To the Graduate Council:

I am submitting herewith a thesis written by David Eugene Vinson entitled "Simpson's Paradox." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

Carl G. Wagner

Carl G. Wagner, Major Professor

We have read this thesis
and recommend its acceptance:

Robert M. McConnel

Lawrence A. Bales

Accepted for the Council:

Lee Minkel

Vice Provost
and Dean of The Graduate School

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master of Science degree at The Univeristy of Tennessee, Knoxville, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of the source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his absence, by the Head of Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature David Eugene Vinson

Date November 26, 1985

SIMPSON'S PARADOX

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

David Eugene Vinson

December 1985

ACKNOWLEDGMENTS

The author is very thankful to his major professor, Dr. Carl Wagner, for his helpfulness and support as well as for his unending patience.

Further, the author is grateful to Drs. L. Bales, R. McConnel, and C. Wagner for the giving of their time and encouragement for the completion of this project.

Finally, the author wishes to thank both his wife, Carol Vinson, for her love, patience, and constant encouragement, and his parents, Mr. and Mrs. B. E. Vinson for their support throughout the years.

ABSTRACT

In this thesis, we present Simpson's Paradox, preliminary results necessary to understanding the paradox, sufficient conditions for blocking the paradox, and interpretations of the occurrence of the paradox in real life situations.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	1.1 Introduction.	1
II.	PRELIMINARY RESULTS.	6
	2.1 Subjective Probability.	6
	2.2 Reducing the Sample Space	7
	2.3 Probablistic Independence	8
	2.4 Relevance and Irrelevance	11
	2.5 Properties of Relevance	13
	2.6 Bayes' Rule	19
	2.7 Conditional Independence and Conditional Relevance	22
III.	SIMPSON'S PARADOX.	25
	3.1 Blocking Simpson's Paradox.	25
	3.2 The Berkeley Admissions Case.	33
	3.3 The Logical Status of Statistical Paradoxes	35
BIBLIOGRAPHY		37
VITA		39

CHAPTER I

INTRODUCTION

1.1 Introduction. Let X be a set of possible states of nature and let P be a probability measure defined on some algebra of events (i.e., subsets) in X. Suppose that X may be partitioned into a union of pairwise disjoint events $x_1,\ldots,x_n$ such that $\bigcup_{i=1}^{n} x_i = X$ and $P(x_i) > 0$ $\forall$ i=1,...,n. Let A and B be events in X with $P(B) > 0$ and $P(B^c) > 0$. Suppose that one calculated and compared the conditional probabilities $p_i = P(A|B\cap x_i)$ and $p'_i = P(A|B^c\cap x_i)$ $\forall$ i=1,...,n. One might expect that if $p_i = p'_i$, i=1,...,n, then $P(A|B) = P(A|B^c)$, and that a unidirectional inequality between these quantities

$$
(1.1a)\qquad \begin{cases} p_i > p'_i & i = 1,\ldots,n \quad ; \quad \text{or} \\ p_i \geq p'_i & i = 1,\ldots,n \quad ; \quad \text{or} \\ p_i < p'_i & i = 1,\ldots,n \quad ; \quad \text{or} \\ p_i \leq p'_i & i = 1,\ldots,n \quad ; \quad \text{or} \\ p_i = p'_i & i = 1,\ldots,n \quad ; \end{cases}
$$

would yield the corresponding inequality (or equality) between $P(A|B)$ and $P(A|B^c)$. The cases in which this expectation is unfulfilled go by the collective name "Simpson's paradox." The following examples illustrate the occurrence of this paradox in a diversity of situations:

Example 1. In an attempt to encourage students not to rely on intuition, a teacher makes the following presentation. On each of two tables there is a red and a green box. On the first table, he explains, the red box contains five blue marbles and six white marbles. The green box contains three blue marbles and four white marbles. The teacher chooses a student and asks the following question. "If you are allowed to randomly choose one marble from either box, which box should you draw from to maximize your chance of obtaining a blue marble?" The student correctly reasons that her chance of obtaining a blue marble from the red box is $\frac{5}{11}$, or approximately .45, while her chance of drawing a blue marble from the green box is $\frac{3}{7}$, or approximately .43, so she chooses the red box.

