logic and on objective analysis of cause and effect. This trait is regarded as necessary to establish reasoned conclusions. People in the Feeling category tend to base decisions primarily on values and on subjective evaluation of person-centered concerns. The results show that those with Thinking personality performed better than those with Feeling personality in the subjective probability estimation, especially in the tails of the distribution.

The coefficient of dummy variable MD4 (1 = Judging, 0 = Perceiving) was negative and significant in all but the left tail evaluation. People in the Judging category like a planned and organized approach to life and prefer to have things settled. People in category Perceiving like a flexible and spontaneous approach to life and prefer to keep options open. These findings show that being dependent on reasoned judgment contributes significantly to the accuracy of subjective probability estimation.

In short, according to our regression model, people with Extroversion, Intuition, Thinking, and Judging types tend to be more accurate at subjective probability estimation than people with Introversion, Sensing, Feeling and Perceiving types.

Risk Preference Effect

Besides personality type, risk preference was another psychological factor which was examined in this study. The null hypothesis was that risk aversion people tend to produce subjective probability estimation with greater error because they fear the uncertainty of the decision process.

Table 5.1 shows the mean number among the total 56 valid observations was 5 demonstrating that on average participants exhibited slight risk aversion in our experiment. Table 5.7 shows the coefficient of variable RISK was not statistically significant in any error measurement scenarios which means risk attitude did not affect the accuracy of subjective probability estimation.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

This study uses a combination of statistical sampling techniques and experimental economics to evaluate two hypotheses with regard to eliciting and combining subjective probability estimates. The experimental approach begins with four different datasets drawn from four known population distributions which allow evaluation of elicitation procedures relative to known probabilities. Participants were allowed to analyze one sample dataset in the laboratory and informed of the underlying population distribution from which the sample dataset was drawn and with which their estimates about the population distribution will be compared in advance to the formal eliciting procedure. Participants attended one of three different treatments. Treatment One precluded any kind of interaction between participants so there was no knowledge sharing. Treatment Two allowed participants access other group members' responses so there was indirect knowledge sharing. Treatment Three not only allowed participants access other group members' responses but also allowed them to communicate with each other through a chat room. So in the last treatment there were indirect and direct knowledge sharing. The web-based experimental software provides incentive-compatible rewards to participants in order to induce rational decision-making. Demographic characteristics of participants including age, gender, classification, statistical backgrounds, risk preference, and Myers-Briggs Type Indicator have been collected. These make it possible to investigate whether knowledge and psychological factors such as risk attitude, and personality type are related to the accuracy of subjective probability estimation.

Aggregated Subjective Estimates

Empirical results show compared with the composite method, the composite with feedback and the consensus methods produced less mean error of aggregated estimate. This supports the use of several rounds of feedback in the Delphi approach. Interaction between participants probably helped reduce the heuristic error in subjective judgment environments.

The effectiveness of different datasets shows that given the same underlying normal population distribution, the more data provided, the less mean error of aggregated estimate generated. This is self evident since more data contains more information so that the probability for participants to reach accurate estimates has been increased.

If the datasets contain the same number of observations but with different underlying known population distribution, the model shows that the dataset drawn from the skewed beta distribution generated less mean error of aggregated estimate in some cases, but greater mean error of aggregated estimate in one case. These mixed results suggest not all subjective probability problems are equally difficult to generate accurate aggregated estimate, but clear patterns did not emerge.

The weighting scheme used to aggregate individual responses did not make any statistical difference in our regression model. This emphasizes that even though there are many studies on how to aggregate subjective probability estimates, no one weighting method is superior over another.

Individual Subjective Estimates

When it comes to evaluating individual subjective estimates, compared with treatment one, treatments two and three produced less error in individual subjective estimates holding other variables constant. The dataset effects are the same in individual subjective estimates as in aggregated subjective estimates.

The knowledge effect shows that people with more knowledge tend to generate less error of subjective estimate in their domain of expertise. This finding rigorously emphasizes the theoretical logic underlying using experts to assist decision making.

The regression results for the Myers-Briggs Type Indicator show that people with different personality types performed differently in subjective probability assessment. The best assessor has the Extroversion, Intuition, Thinking, and Judging types.

Another psychological factor in our model is risk preference. However, no evidence here shows that risk attitude played any significant influence on the accuracy of individual subjective estimates.

Conclusion

This analysis presented a scientific means to evaluate alternative subjective probability estimate performance and causal factors. The empirical results provided critical, solidly rigorous information on appropriate elicitation and aggregation techniques that will yield the most accurate subjective estimates of unknown probabilities.

Knowledge is the core of experts' decision support system. The generic nature of expert makes it universally acceptable in decision making process with regard to uncertain events. Knowledge sharing and feedback extend the knowledge space upon which experts formulate subjective probability assessments so that the accuracy of subjective estimates generated under these situations tends to be more reliable.

Myers-Briggs Type Indicator test provides a scientific technique to distinguish different people. The Extroversion, Intuition, Thinking, and Judging type's people are the best to perform subjective probability assessment.

When it comes to performing subjective probability assessment, the expertise level and the personality type of expert should be taken into account. Several rounds of feedback will increase the forecast accuracy.

REFERENCES

Arrow, K. J. (1971). Essays in the Theory of Risk-Bearing. Amsterdam: North-Holland Publishing Company.

Bedford T. and Cooke, R. T. (2001). Probabilistic Risk Analysis: Foundations and Methods. New York: Cambridge University Press.

Clemen, R.T., and Winkler, R.L. (1999). Combining Probability Distributions from Experts in Risk Analysis. *Risk Analysis,* 19 (2), 187-203.

Cooke, R. M. (1991). Experts in Uncertainty: Opinion and Subjective Probability in Science. New York: Oxford University Press.

Cooke, R.M., and Goossens, L.H.J. (2000). Procedures Guide for Structured Expert Judgment in Accident Consequence Modeling. *Radiation Protection Dosimetry,* 90 (3), 303-309.

Genest, C., and Zidek, J. V. (1986). Combining Probability Distribution: A Critique and Annotated Bibliography. *Statistical Science*, 1, 114-148.

Hogarth, R. M. (1987). Judgment and Choice. 2nd ed. New York: John Wiley & Sons, Inc.

Holt, C.A., and Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92, 1644-1655.

Kahneman, D., Slovic, P., and Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Merriam-Webster's Collegiate Dictionary. (2001). 10th ed. Massachusetts: Merriam-Webster Inc.

Mestelman, Stuart, and David Feeny. (1988). Does Ideology Matter? Anecdotal Experimental Evidence on the Voluntary Provision of Public Goods. *Public Choice,* 57, 281-286.

Myers, I. B., and Myers, P. B. (1980). Gift Differing: Understanding Personality Type. Palo Alto, California: Davies-Black Publishing.

Morgan, G., and Henrion, M. (1990). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. New York: Cambridge University Press.

NASA Reference Publication 1358: System Engineering "Toolbox" for Design-Oriented Engineers, http://pbma.nasa.gov/docs/public/pbma/general/toolbox.pdf.

Ouchi, F. (2004). A Literature Review on the Use of Expert Opinion in Probabilistic Risk Analysis. World Bank Policy Research Working Paper 3201.

Pindyck, R. S. and Rubinfeld, D. L. (2001). Microeconomics, 5th ed. New Jersey: Prentice Hall, Inc.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica,* 32(1-2), January-April: 122-136.

Smith, V., G. Suchanek, and A. Williams. (1988). Bubles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets. *Econometrica,* 56, 1119-1151.

Spetzler, C. S., and Carl-Axel S. Stael von Holstein. (1975). Probability Encoding in Decision Analysis. *Management Science,* 22, 340-58.

Thucydides. (431. B. C.); translated, with introduction, notes, and glossary by Steven Lattimore. The Peloponnesian War. Indianapolis/Cambridge: Hackett Publishing Company, Inc.

Tversky, A. and Kahneman, D. (1982a). Judgments of and by Representativeness. In Kahneman, D., Slovic, P., and Tversky, A. (Eds.), Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Von Neumann J. and Morgenstern, O. (1944). Theory of Games and Economic Behavior. New Jersey: Princeton University Press.

APPENDIX A

EXPERIMENT PROCEDURE

AEC Subjective Probability Analysis

Thank you for agreeing to participate in our experiment. This experiment should take approximately 2 hours to complete. All information you provide is ***strictly confidential.***

- Today's experiment consists of four parts, the Myers-Briggs survey, the statistical knowledge survey, the risk preference survey, and the subjective probability solicitation section.
- You are being provided a fee of $5.00 for participation, and you will have six chances in the statistical knowledge survey, the risk preference survey, and the subjective probability solicitation section to earn extra money.
- Your participation is ***strictly voluntary***. Your participation does not require you to consume foods or be exposed to physical risks. This project has been reviewed by the Institutional Review Board for the protection of human subjects (IRB Docket #04-325). If you have questions about your rights as a human subject, please contact the IRB office at 325-3294 or Dr. Keith Coble at coble@agecon.msstate.edu.
- To participate in this experiment, you must provide the Test ID for the test to which you have been assigned. Now please insert your Test ID.
- By clicking the "Begin" button below, you are consenting to be a participant in this experiment under the conditions outlined above.

Test ID:

Begin

Thank you for agreeing to participate in our experiment. This experiment should take approximately 2 hours to complete. All information you provide is ***strictly confidential.***

- Today's experiment consists of four parts, the Myers-Briggs survey, the statistical knowledge survey, the risk preference survey, and the subjective probability solicitation section.

- You are being provided a fee of $5.00 for participation, and you will have six chances in the statistical knowledge survey, the risk preference survey, and the subjective probability solicitation section to earn extra money.

- Your participation is ***strictly voluntary***. Your participation does not require you to consume foods or be exposed to physical risks. This project has been reviewed by the Institutional Review Board for the protection of human subjects (IRB Docket #04-325). If you have questions about your rights as a human subject, please contact the IRB office at 325-3294 or Dr. Keith Coble at coble@agecon.msstate.edu.

- To participate in this experiment, you must provide the Test ID for the test to which you have been assigned. Now please insert your Test ID.

- By clicking the "Begin" button below, you are consenting to be a participant in this experiment under the conditions outlined above.

Test ID: 4194E.
User ID: User5.

-----Voluntary

Your participation is completely voluntary. If you do not wish to participate in the experiment, please say so at any time. Non-participants will not be penalized in any way.

-----Confidential

Your responses will be kept strictly confidential and used **ONLY** for the purposes of this research.

-----Rules

Two rules should be observed in this experiment.

1. Sit some distance from any of the other participants.
2. No talking.

Failure to comply with these two rules will result in immediate disqualification from this experiment.

If you have any question during this experiment, please raise your hand.

Please click on the "Next" button to go to next page.

Initial Survey

Below are some basic questions about yourself. Please answer each question.
Remember, all answers are *strictly confidential*.

1. What is your age? ____ years

2. What is your classification? [Freshman ▾]

3. What is your gender? [Male ▾]

4. Have you ever taken a statistics class? [No ▾]
   If YES, how many? ____ classes

Subjective Probability Analysis

Test ID: 4194E.
User ID: User5.

This part will be done on paper. You will be given a copy of MBTI Booklet and an answer sheet. Please put your Test ID and User ID on the answer sheet. The Test ID and User ID are used to ensure information ***consistency and confidentiality.***

The Booklet contains 93 paired-choice questions. Don't make any marks on the Booklet. Please mark your choices on the answer sheet only.

Your responses to this part don't account for your payment. There are no "right" or "wrong" answers. Feel free to make the choice with which you feel most comfortable.

You will be allowed a maximum of 20 minutes to complete this part.

Are there any questions before we begin?

To be read after completion of the survey.

Has everyone completed the survey? Please return the Booklet and your answer sheet to me.

Now, you will go through the remaining three sections of today's experiment which you will have the opportunity to earn money. All these three sections will be done on computer.

Please click on the "Next" button to go to next page.

Subjective Probability Analysis

Test ID: 4194E.
User ID: User5.

- This part consists of 10 multiple-choice questions. You will be allowed a maximum of 10 minutes to complete the entire section.
- You will earn $0.50 per question if your answer is correct.
- When you complete, please click on "Submit" button to submit your answers and go to the next page.

Please click on the "Next" button to go to next page.

Statistical Evaluation

1. The most often observed point of a distribution of numbers can be called:

- ☐ Mean
- ☐ Median
- ☐ Mode
- ☐ None of the Above

2. The standard deviation measures:

- ☐ The dispersion of numbers around a critical point
- ☐ The normal means of changing numbers
- ☐ The typical change required to make something important
- ☐ All of the Above

3. If there are three green balls in an urn with 100 total balls, what is the probability that you will draw a green ball from the urn on the first draw?

- ☐ 15%
- ☐ 1%
- ☐ 3%
- ☐ 30%

4. If Fred has a 50% chance of wearing an orange polka-dot tie to work and Barney has a 50% chance of wearing an orange polka-dot tie to work, what is the probability that both will wear polka-dot ties to work?