Continuing in this manner, the teacher explains that on the second table, the red box contains six blue marbles and three white marbles, while the green box contains nine blue and five white marbles. The teacher then poses the same question as before. The student reasons that her chance of drawing a blue marble from the red box is $\frac{6}{9}$, or approximately .67, while her chance of drawing a blue marble from the green box is $\frac{9}{14}$, or approximately .64, and so once again she chooses the red box.

"You made the correct choice in each case," said the teacher, who then proceeded to combine the contents from the two red boxes into a single red box. She also combined the two green boxes into a single green box, and asked, "Now, if you wish to obtain a blue marble, and you can randomly choose one marble from either box, which box

should you choose?" The student reasons that the correct choice of the red box in both previous cases indicates the red box should be chosen again, and she does so. "Your choice is incorrect, and had you checked the figures you would have realized it," said the teacher. Indeed, the green box would have been the correct choice, since the chance of drawing a blue marble from the red box is $\frac{11}{20}$, or .55, while the chance of obtaining a blue marble from the green box is $\frac{12}{21}$, or approximately .57.

Example 2. A "fantasy league" baseball game is one in which players form teams and base results on actual statistics from professional ball players. Two participants, A and B, draft players at the start of a season, each attempting to create the team with better statistics. At seasons end, they compare the statistics that their respective players have accumulated. Player A declares that he had the better hitting team, since his players had a higher batting average than B's players at each position. B argues that he clearly has the better hitting team, since his team has a higher batting average (see Table 1).

Example 3. Two medical researchers conduct a joint research project to determine trends in the incidence of tuberculosis among different racial groups. They acquire TB data from city A and city B. Each of these cities has an estimated population of 400,000. The first researcher bases his results on data separated by cities (see Table 2; all numbers given in thousands). The TB rate in city A

Table 1. "Fantasy League" Baseball Teams' Batting Averages.

Player A					Player B				
POS	Player	H	AB	BA	POS	Player	H	AB	BA
OF	Thompson	55	182	.302	OF	Murphy	185	616	.300
OF	Strawberry	109	393	.278	OF	Washington	110	398	.276
OF	Raines	184	575	.320	OF	Gwynn	197	622	.317
18	Hernandez	183	593	.309	18	Lopes	78	275	.281
28	Sandberg	186	609	.305	28	Herr	180	596	.302
55	Zuvella	48	190	.253	55	Ramirez	141	568	.248
38	Oberkfell	112	412	.272	38	Horner	129	483	.267
C	Cerone	61	282	.216	C	Benedict	42	208	.202
DH	Ford	6	12	.500	DH	McGee	216	612	.302
	Composite	944	3,248	.291		Composite	1,278	4,378	.292

Table 2. Incidence of Tuberculosis Among Different Racial Groups.

City A*			City B*		
	Healthy	Diseased		Healthy	Diseased
Black	313.6	6.4	Black	79.2	.2
White	78.4	1.6	White	316.8	3.2

*All figures given in thousands.

is 2% for both blacks and whites, while the rate is 1% in city B for each racial group. Based on these statistics, the first researcher concludes that while location seems to have an effect on TB frequency, race and TB susceptibility appear to be independent.

The second researcher aggregates the data into a single table without regard to location and draws a very different conclusion (see Table 3). These data indicate that 1.8% of blacks contract TB, while only 1.2% of whites do so, and the second researcher concludes that since blacks are 50% more likely to contract TB, race and TB susceptibility are not independent.

Table 3. Aggregate Tuberculosis Data.

	Healthy	Diseased
Black	392.8	7.2
White	395.2	4.8

These examples present hypothetical situations in which Simpson's paradox occurs. In fact, the paradox is known to actually occur in "real life" situations. This paper will investigate this phenomenon, starting with some preliminary observations, specify conditions for blocking the paradox, and examine philosophical issues raised by this paradox.

CHAPTER II

PRELIMINARY RESULTS

2.1 Subjective probability. Listen to a radio or television broadcast and you might hear a statement like, " . . . the prosecution team believes that they have a strong case, and say they feel they have about a 75% chance of conviction . . . ." This 75% chance that the attorneys speak of is quite different from the 75% chance we would say we have of obtaining at least one head from the toss of two unbiased coins. The coin toss model is a precise mathematical construct which is quite easy to justify and is subject to empirical verification. The attorneys' 75% may take into account many unquantifiable factors such as reliability of witnesses, attitudes of the judge, and mental or physical makeup of the jurors. Furthermore, empirical verification is impossible since even the simplest cases can not be repeated precisely.