- ☐ 5%
- ☐ 15%
- ☐ 25%
- ☐ None of the Above

5. If there is a 10% chance of rain today, what is the probability that

	it will not rain?
	☐ 90% ☐ 60% ☐ 20% ☐ None of the Above

6.	A sample can be described as what?
	☐ A collection of items with no relationship to one another ☐ A small subset drawn from a larger population of items ☐ A taste test ☐ None of the Above

7.	Skewness of a distribution comes from what?
	☐ An incorrect assumption about the distribution ☐ Extreme values at one end of the distribution ☐ An incorrect use of statistical software ☐ None of the Above

8.	If two random variables tend to move together such that if one goes up the other also goes up, this is called:
	☐ independence ☐ positive correlation ☐ negative correlation ☐ kurtosis

9.	If the probability of rolling a 6 with a single die on the first toss is 1/6 (or 16.67%), what is the probability of rolling a 6 on the second roll of the die?
	☐ 1/6 ☐ 1/2 ☐ 1/32 ☐ None of the Above

10.	Assume that there is a 50% chance that you may get infected with the flying gygoutous. Assume that if you are infected, the probability of dying from the disease is 10%. What is the probability that you will die from the flying gygoutous?
	☐ 10% ☐ 15% ☐ 5% ☐ None of the Above

Part Three: The Risk Preference Survey

Subjective Probability Analysis

Test ID: 4194E.
User ID: User5.

This part consists of 10 **paired-choice** questions. There is no standard answer to each question. You will be allowed a maximum of 10 minutes to complete the entire section.

Before you starting making your ten choices, please let me explain how these choices will affect your payment in this part.

Question	**Option A**	**Option B**	**Which option is preferred?**
1	10% chance of $10.00, 90% chance of $8.00	10% chance of $19.00, 90% chance of $1.00	
2	20% chance of $10.00, 80% chance of $8.00	20% chance of $19.00, 80% chance of $1.00	
3	30% chance of $10.00, 70% chance of $8.00	30% chance of $19.00, 70% chance of $1.00	
4	40% chance of $10.00, 60% chance of $8.00	40% chance of $19.00, 60% chance of $1.00	
5	50% chance of $10.00, 50% chance of $8.00	50% chance of $19.00, 50% chance of $1.00	
6	60% chance of $10.00, 40% chance of $8.00	60% chance of $19.00, 40% chance of $1.00	
7	70% chance of $10.00, 30% chance of $8.00	70% chance of $19.00, 30% chance of $1.00	
8	80% chance of $10.00, 20% chance of $8.00	80% chance of $19.00, 20% chance of $1.00	
9	90% chance of $10.00, 10% chance of $8.00	90% chance of $19.00, 10% chance of $1.00	
10	100% chance of $10.00, 0% chance of $8.00	100% chance of $19.00, 0% chance of $1.00	

This table contains the ten questions you will be required to make decisions. In each question there are two choices, “Option A” or “Option B”. Option A has two paired events (Event A: $10 or Event B: $8). Option B has two paired events (Event C: $19 or Event D: $1). Each event has been allocated with different probability. You make the decision, either Option A or Option B as you prefer.

After you submit your responses, the computer will make two round random selections to decide your payment.

The first round random selection the computer will make is to select which of the ten questions as well as your decisions will be used. So you will make ten decisions, but only one of them will be used in the end to affect your payment. You will not know which one will be used in advance. The programming is insured that each question as well as each of your decision has an equal chance of being used.

The second round random selection the computer will make is to select which event of your decision in the chosen question will be used to determine your payment. The programming is insured that each event will be selected based on its probability.

Question	Option A	Option B	Your decision
1	10% chance of $10.00, 90% chance of $8.00 which means: Event A: $10; Probability: 10% Event B: $8; Probability: 90%	10% chance of $19.00, 90% chance of $1.00 which means: Event C: $19; Probability: 10% Event D: $1; Probability: 90%	Option B
10	100% chance of $10.00, 0% chance of $8.00 which means: Event A: $10.00; Probability: 100% Event B: $8.00; Probability: 0%	100% chance of $19.00, 0% chance of $1.00 which means: Event C: $19; Probability: 100% Event D: $1; Probability: 0%	Option A

Suppose in the first round random selection, Question 1 has been selected by the computer. If your decision is Option B, in the second round random selection, the computer will select either “Event C: 10% chance of $19.00” or “Event D: 90% chance of $1.00” in Option B as your payment. If you are lucky, you may earn $19. Here, Event C has 10% chance to be selected. If you are not lucky, you may earn $1 since Event D has 90% chance to be selected too.

If your decision is Option A, in the second round random selection, you may earn either $10 (if Event A has been selected) or $8 (if Event B has been selected). In this case, Event A has a probability of 10% while Event B has a probability of 90% to be chosen.

Suppose in the first round random selection, Question 10 has been selected by the computer. If your decision is Option A, you will earn $10 definitely since only Event A will occur. If your decision is Option B, you will earn $19 since Event C is the only choice.

Generally speaking, each question consists of four potential events ($10, $8, $19, $1). After the first round random selection, there will be only two potential events left ($10/$8 or $19/$1). After the second round random selection, the final event will be chosen and your payment will be determined. Obviously, your payment will be one of these four numbers: $10, $8, $19, $1. Your decision and the chance will jointly determine your payment.

When you complete, please click on "Submit" button to submit your answers and go to the next page.

Are there any questions before we begin?

Please click on the "Next" button to go to next page.

Holt and Laury Procedure

1 ☐ Option A: 10% chance of $10.00, 90% chance of $8.00
☐ Option B: 10% chance of $19.00, 90% chance of $1.00

2 ☐ Option A: 20% chance of $10.00, 80% chance of $8.00
☐ Option B: 20% chance of $19.00, 80% chance of $1.00

3 ☐ Option A: 30% chance of $10.00, 70% chance of $8.00
☐ Option B: 30% chance of $19.00, 70% chance of $1.00

4 ☐ Option A: 40% chance of $10.00, 60% chance of $8.00
☐ Option B: 40% chance of $19.00, 60% chance of $1.00

5 ☐ Option A: 50% chance of $10.00, 50% chance of $8.00
☐ Option B: 50% chance of $19.00, 50% chance of $1.00

6 ☐ Option A: 60% chance of $10.00, 40% chance of $8.00
☐ Option B: 60% chance of $19.00, 40% chance of $1.00

7 ☐ Option A: 70% chance of $10.00, 30% chance of $8.00
☐ Option B: 70% chance of $19.00, 30% chance of $1.00

8 ☐ Option A: 80% chance of $10.00, 20% chance of $8.00
☐ Option B: 80% chance of $19.00, 20% chance of $1.00

9 ☐ Option A: 90% chance of $10.00, 10% chance of $8.00
☐ Option B: 90% chance of $19.00, 10% chance of $1.00

10 ☐ Option A: 100% chance of $10.00, 0% chance of $8.00

	Option B: 100% chance of $19.00, 0% chance of $1.00

Holt and Laury Pay Out

Below is a summary of your choices for the Holt and Laury procedure along with the randomly selected value. Your choices are marked in green and the computer-chosen value is shown in red. Your final payout is shown at the bottom of this page.