We refer to this use of probability as a measure of personal belief as subjective probability. Despite the ambiguities and problems the use of subjective probability presents, we would be wrong to de-emphasize the importance of attempting to extend probability models to everyday situations. Oftentimes in elementary probability courses, this subject is given only a cursory treatment, or avoided altogether in favor of the less troublesome objective approach.

As a simple illustration of how the use of subjective probabilities can be useful in applications, however, consider the following example:

Example 4. A city council is set to take a vote to decide which of three construction companies will receive a contract to build a new city office building. Company A submits a bid of 630 thousand; company B submits a bid of 645 thousand; and company C submits a bid of 670 thousand. The decision is not based entirely on low bid, but is also to include consideration of quality of workmanship, estimated completion times, and companies' reputations.

When legal problems force a delay in the vote, the city manager is forced to plan a fiscal budget without foreknowledge of the correct amount of this considerable expenditure. After consulting his aides, he decides that the probability of acceptance of A, B, or C's bid is .50, .30, or .20 respectively. Using basic principles of probability, he concludes that the expected amount of the accepted bid will be ($630,000 $\cdot$.50 + $645,000 $\cdot$.30 + $670,000 $\cdot$.20) = $642,500, and so he includes this figure in his budget.

Throughout this paper, we will consider subjective as well as objective interpretations of the probabilistic results at issue.

2.2 Reducing the sample space. Computing probabilities often requires large-scale combinatorial problems used to measure the size of the sample space. Some problems require computational work of prohibitive proportions, and it is desirable to work with a reduced

sample space to make the problem more manageable. It is sometimes possible to achieve this reduction, if some partial information is given, by using conditional probabilities. For instance, there are 36 possible outcomes if we toss 2 unbiased dice of different colors, but if we are told after the toss that one die displays an even number, the number of possible outcomes is 18, or half the original number. In such a trivial example, this reduction is insignificant, but for extremely large problems it can be a considerable saving of computational work.

2.3 Probablistic independence. We can use conditional probabilities to solve problems in two ways. First, we may use the given condition to revise the sample space as mentioned in section 2.2, and then compute probabilities based on the new space. Second, we may use the formal definition of conditional probability, that is, if $P(F) > 0$, then

$$(2.3a) \qquad P(E|F) = \frac{P(EF)}{P(F)}$$

where $P(E|F)$ is the probability of E given that F has occurred. (Throughout the remainder of this paper, $P(EF)$ is equivalent to $P(E \cap F)$.

In general, $P(E|F) \neq P(E)$, that is, prior knowledge of the occurrence of an event F will often have some effect on the probability of the occurrence of E. Examples are easily constructed to demonstrate

this. If $P(E|F) = P(E)$, prior knowledge of the occurrence of F has no effect on the probability of the occurrence of E, and we say that event E is independent of event F. By substituting $P(E)$ for $P(E|F)$ in (2.3a), we can derive a second definition for the independence of E from F; that is, suppose $P(E) > 0$ and $P(F) > 0$, then E is independent of F iff:

$$(2.3b) \qquad P(E) = \frac{P(EF)}{P(F)}$$

or equivalently,

$$(2.3c) \qquad P(E)P(F) = P(EF).$$

Since (2.3c) is clearly equivalent to $P(F) = P(EF)/P(E) = P(F|E)$, we see that E is independent of F iff F is independent of E, so that we may speak of the independence of the events E and F. Such independence will be denoted by $E \perp F$.

It is easy to establish relationships between independence of events and independence of the complements of those events.

Proposition 1. $E \perp F$ is equivalent to $E \perp F^c$.

Proof. Suppose $E \perp F$. We can write $E = EF \cup EF^c$, so

$$\begin{aligned} P(E) &= P(EF \cup EF^c) \\ &= P(EF) + P(EF^c) - P(EF \cap EF^c) \end{aligned}$$

but $EF \cap EF^C = \emptyset$, so $P(EF \cap EF^C) = 0$.