1	Option A: 10% chance of $10.00, 90% chance of $8.00
	Option B: 10% chance of $19.00, 90% chance of $1.00
2	Option A: 20% chance of $10.00, 80% chance of $8.00
	Option B: 20% chance of $19.00, 80% chance of $1.00
3	Option A: 30% chance of $10.00, 70% chance of $8.00
	Option B: 30% chance of $19.00, 70% chance of $1.00
4	Option A: 40% chance of $10.00, 60% chance of $8.00
	Option B: 40% chance of $19.00, 60% chance of $1.00
5	Option A: 50% chance of $10.00, 50% chance of $8.00
	Option B: 50% chance of $19.00, 50% chance of $1.00
6	Option A: 60% chance of $10.00, 40% chance of $8.00
	Option B: 60% chance of $19.00, 40% chance of $1.00
7	Option A: 70% chance of $10.00, 30% chance of $8.00
	Option B: 70% chance of $19.00, 30% chance of $1.00
8	Option A: 80% chance of $10.00, 20% chance of $8.00
	Option B: 80% chance of $19.00, 20% chance of $1.00
9	Option A: 90% chance of $10.00, 10% chance of $8.00
	Option B: 90% chance of $19.00, 10% chance of $1.00
10	Option A: 100% chance of $10.00, 0% chance of $8.00
	Option B: 100% chance of $19.00, 0% chance of $1.00

Part Four: The Subjective Probability Solicitation Section

Subjective Probability Analysis

Test ID: 4194E.
User ID: User5.

In this section, you will have four chances to earn extra money. Each time you will be provided one dataset which is drawn randomly from a known population. You need to download the dataset, save it at the Desktop and after that, use Excel to estimate the population probability distribution based on the given dataset.

For example, there are 50 numbers in the sample dataset randomly drawn from uniform distributed population within range between 75 and 125.

SAMPLE DATA

◆	95.02624592	◆	102.9976806	◆	119.9781793
◆	84.13113804	◆	82.39921262	◆	117.5122837
◆	106.3867611	◆	119.8164312	◆	106.7056795
◆	120.2986236	◆	105.9625538	◆	117.4359874
◆	113.3587146	◆	101.5221107	◆	104.197058
◆	124.2996002	◆	80.27054659	◆	100.4203925
◆	112.5988037	◆	90.49729911	◆	77.16681417
◆	84.89715262	◆	117.1430097	◆	120.8586383
◆	79.71510971	◆	99.96871853	◆	82.41752373
◆	115.1654103	◆	83.98159734	◆	101.5602588
◆	108.9442122	◆	97.1396527	◆	112.5270852
◆	122.7141636	◆	99.06079287	◆	109.8673971
◆	84.63316752	◆	83.15454573	◆	95.84719382
◆	98.53282266	◆	94.19309061	◆	122.727897
◆	104.8089541	◆	97.45857112	◆	106.1532945
◆	76.21921445	◆	97.45857112	◆	106.1532945
◆	113.8027589	◆	124.4934	◆	83.096560
		◆	87.37983		

The Datasets you will be given later in the solicitation procedure will be something like this sample. But they are not definitely drawn from a uniform distributed population.

Please click on the "Next" button to go to next page.

Test ID: 4194E.
User ID: User5.

Now let me explain how your payment will be determined based on your estimation.

Payment = Max ($0.00, $10.00 – 0.025 * Total Squared Deviation)

Example:

%	**Below 80**	**80-90**	**90-100**	**100-110**	**110-120**	**Above 120**
True distribution	10	20	20	20	20	10
Your response	5	20	28	25	20	2
Squared deviation	(5-10)^2=25	(20-20)^2 =0	(28-20)^2 =64	(25-20)^2 =25	(20-18)^2 =4	(2-10)^2 =64

Total Squared Deviation: 25 + 0 + 64 + 25 + 4 + 64 = 182
$10.00 – 0.025 * 182 = $5.45
Your Payment: Max ($0.00, $5.45) = **$5.45**

In the example, the known population probability the variable will fall below 80 next period is 10 (%). Suppose your estimation is 5 (%). Then the deviation is -5 (%). The squared deviation is 25 (%%). In the range between 80 and 90, the known population probability the variable will fall next period is 20 (%). Suppose your estimation is 20 (%). Then the deviation is 0 (%). The squared deviation is 0(%%). Overall the total squared deviation is 182 (%%). The payment is Max ($0.00, $5.45) = $5.45.

According to the payment rule, you will earn money if your total squared deviation is less than 400. If your total squared deviation is greater than 400, your payment will be 0. In any case, you will not lose money. The rule of thumb is “the smaller your total squared deviation, the more money you earn”.

In the end of this part, there will be a number indicating your show up fee plus 6 payments. This will be your total compensation in today’s experiment. Please write this number on the payment sheet. Sign the payment sheet and return it to me.

When you complete, please click on "Submit" button to submit your answers and go to the next page.

Are there any questions before we begin?

Please click on the "Next" button to go to next page.

Data Analysis Dataset: 1 (Total Datasets:4) Round: 1

Subjective Probability Analysis

Please click "Download and Save" button to download the *DATASET1*, and save it as dataset1.xls at the Desktop.Then open the dataset1.xls file, it should be an Excel spreadsheet containing given data. Now you can use any techniques to analyse the data and answer the following six questions.

You will be allowed a maximum of 10 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.

Download and Save

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:09:41

Data Analysis Dataset: 1 (Total Datasets:4) Round: 2

Subjective Probability Analysis

You will be allowed a maximum of 4 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat showothers

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:51

Data Analysis Dataset: 1 (Total Datasets:4) Round: 3

Subjective Probability Analysis

Your answers in this round will be used to calculate your payout. You will be allowed a maximum of 4 minute to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn. Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat

showothers

What is the chance the variable will fall below 80 next period? %

What is the chance the variable will fall at or above 80 and below 90 next period? %

What is the chance the variable will fall at or above 90 and below 100 next period? %

What is the chance the variable will fall at or above 100 and below 110 next period? %

What is the chance the variable will fall at or above 110 and below 120 next period? %

What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:53

Data Analysis Dataset: 2 (Total Datasets:4) Round: 1

Subjective Probability Analysis

Please click "Download and Save" button to download the *DATASET2*, and save it as dataset2.xls at the Desktop.Then open the dataset2.xls file, it should be an Excel spreadsheet containing given data. Now you can use any techniques to analyse the data and answer the following six questions.

You will be allowed a maximum of 10 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.

Download and Save

What is the chance the variable will fall below 80 next period? %

What is the chance the variable will fall at or above 80 and below 90 next period? %

What is the chance the variable will fall at or above 90 and below 100 next period? %

What is the chance the variable will fall at or above 100 and below 110 next period? %

What is the chance the variable will fall at or above 110 and below 120 next period? %

What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:09:49

Data Analysis Dataset: 2 (Total Datasets:4) Round: 2

Subjective Probability Analysis

You will be allowed a maximum of 4 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat showothers

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:56

Data Analysis Dataset: 2 (Total Datasets:4) Round: 3

Subjective Probability Analysis

Your answers in this round will be used to calculate your payout. You will be allowed a maximum of 4 minute to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn. Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat showothers

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:54

Data Analysis Dataset: 3 (Total Datasets:4) Round: 1

Subjective Probability Analysis

Please click "Download and Save" button to download the *DATASET3*, and save it as dataset3.xls at the Desktop.Then open the dataset3.xls file, it should be an Excel spreadsheet containing given data. Now you can use any techniques to analyse the data and answer the following six questions.

You will be allowed a maximum of 10 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.

Download and Save

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:09:53

Data Analysis Dataset: 3 (Total Datasets:4) Round: 2

Subjective Probability Analysis

You will be allowed a maximum of 4 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat showothers

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:56

Data Analysis Dataset: 3 (Total Datasets:4) Round: 3

Subjective Probability Analysis

Your answers in this round will be used to calculate your payout. You will be allowed a maximum of 4 minute to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn. Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat showothers

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:56

Data Analysis Dataset: 4 (Total Datasets:4) Round: 1

Subjective Probability Analysis

Please click "Download and Save" button to download the *DATASET4*, and save it as dataset4.xls at the Desktop.Then open the dataset4.xls file, it should be an Excel spreadsheet containing given data. Now you can use any techniques to analyse the data and answer the following six questions.

You will be allowed a maximum of 10 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.

Download and Save

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:09:55

Data Analysis Dataset: 4 (Total Datasets:4) Round: 2

Subjective Probability Analysis

You will be allowed a maximum of 4 minutes to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn.Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat

showothers

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:57

Data Analysis Dataset: 4 (Total Datasets:4) Round: 3

Subjective Probability Analysis

Your answers in this round will be used to calculate your payout. You will be allowed a maximum of 4 minute to complete this round. *Be careful* ! you are asked to estimate the distribution of the population from which the dataset was drawn. Click on "Show Others" button below to see the answers of all users of your group. Click on "Join Chat" button to chat with your group members.

Join Chat

showothers

What is the chance the variable will fall below 80 next period? %	What is the chance the variable will fall at or above 80 and below 90 next period? %	What is the chance the variable will fall at or above 90 and below 100 next period? %
What is the chance the variable will fall at or above 100 and below 110 next period? %	What is the chance the variable will fall at or above 110 and below 120 next period? %	What is the chance the variable will fall at or above 120 next period? %

Submit

Note: Sum of the your Six Answers should be 100. At present sum of your Six Answers is : 0.00

00:03:57

APPENDIX B

RAW DATA OF THE EXPERIMENT

Treatment 1

Group One		B93EB/4	BF62C/3	1F95A/5	688CD/5	7A754/3
		Subject (Test ID/User ID)				
MBTI	E/I	E	E	E	I	I
	S/N	N	N	N	S	S
	T/F	T	T	F	F	T
	J/P	P	P	J	J	J
Statistical Score		8	7	8	10	5
Risk Preference		Switch				Switch
Number of safe choice		NA	6	6	5	NA
Dataset One	(-∞, 80)	13	5	13.3	13.33	10
	[80, 90)	17	60	16.7	16.67	17
	[90, 100)	30	10	36.7	36.67	37
	[100, 110)	15	10	6.7	6.67	7
	[110, 120)	13	10	13.3	13.33	14
	[120,+∞)	12	5	13.3	13.33	15
Dataset Two	(-∞, 80)	20	10	20	20	20
	[80, 90)	14	20	16	16	16
	[90, 100)	18	20	18	18	18
	[100, 110)	22	20	20	22	20
	[110, 120)	17	20	8	6	8
	[120,+∞)	9	10	18	18	18
Dataset Three	(-∞, 80)	1	10	1	1	1
	[80, 90)	11	20	9	9	9
	[90, 100)	36	20	37	37	38
	[100, 110)	39	20	40	40	39
	[110, 120)	11	20	11	11	11
	[120,+∞)	2	10	2	2	2
Dataset Four	(-∞, 80)	3	5	20	20	20
	[80, 90)	50	10	16.7	16.67	17
	[90, 100)	20	35	20	20	17
	[100, 110)	6	35	20	20	20
	[110, 120)	1	10	3.3	3.33	6
	[120,+∞)	20	5	20	20	20

Treatment 1

Group Two		Subject (Test ID/User ID) B93EB/1	B93EB/5	BF62C/4	BF536/4	688CD/2
MBTI	E/I	I	I	I	I	E
	S/N	N	N	N	N	N
	T/F	T	F	F	T	F
	J/P	J	J	P	J	P
Statistical Score		7	8	9	9	8
Risk Preference					Switch	
Number of safe choice		4	8	4	NA	6
Dataset One	(-∞, 80)	16	13	12	13	10
	[80, 90)	18	16	19	17	30
	[90, 100)	20	36	35	37	40
	[100, 110)	12	6	8	7	3
	[110, 120)	18	13	15	13	3
	[120,+∞)	16	16	11	13	14
Dataset Two	(-∞, 80)	16	20	19	20	20
	[80, 90)	17	16	16	16	15
	[90, 100)	18	18	19	18	20
	[100, 110)	17	20	20	20	20
	[110, 120)	16	8	7	8	5
	[120,+∞)	16	18	19	18	20
Dataset Three	(-∞, 80)	2	1	1	1	0
	[80, 90)	10	9	9	9	5
	[90, 100)	36	37	37	37	40
	[100, 110)	40	40	40	40	45
	[110, 120)	10	11	11	11	5
	[120,+∞)	2	2	2	2	5
Dataset Four	(-∞, 80)	20	20	20	20	15
	[80, 90)	16	17	16	17	10
	[90, 100)	20	20	20	20	30
	[100, 110)	20	20	22	20	30
	[110, 120)	4	9	2	3	0
	[120,+∞)	20	14	20	20	15

Treatment 1

Group Three		BF62C/1	BF62C/2	1F95A/2	1F95A/4	7A754/4
		Subject (Test ID/User ID)				
MBTI	E/I	E	E	E	E	E
	S/N	N	N	N	N	S
	T/F	F	F	F	F	F
	J/P	J	P	J	P	P*
Statistical Score		7	10	8	7	7
Risk Preference		Switch				
Number of safe choice		NA	5	4	9	3
Dataset One	(-∞, 80)	14	13	13.32	13.3333	10
	[80, 90)	17	17	16.65	16.6667	20
	[90, 100)	36	37	36.63	36.6667	20
	[100, 110)	7	7	6.66	6.6667	20
	[110, 120)	13	13	13.32	13.3333	20
	[120,+∞)	13	13	13.42	13.3333	10
Dataset Two	(-∞, 80)	20	20	20	20	5
	[80, 90)	14	16	16	16	20
	[90, 100)	18	18	18	18	30
	[100, 110)	20	20	20	20	20
	[110, 120)	8	8	8	8	15
	[120,+∞)	20	18	18	18	10
Dataset Three	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	10
	[90, 100)	37	38	37	37	40
	[100, 110)	40	39	40	40	35
	[110, 120)	11	11	11	11	10
	[120,+∞)	2	2	2	2	4
Dataset Four	(-∞, 80)	20	20	19.98	23.3333	25
	[80, 90)	17	17	16.65	13.3333	10
	[90, 100)	20	20	19.98	20	10
	[100, 110)	20	20	19.98	20	20
	[110, 120)	3	3	3.33	3.3334	1
	[120,+∞)	20	20	20.08	20	34