Therefore,

$$P(E) = P(EF) + P(EF^C)$$
$$= P(E)P(F) + P(EF^C) \qquad \text{(by } E \perp F\text{)}.$$

Rearranging terms, we obtain

$$P(EF^C) = P(E) - P(E)P(F)$$
$$= P(E)(1 - P(F))$$
$$= P(E)P(F^C), \text{ or equivalently}$$

$E \perp F^C$.

Proposition 2. $E \perp F$ is equivalent to $E^C \perp F$.

Proof. Proof is immediate from proposition 1 and the symmetry of independence.

Proposition 3. $E \perp F$ is equivalent to $E^C \perp F^C$.

Proof. Immediate from symmetry of independence and repeated application of Proposition 1.

The foregoing observations suggest a reformulation of independence as a relation between the partitions $[E,E^C]$ and $[F,F^C]$. Suppose, for example, E = the set of all females in some population and F = the set of individuals with incomes over \$20,000, and $E \perp F$. This is often expressed by saying sex and income are independent, since it follows from $E \perp F$, that $E \perp F^C$, $E^C \perp F$, and $E^C \perp F^C$.

2.4 Relevance and irrelevance. We may easily extend the notion of independence to subjective probability. If the addition of information has no effect on the probability of an event, we will say that this information is irrelevant to our hypothesis. Added information that is not irrelevant is said to be relevant. Suppose a defendant is being tried in a car theft case, and the prosecuting attorney estimates the probability of the suspect's guilt to be approximately .25. If the attorney discovers that the suspect had a set of keys in his possession upon his arrest, he will more than likely dismiss the evidence as irrelevant, and so disregard it. If, upon further investigation, it is later learned that the keys fit the victim's car, the subjective probability of guilt will be revised upward. Such information is considered relevant.

Irrelevant information has no effect on the probability of a given event. Relevant information may lead to either an upward or downward revision of the probability of a given hypothesis. Any condition B that increases the probability of occurrence of an event A is said to be positively relevant to A, and can be written $P(A|B) > P(A)$. If the condition B decreases the probability of occurrence of the event A, we say B is negatively relevant to A, and write $P(A|B) < P(A)$. Irrelevance, positive relevance, and negative relevance may be expressed in many equivalent ways, some of which are presented in Table 4. We will show the equivalence of the foregoing ways of expressing positive relevance. The proofs for irrelevance and negative relevance are similar.

Table 4. Equivalent Expressions for Irrelevance, Positive Relevance, and Negative Relevance.

Irrelevance	Positive Relevance	Negative Relevance
$P(A\|B) = P(A)$	$P(A\|B) > P(A)$	$P(A\|B) < P(A)$
$P(AB) = P(A)P(B)$	$P(AB) > P(A)P(B)$	$P(AB) < P(A)P(B)$
$P(A\|B) = P(A\|B^C)$	$P(A\|B) > P(A\|B^C)$	$P(A\|B) < P(A\|B^C)$
$P(B\|A) = P(B)$	$P(B\|A) > P(B)$	$P(B\|A) < P(B)$
$P(B\|A) = P(B\|A^C)$	$P(B\|A) > P(B\|A^C)$	$P(B\|A) < P(B\|A^C)$

Proposition 4. If $P(A)$, $P(A^C)$, $P(B)$, and $P(B^C) > 0$, then the following are equivalent:

a) $P(A|B) > P(A)$

b) $P(AB) > P(A)P(B)$

c) $P(A|B) > P(A|B^C)$

d) $P(B|A) > P(B)$

e) $P(B|A) > P(B|A^C)$.

Proof. i). a) <=> d).

$$P(A|B) > P(A)$$

$$<=> \frac{P(BA)}{P(B)} > P(A)$$

(2.4a)
$$<=> P(BA) > P(A)P(B)$$

$$<=> \frac{P(BA)}{P(A)} > P(B)$$

$$<=> P(B|A) > P(B)$$

ii). a) <=> c).