* means it is equal between this dichotomy

Treatment 1

Group Four		14FF5/3	14FF5/5	BF536/5	688CD/1	688CD/3
		Subject (Test ID/User ID)				
MBTI	E/I	I	E	I	E	I
	S/N	S	N	S	S	S
	T/F	T	F	T	T	F
	J/P	J	J	P	P	P
Statistical Score		6	8	9	10	8
Risk Preference		Switch				
Number of safe choice		NA	4	5	3	5
Dataset One	(-∞, 80)	15	13	13	13	20
	[80, 90)	25	17	17	17	10
	[90, 100)	15	37	37	37	15
	[100, 110)	15	7	7	7	15
	[110, 120)	20	13	13	13	20
	[120,+∞)	10	13	13	13	20
Dataset Two	(-∞, 80)	15	20	20	20	10
	[80, 90)	25	16	16	16	15
	[90, 100)	25	18	18	18	20
	[100, 110)	25	20	20	20	20
	[110, 120)	5	8	8	8	10
	[120,+∞)	5	18	18	18	25
Dataset Three	(-∞, 80)	10	1	1	1	5
	[80, 90)	15	9	9	9	15
	[90, 100)	20	37	37	37	20
	[100, 110)	30	40	40	40	15
	[110, 120)	15	11	11	11	15
	[120,+∞)	10	2	2	2	30
Dataset Four	(-∞, 80)	25	20	20	20	5
	[80, 90)	25	17	17	16.7	10
	[90, 100)	10	20	20	20	15
	[100, 110)	25	20	20	20	15
	[110, 120)	10	3	3	3.3	5
	[120,+∞)	5	20	20	20	50

Treatment 1

Group Five		Subject (Test ID/User ID) 14FF5/4	BF536/2	1F95A/3	688CD/4	7A754/2
MBTI	E/I	E	E	E	I	E
	S/N	N	S	N	N	S
	T/F	T	T	F	T	T
	J/P	J	J	J	J	J
Statistical Score		10	9	9	8	8
Risk Preference Number of safe choice		5	7	8	4	4
Dataset One	(-∞, 80)	13.33	13	13	10	5
	[80, 90)	16.67	17	17	18	10
	[90, 100)	36.67	37	37	50	50
	[100, 110)	6.67	7	7	3	5
	[110, 120)	13.33	13	13	5	10
	[120,+∞)	13.33	13	13	14	20
Dataset Two	(-∞, 80)	22	20	20	18	30
	[80, 90)	14	16	16	16	10
	[90, 100)	18	18	18	18	15
	[100, 110)	18	20	20	20	15
	[110, 120)	10	8	8	6	5
	[120,+∞)	18	18	18	22	25
Dataset Three	(-∞, 80)	2	1	1	1	1
	[80, 90)	9	9	9	8	9
	[90, 100)	36	37	37	35	35
	[100, 110)	34	40	40	40	35
	[110, 120)	12	11	11	11	10
	[120,+∞)	7	2	2	5	10
Dataset Four	(-∞, 80)	20	20	20	20	20
	[80, 90)	17	17	17	17	15
	[90, 100)	20	20	20	20	20
	[100, 110)	20	20	20	35	20
	[110, 120)	3	3	3	0	10
	[120,+∞)	20	20	20	8	15

Treatment 2

Group One		Subject (Test ID/User ID) D2E33/1	D2E33/2	D2E33/3	D2E33/4	D2E33/5
MBTI	E/I	I	I	E	I	I
	S/N	N	N	S	N	N
	T/F	T	T	F	T	F
	J/P	J	J	P	J	P
Statistical Score		8	10	7	10	10
Risk Preference			Switch	Switch		
Number of safe choice		5	NA	NA	4	4
Dataset One	(-∞, 80)	18	13.33	13	13.33	13.33
	[80, 90)	30	16.67	16	16.67	16.67
	[90, 100)	30	36.67	35	36.67	36.67
	[100, 110)	9	6.67	11	6.67	6.67
	[110, 120)	8	13.33	12.5	13.33	13.33
	[120,+∞)	5	13.33	12.5	13.33	13.33
Dataset Two	(-∞, 80)	20	20	20	20	20
	[80, 90)	25	16	16	16	16
	[90, 100)	20	18	19	18	18
	[100, 110)	13	20	20	20	20
	[110, 120)	10	8	8	8	8
	[120,+∞)	12	18	17	18	18
Dataset Three	(-∞, 80)	8	1	1	1	1
	[80, 90)	18	9	9.5	9	9
	[90, 100)	24	37	36.5	37	37
	[100, 110)	24	40	40	40	40
	[110, 120)	18	11	11	11	11
	[120,+∞)	8	2	2	2	2
Dataset Four	(-∞, 80)	20	20	19.5	20	20
	[80, 90)	16	16.67	16	16.67	16.67
	[90, 100)	16	20	20	20	20
	[100, 110)	16	20	21	20	20
	[110, 120)	14	3.33	3.5	3.33	3.33
	[120,+∞)	18	20	20	20	20

Treatment 2

Group Two		Subject (Test ID/User ID)				
		DC98A/1	DC98A/2	DC98A/3	DC98A/4	DC98A/5
MBTI	E/I	E	E	I	E	E
	S/N	N	S	S	S	S
	T/F	F	T	F	T	T
	J/P	P	P	J	P*	P*
Statistical Score		9	10	6	9	9
Risk Preference				Switch		
Number of safe choice		3	4	NA	4	6
Dataset One	(-∞, 80)	13.33	13.33	13.33	13.33	13.33
	[80, 90)	16.67	16.66	16.67	16.66	16.67
	[90, 100)	36.67	36.67	33.66	36.66	36.67
	[100, 110)	6.67	6.66	6.66	6.66	6.67
	[110, 120)	13.33	13.33	13.66	13.33	13.33
	[120,+∞)	13.33	13.35	16.02	13.36	13.33
Dataset Two	(-∞, 80)	20	20	20	20	20
	[80, 90)	16	16	16	16	16
	[90, 100)	18	18	18	18	18
	[100, 110)	20	20	20	20	20
	[110, 120)	8	8	8	8	8
	[120,+∞)	18	18	18	18	18
Dataset Three	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	9
	[90, 100)	37	37	37	37	37
	[100, 110)	40	40	40	40	40
	[110, 120)	11	11	11	11	11
	[120,+∞)	2	2	2	2	2
Dataset Four	(-∞, 80)	20	20	20	20	20
	[80, 90)	16.67	16.66	16	16.66	16.67
	[90, 100)	20	20	20	20	20
	[100, 110)	20	20	20	20	20
	[110, 120)	3.33	3.34	3.34	3.33	3.33
	[120,+∞)	20	20	20.66	20.01	20