As in (2.4a), $P(A|B) > P(A)$

$\iff P(AB) > P(A)P(B)$

$\iff P(AB) - P(B)P(AB) > P(A)P(B) - P(B)P(AB)$

$= P(B)(P(A) - P(AB))$

$= P(B)P(AB^C)$

$\iff P(AB)(1 - P(B)) > P(B)P(AB^C)$

$\iff P(AB)P(B^C) > P(B)P(AB^C)$

$\iff \frac{P(AB)}{P(B)} > \frac{P(AB^C)}{P(B^C)}$

$\iff P(A|B) > P(A|B^C)$

iii).a).<=>b).

Proof is immediate from (2.4a).

iv).a) = e. By (iii), $P(A|B) > P(A)$

$\iff P(AB) > P(A)P(B)$

$\iff P(AB) - P(A)P(AB) > P(A)P(B) - P(A)P(AB)$

$\iff P(AB)(1 - P(A)) > P(A)(P(B) - P(AB))$

$\iff P(AB)P(A^C) > P(A)P(A^CB)$

$\iff \frac{P(AB)}{P(A)} > \frac{P(A^CB)}{P(A^C)}$

$\iff P(B|A) > P(B|A^C)$

2.5 Properties of relevance. Falk and Bar-Hillel [3] introduce a simple device to denote probablistic dependence. If B is positively relevant to A, we write $B \nearrow A$, and say that B supports A. If B is negatively relevant to A, we write $B \searrow A$, and say B weakens A. As

before, if A and B are independent, we write $A \perp B$. We will refer to $\nearrow$, $\searrow$, and $\perp$ as dependence relationships. The following proposition regarding the symmetry of the dependence relationships is a reformulation of parts of proposition 4 in terms of $\nearrow$, $\searrow$, and $\perp$.

Proposition 5. The dependence relationships are symmetric, i.e.,

i) $A \nearrow B <=> B \nearrow A$

ii) $A \searrow B <=> B \searrow A$

iii) $A \perp B <=> B \perp A$.

While we see that the dependence relationships are symmetric, they fail to fulfill other intuitive expectations. For instance, while $\nearrow$ is reflexive on the set of events with probabilities strictly between 0 and 1, $\searrow$ and $\perp$ fail to be so, and none of the relations is transitive.

Proposition 6. For all events of probability < 1, the supportive dependence relationship is reflexive, i.e., $A \nearrow A$.

Proof. Given $P(A) < 1$, for any event A,

$P(A|A) = 1 > P(A)$

$<=> A \nearrow A$.

Proposition 7. i) For all events of nonzero probability, $A \searrow A$ is false.

ii) For all events with nonzero probability less than 1, $A \perp A$ is false.

Proof. i). If $P(A) \neq 0$, $P(A|A) = 1 \geq P(A)$, and so it is not the case that $P(A|A) < P(A)$.

ii). Suppose $0 < P(A) < 1$, then $P(A|A) = 1 > P(A)$, and so it is not the case that $P(A|A) = P(A)$.

Proposition 8. The dependence relationships $\perp$, $\not\perp$, and $\searrow$ are not transitive, i.e., the following statements are not true in general.

i). $A \perp B$ and $B \perp C \Rightarrow A \perp C$

ii). $A \searrow B$ and $B \searrow C \Rightarrow A \searrow C$

iii). $A \not\perp B$ and $B \not\perp C \Rightarrow A \not\perp C$

Proof. 1). A counter example to (i) may be constructed by choosing any A with $0 < P(A) < 1$ and any B such that $A \perp B$. If (i) were true in general it would follow that $A \perp A$, which is not the case, since $0 < P(A) < 1$.

ii). Suppose we have an unbiased die, with events A, B, and C defined as follows:

A - a roll of 6

B - an even number

C - a roll of 6. Then

$$P(A|B) = \frac{1}{3} > P(A) = \frac{1}{6} \Rightarrow A \searrow B,$$

$$P(C|B) = \frac{1}{3} > P(C) = \frac{1}{6} \Rightarrow C \searrow\!\nearrow B, \text{ or}$$

equivalently $B \searrow\!\nearrow C$, but

$$P(A|C) = 1 < P(C) = \frac{1}{6} \Rightarrow A \nearrow C.$$

iii). Counterexample. Suppose we have two unbiased dice, one white and the other yellow, with events A, B, and C defined as follows:

A - a roll of 6

B - an even number

C - a roll of 6. Then

$$P(A|B) = \frac{1}{3} > P(A) = \frac{5}{36} \Rightarrow A \nearrow B$$
$$P(B|C) = \frac{1}{3} > P(B) = \frac{1}{6} \Rightarrow B \nearrow C$$
$$P(A|C) = 0 < P(A) = \frac{5}{36} \Rightarrow A \searrow\!\nearrow C.$$

There is a simple relationship between positive and negative relevance, as seen in Proposition 9.

Proposition 9. $A \nearrow B \iff A \searrow\!\nearrow B^C$

Proof. $A \nearrow B$

$$\iff P(B|A) > P(B)$$
$$\iff \frac{P(AB)}{P(A)} > P(B)$$
$$\iff P(A)P(B) < P(AB) = P(A) - P(AB^C)$$
$$\iff P(A) - P(AB^C) > P(A)(1 - P(B^C))$$
$$\iff P(A) - P(AB^C) > P(A) - P(A)P(B^C)$$

$$<=> \quad P(A)P(B^C) > P(AB^C)$$

$$<=> \quad P(B^C) > \frac{P(AB^C)}{P(A)}$$

$$<=> \quad P(B^C) > P(B^C|A)$$

$$<=> \quad A \searrow B^C.$$

Given two events A and B that support a third event C, it does not necessarily follow that the intersection or union of A and B do so, as will be shown below.

Proposition 10. If $A \nearrow C$ and $B \nearrow C$, it does not generally follow that $AB \nearrow C$.

Proof. (Counterexample). Suppose we have an unbiased pair of dice, one blue and one green. Let the events A, B, and C be defined as follows:

A - a 2 on the blue die

B - a 2 on the green die

C - a total of 8, then

$A \nearrow C$ since $P(C|A) = \frac{1}{6} > \frac{5}{36} = P(C)$

$B \nearrow C$ since $P(C|B) = \frac{1}{6} > \frac{5}{36} = P(C)$, but

$AB \searrow C$ since $P(C|AB) = 0 < \frac{5}{36} = P(C)$.

The proof of proposition 11 will require the following lemma.

Lemma 1. $A \not\perp B$ is equivalent to $A^C \not\perp B^C$.

Proof. $A \not\perp B \iff A \not\perp B^C \iff B^C \not\perp A \iff B^C \not\perp A^C \iff A^C \not\perp B^C$.

Proposition 11. If $A \not\perp C$ and $B \not\perp C$, it does not generally follow that $A \cup B \not\perp C$.

Proof. If, for all A, B, and C, $A \not\perp C$ and $B \not\perp C$ implied $A \cup B \not\perp C$, it would follow that $A \not\perp C$ and $B \not\perp C$ implied $AB \not\perp C$, contrary to Proposition 10. For $A \not\perp C$ and $B \not\perp C$ yield $A^C \not\perp C^C$ and $B^C \not\perp C^C$ by Lemma 1. Hence if $A^C \cup B^C \not\perp C^C$, then again by Lemma 1, $(A^C \cup B^C)^C \not\perp C^C$, i.e., $AB \not\perp C$.

Proposition 12. If $C \not\perp A$ and $C \not\perp B$, it does not generally follow that $C \not\perp AB$.

Proof. Proof follows from Proposition 10 along with symmetry of $\not\perp$.

Proposition 13. If $C \not\perp A$ and $C \not\perp B$, it does not generally follow that $C \not\perp A \cup B$.

Proof. Proof follows from proposition 11 along with symmetry of $\not\perp$.

Analogues of Propositions 11-13 also hold for the dependence relationships $\not\searrow$ and $\perp$.

2.6 Bayes' Rule. Recall the derivation of Bayes' rule. Given occurrences E and F, E can be written as the disjoint union of EF and EF^C. Therefore,

(2.6a) $$P(E) = P(EF) + P(EF^C).$$

Now, from the definition of conditional probability, $P(E|F) = \frac{P(EF)}{P(F)}$, of equivalently $P(EF) = P(E|F)P(F)$, so we can make a substitution in (2.6a) to obtain

(2.6b) $$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C).$$

We can then obtain the following equation, known as Bayes' rule, to indicate how the probability of the occurrence of an event F is altered when new information is given.