* means it is equal between this dichotomy

Treatment 2

Group Three		Subject (Test ID/User ID) 76444/1	76444/2	76444/3	76444/4	76444/5
MBTI	E/I	I	I	I	E	I
	S/N	S	N	S	N*	N*
	T/F	T	T	T	T	T
	J/P	J	J	J	J	P
Statistical Score		10	9	9	8	8
Risk Preference Number of safe choice		4	4	4	4	4
Dataset One	(-∞, 80)	13.33	13.33	13.33	13.3	13.33
	[80, 90)	16.67	16.66	16.66	16.67	16.67
	[90, 100)	36.67	36.66	6.66	36.7	36.67
	[100, 110)	6.67	6.66	36.69	6.73	6.67
	[110, 120)	13.33	13.33	13.33	13.3	13.33
	[120,+∞)	13.33	13.36	13.33	13.3	13.33
Dataset Two	(-∞, 80)	20	20	20	20	20
	[80, 90)	16	16	16	16	16
	[90, 100)	18	18	18	18	18
	[100, 110)	20	20	20	20	20
	[110, 120)	8	8	8	8	8
	[120,+∞)	18	18	18	18	18
Dataset Three	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	9
	[90, 100)	37	37	37	37	37
	[100, 110)	40	40	40	40	40
	[110, 120)	11	11	11	11	11
	[120,+∞)	2	2	2	2	2
Dataset Four	(-∞, 80)	20	20	20	20	20
	[80, 90)	16.67	16.67	16.67	16.67	16.67
	[90, 100)	20	20	20	20	20
	[100, 110)	20	20	20	20	20
	[110, 120)	3.33	3.33	3.33	3.33	3.33
	[120,+∞)	20	20	20	20	20

* means it is equal between this dichotomy

Treatment 2

Group Four		Subject (Test ID/User ID) 66383/1	66383/2	66383/3	66383/4	66383/5
	E/I	I	I	E	I	I
MBTI	S/N	N	N	N	N	N
	T/F	F*	F	F	F	T
	J/P	J	P	J	P	P
Statistical Score		8	7	4	9	9
Risk Preference				Switch	Switch	
Number of safe choice		4	8	NA	NA	3
	(-∞, 80)	13.33	15	15	13.33	13.33
	[80, 90)	16.67	15	15	16.67	16.67
Dataset One	[90, 100)	36.67	35	35	36.67	36.67
	[100, 110)	6.67	5	5	6.67	6.67
	[110, 120)	13.33	15	15	13.33	13.33
	[120,+∞)	13.33	15	15	13.33	13.33
	(-∞, 80)	18	18	20	20	20
	[80, 90)	14	16	15	16	16
Dataset Two	[90, 100)	18	16	15	18	18
	[100, 110)	18	20	20	20	20
	[110, 120)	14	10	10	8	8
	[120,+∞)	18	20	20	18	18
	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	9
Dataset Three	[90, 100)	37	37	37	37	37
	[100, 110)	40	40	40	40	40
	[110, 120)	11	11	11	11	11
	[120,+∞)	2	2	2	2	2
	(-∞, 80)	20	20	20	20	20
	[80, 90)	16.67	16	16.67	16.67	16.67
Dataset Four	[90, 100)	20	20	20	20	20
	[100, 110)	20	20	20	20	20
	[110, 120)	3.33	4	3.33	3.33	3.33
	[120,+∞)	20	20	20	20	20

* means it is equal between this dichotomy

Treatment 2

Group Five		Subject (Test ID/User ID)				
		5D149/1	5D149/2	5D149/3	5D149/4	5D149/5
MBTI	E/I	E	E	E	E	I
	S/N	N	N	N	S	S
	T/F	F	F	F	F*	T
	J/P	P	J	J	P*	J
Statistical Score		9	7	8	10	8
Risk Preference			Switch			
Number of safe choice		6	NA	8	4	6
Dataset One	(-∞, 80)	13.33	12	13	13.33	13.33
	[80, 90)	16.67	16	16.6	16.67	16.67
	[90, 100)	36.67	35	35.3	36.67	36.67
	[100, 110)	6.67	8	8.1	6.67	6.67
	[110, 120)	13.33	13.5	13.5	13.33	13.33
	[120,+∞)	13.33	15.5	13.5	13.33	13.33
Dataset Two	(-∞, 80)	20	17	19.25	20	20
	[80, 90)	15	16	16	16	16
	[90, 100)	15	17	18	18	18
	[100, 110)	15	17	19.25	20	20
	[110, 120)	15	16	9.5	8	8
	[120,+∞)	20	17	18	18	18
Dataset Three	(-∞, 80)	2	1.5	1.5	1	1
	[80, 90)	10	10	9	9	9
	[90, 100)	38	38.5	37	37	37
	[100, 110)	38	38.5	39.75	40	40
	[110, 120)	10	10	10	11	11
	[120,+∞)	2	1.5	2.75	2	2
Dataset Four	(-∞, 80)	19	17	19.25	20	20
	[80, 90)	19	17	16.5	16.67	16.67
	[90, 100)	19	17	19.25	20	20
	[100, 110)	19	16.5	19.5	20	20
	[110, 120)	5	16.5	6	3.33	3.33
	[120,+∞)	19	16	19.5	20	20

* means it is equal between this dichotomy

Treatment 3

Group One		B5159/1	B5159/2	B5159/3	B5159/4	B5159/5
		Subject (Test ID/User ID)				
MBTI	E/I	E	E	E	I	I
	S/N	N	N	N	S	S
	T/F	F	T	F	T	T
	J/P	P	J	P	J	P
Statistical Score		6	7	8	6	7
Risk Preference		Switch		Switch	Switch	
Number of safe choice		NA	10	NA	NA	4
Dataset One	(-∞, 80)	10	10	10	10	10
	[80, 90)	20	30	15	15	30
	[90, 100)	25	10	15	20	20
	[100, 110)	15	15	20	30	10
	[110, 120)	20	25	15	15	20
	[120,+∞)	10	10	25	10	10
Dataset Two	(-∞, 80)	20	20	20	20	20
	[80, 90)	16	16	16	16	16
	[90, 100)	20	18	18	18	18
	[100, 110)	20	20	20	20	20
	[110, 120)	8	8	8	8	8
	[120,+∞)	16	18	18	18	18
Dataset Three[a]	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	9
	[90, 100)	46	46	46	46	46
	[100, 110)	31	31	31	31	31
	[110, 120)	11	11	11	11	11
	[120,+∞)	2	2	2	2	2
Dataset Four	(-∞, 80)	18	18	20	20	20
	[80, 90)	18	18	17	17	17
	[90, 100)	18	18	20	20	20
	[100, 110)	18	18	20	20	20
	[110, 120)	10	10	3	3	3
	[120,+∞)	18	18	20	20	20