(2.6c) Bayes' Rule

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)}$$

It is possible to use Bayes' rule in a subjective context. We may think of using it in a legal framework if we consider F to be a certain hypothesis, and let E be an item of new evidence. Then $P(F|E)$ gives an idea of how the introduction of this new evidence affects the subjective probability of the truth of F. If E is positively relevant to F, then the introduction of E strengthens the case. Evidence that is irrelevant is not introduced, and evidence that is negatively relevant is avoided, as it weakens the case.

These points are discussed in Lempert's Article [5] from the Michigan Law Review, in which he discusses attempts to incorporate techniques of mathematical decision theory into jurors' decision processes. He is pessimistic about the notion; in fact, he quotes from a paper by Tribe [9], " . . . the costs of attempting to integrate mathematics into the fact finding process of a legal trial outweigh the benefits." Lempert is convinced, however, that mathematical models may be used in legal situations as heuristic devices. He also suggests that laymen may be better able to deal with mathematical ideas in the context of an odds formulation rather than a probablistic one.

Suppose H is an hypothesis with $P(H) = \rho$. Then the odds of the occurrence of H, or $O(H) = \rho:(1-\rho)$. Probabilities with their corresponding odds are given in Table 5. Since these probabilities and odds are presented in a subjective context, we don't expect them to be exact, and so allow for minor inconsistencies. For instance, if P(H) = .83, we may say O(H) = 5:1 rather than .83:.17.

Table 5. Probabilities and Odds of Occurrence of H.

P(H)	O(H)
.20	1:4
.50	1:1
.67	2:1
.75	3:1

We may also introduce the concept of conditional odds. As might be expected

$$O(G|E) = \frac{P(G|E)}{P(G^C|E)}$$

is defined as the conditional odds of G given E.

We now may present an alternate formulation of Bayes' rule in the context of odds and conditional odds, as follows:

Proposition 14. If $P(E|G^C) \neq 0$, then

(2.6d) $$O(G|E) = \frac{P(E|G)}{P(E|G^C)} O(G).$$

Proof. $O(G|E)$

$$= \frac{P(G|E)}{P(G^C|E)}$$

$$= \frac{P(E)P(G|E)}{P(E)P(G^C|E)}$$

$$= \frac{P(EG)}{P(EG^C)}$$

$$= \frac{P(E|G)P(G)}{P(E|G^C)P(G^C)}$$

$$= \frac{P(E|G)}{P(E|G^C)} O(G).$$

$E \nearrow G$ if $O(G|E) > O(G)$, or equivalently if $P(E|G) > P(E|G^c)$, $E \searrow G$ if $P(E|G) < P(E|G^c)$, and $E \perp G$ if $P(E|G) = P(E|G^c)$. The ratio

$$\frac{P(E|G)}{P(E|G^c)} \tag{2.6e}$$

is called the likelihood ratio, and the value of the ratio gives the magnitude of the increase or decrease in $O(G)$. For instance, if $P(E|G) = .67$, then $P(E|G^c) = .33$. The likelihood ratio of $\frac{.67}{.33} \cong 2$ indicates that the odds of the occurrence of G are roughly doubled.

Evidence is considered irrelevant not only if the likelihood ratio equals one, but also if the ratio is "sufficiently" close to one. The degree to which proximity of the ratio to one is "sufficiently" close may differ from one situation to another.

2.7 Conditional Independence and Conditional Relevance. We can extend the notion of independence to conditional probabilities. We say the events A and B are conditionally independent given C, if given that C occurs, the probability of the occurrence of A is unchanged by the knowledge of the occurrence of B, that is A and B are conditionally independent given C if $P(A|BC) = P(A|C)$. Equivalent forms are given in Proposition 15.

Proposition 15. The following statements are equivalent:

i). $P(AB|C) = P(A|C)P(B|C)$

ii). $P(A|BC) = P(A|C)$

iii). $P(B|AC) = P(B|C)$.