[a] means a consensus

Treatment 3

Group Two		Subject (Test ID/User ID) EF453/1	EF453/2	EF453/3	EF453/4	EF453/5
MBTI	E/I	E	I	I	E	E
	S/N	S	N	N	S	S
	T/F	F	F	T	T	T
	J/P	J	P	J	P	J
Statistical Score		5	10	8	9	6
Risk Preference						Switch
Number of safe choice		6	6	3	6	NA
Dataset One	(-∞, 80)	13	15	13.33	13.33	13.33
	[80, 90)	17	25	16.67	16.67	16.67
	[90, 100)	37	35	36.67	36.67	36.67
	[100, 110)	7	15	6.67	6.67	6.67
	[110, 120)	13	5	13.33	13.33	13.33
	[120,+∞)	13	5	13.33	13.33	13.33
Dataset Two[a]	(-∞, 80)	20	20	20	20	20
	[80, 90)	16	16	16	16	16
	[90, 100)	18	18	18	18	18
	[100, 110)	20	20	20	20	20
	[110, 120)	8	8	8	8	8
	[120,+∞)	18	18	18	18	18
Dataset Three[a]	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	9
	[90, 100)	37	37	37	37	37
	[100, 110)	40	40	40	40	40
	[110, 120)	11	11	11	11	11
	[120,+∞)	2	2	2	2	2
Dataset Four[a]	(-∞, 80)	20	20	20	20	20
	[80, 90)	16.67	16.67	16.67	16.67	16.67
	[90, 100)	20	20	20	20	20
	[100, 110)	20	20	20	20	20
	[110, 120)	3.33	3.33	3.33	3.33	3.33
	[120,+∞)	20	20	20	20	20

[a] means a consensus

Treatment 3

Group Three		7B88A/1	7B88A/2	7B88A/3	7B88A/4	7B88A/5
		Subject (Test ID/User ID)				
MBTI	E/I	E	E	E	I	I
	S/N	N	N	N	S	S
	T/F	F	T	F	T	T
	J/P	P	J	P	J	P
Statistical Score		9	8	8	8	7
Risk Preference				Switch		Switch
Number of safe choice		7	4	NA	3	NA
Dataset One	(-∞, 80)	13.3	13.34	13.33	13.33	13.34
	[80, 90)	16.7	16.66	16.67	16.67	16.67
	[90, 100)	36.7	36.66	33.33	36.67	33.33
	[100, 110)	6.7	10	6.67	6.67	6.67
	[110, 120)	13.3	10	13.33	13.33	16.66
	[120,+∞)	13.3	13.34	16.67	13.33	13.33
Dataset Two[a]	(-∞, 80)	20	20	20	20	20
	[80, 90)	16	16	16	16	16
	[90, 100)	18	18	18	18	18
	[100, 110)	20	20	20	20	20
	[110, 120)	8	8	8	8	8
	[120,+∞)	18	18	18	18	18
Dataset Three[a]	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	9
	[90, 100)	37	37	37	37	37
	[100, 110)	40	40	40	40	40
	[110, 120)	11	11	11	11	11
	[120,+∞)	2	2	2	2	2
Dataset Four	(-∞, 80)	20	20	20	20	20
	[80, 90)	16.7	16.67	16.67	16.6	16.67
	[90, 100)	20	20	20	20	20
	[100, 110)	20	20	20	20	20
	[110, 120)	3.3	3.33	3.33	3.4	3.33
	[120,+∞)	20	20	20	20	20

[a] means a consensus

Treatment 3

Group Four		FEA25/1	FEA25/2	FEA25/3	FEA25/4	FEA25/5
		Subject (Test ID/User ID)				
MBTI	E/I	I	E	E	I	E
	S/N	S	S	N	S	S
	T/F	T	T	F	T	T
	J/P	J	J	P	J	J
Statistical Score		8	9	6	8	6
Risk Preference Number of safe choice		6	5	3	6	4
Dataset One[a]	(-∞, 80)	13.33	13.33	13.33	13.33	13.33
	[80, 90)	16.67	16.67	16.67	16.67	16.67
	[90, 100)	36.67	36.67	36.67	36.67	36.67
	[100, 110)	6.67	6.67	6.67	6.67	6.67
	[110, 120)	13.33	13.33	13.33	13.33	13.33
	[120,+∞)	13.33	13.33	13.33	13.33	13.33
Dataset Two	(-∞, 80)	20	8	20	20	20
	[80, 90)	16	12	16	16	16
	[90, 100)	18	18	18	18	18
	[100, 110)	20	30	20	20	20
	[110, 120)	8	18	8	8	8
	[120,+∞)	18	14	18	18	18
Dataset Three[a]	(-∞, 80)	1	1	1	1	1
	[80, 90)	9	9	9	9	9
	[90, 100)	37	37	37	37	37
	[100, 110)	40	40	40	40	40
	[110, 120)	11	11	11	11	11
	[120,+∞)	2	2	2	2	2
Dataset Four	(-∞, 80)	20	8	20	20	20
	[80, 90)	16.67	16	16.67	16.67	16.67
	[90, 100)	20	20	20	20	20
	[100, 110)	20	20	20	20	20
	[110, 120)	3.33	20	3.33	3.33	3.33
	[120,+∞)	20	16	20	20	20

[a] means a consensus

Treatment 3

Group Five		Subject (Test ID/User ID)				
		A41B2/1	A41B2/2	A41B2/3	A41B2/4	A41B2/5
MBTI	E/I	I	E	E	E	I
	S/N	N	S	N	N	N
	T/F	F	T	F	T	F
	J/P	P	P	J	P	J
Statistical Score		6	7	6	6	4
Risk Preference		Switch	Switch			
Number of safe choice		NA	NA	7	4	2
Dataset One	(-∞, 80)	12	12	10	10	5
	[80, 90)	15	15	20	20	25
	[90, 100)	30.3	33	20	37	20
	[100, 110)	6.6	6	20	6	20
	[110, 120)	12.4	12	20	13	25
	[120,+∞)	23.7	22	10	14	5
Dataset Two	(-∞, 80)	26	26	10	20	15
	[80, 90)	10.5	10.5	30	16	15
	[90, 100)	10.5	10.5	30	18	20
	[100, 110)	26	26	20	20	20
	[110, 120)	9	9	5	8	15
	[120,+∞)	18	18	5	18	15
Dataset Three	(-∞, 80)	1	1	1	1	10
	[80, 90)	12	12	12	12	10
	[90, 100)	37	37	37	37	30
	[100, 110)	37	37	37	37	20
	[110, 120)	12	12	12	12	20
	[120,+∞)	1	1	1	1	10
Dataset Four	(-∞, 80)	20	20	20	20	22
	[80, 90)	17	15	15	17	18
	[90, 100)	20	20	20	20	15
	[100, 110)	20	20	20	20	20
	[110, 120)	3	3	3	3	15
	[120,+∞)	20	22	22	20	10