Proof. We show i) <=> ii), the proof that i) <=> iii) being similar. By definition of conditional probability,

$$P(AB|C) = P(A|C)P(B|C)$$

$$\iff \frac{P(ABC)}{P(C)} = \frac{P(AC)}{P(C)}\,\frac{P(BC)}{P(C)}$$

$$\iff \frac{P(ABC)}{P(BC)} = \frac{P(AC)}{P(C)}$$

$$\iff P(A|BC) = P(A|C).$$

The independence of two events A and B does not insure their conditional independence given a third event C, as shown by the following example.

Example 5. Suppose an unbiased pair of dice, one blue and one green, are rolled. Let events A, B, and C be defined as follows:

A - a one on the blue die

B - a four on the green die

C - a total of seven

A is independent of B, since $P(A)P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(AB)$, but A and B are not independent given C since $P(AB|C) = 0 \neq \frac{1}{6} \cdot \frac{1}{6} = P(A|C)P(B|C)$.

In a similar manner, we can define conditional relevance. We say that B is positively relevant to A, given C if $P(A|BC) > P(A|C)$, and we say that B is negatively relevant to A, given C if

$P(A|BC) < P(A|C)$. Analogues to Proposition 15 exist for both conditional positive relevance and conditional negative relevance, and are given below. The proofs are similar to the proof of Proposition 15.

Proposition 16. The following statements are equivalent:

i) $P(AB|C) > P(A|C)P(B|C)$

ii) $P(A|BC) > P(A|C)$

iii) $P(B|AC) > P(B|C)$.

Proposition 17. The following statements are equivalent:

i) $P(AB|C) < P(A|C)P(B|C)$

ii) $P(A|BC) < P(A|C)$

iii) $P(B|AC) < P(B|C)$.

CHAPTER III

SIMPSON'S PARADOX

3.1 Blocking Simpson's paradox. Recall Simpson's paradox from section 1.1. A simple case occurs if X is partitioned into just two subsets, C and C^C. In this case, one example of the paradox is as follows. It is possible to have both

$$(3.1a) \qquad \begin{cases} P(A|BC) \geq P(A|B^CC) \text{ and} \\ P(A|BC^C) \geq P(A|B^CC^C) \end{cases}$$

and at the same time have

$$(3.1b) \qquad P(A|B) < P(A|B^C).$$

For the simple 2x2x2 case, we can represent this in tabular form as in Figure 1, where D, E, F, G, d, e, f, and g represent the sizes of corresponding classes of the partition. In these terms, for example, the first inequality from (3.1a) may be rewritten

$$P(A|BC) \geq P(A|B^CC)$$

$$\iff \frac{D}{D+F} \geq \frac{E}{E+G}$$

$$\iff DG \geq EF.$$

	B	
	C	C^C
A	D	E
A^C	F	G

	B^C	
	C	C^C
A	d	e
A^C	f	g

	B + B^C	
	C	C^C
A	D+d	E+e
A^C	F+f	G+g

Figure 1. Tabular Representation of Simpson's Paradox.

Restated in this form, the present example of Simpson's paradoz for the 2x2x2 case is said to occur if

$$\begin{matrix} DG \geq EF, & \text{and} \\ dg \geq ef, & \text{but} \end{matrix} \tag{3.1c}$$

$$Dg + Dg + dG + dg < EF + Ef + eF + ef. \tag{3.1d}$$

This is similar to the original presentation that is given by Simpson [8]. When written in this manner, sufficient conditions for blocking the paradox are obvious.

Proposition 18. If $DG \geq EF$ and $dg \geq ef$, then Simpson's paradox as written in (3.1c) and (3.1d) is blocked provided $dG + Dg \geq EF + eF$.

Proof. If $DG \geq EF$ and $eg \geq ef$ and $dG + Dg \geq Ef + eF$, then summing these inequalities gives $DG + dg + Dg + dG \geq EF + Ef + eF + ef$, so that Simpson's paradox is blocked.

The above theorem, however, is not particularly enlightening, since the stated blocking condition is just the negation of (3.1d), given (3.1c).

We now turn to more substantial and informative blocking conditions for the more general formulation where $X = C_1 \cap C_2 \cap \ldots \cap C_n$, $C_i \cap C_j = \emptyset \quad \forall\ i \neq j$. Three sufficient conditions can be given for blocking Simpson's paradox. These conditions will be interpreted for example 3 from this paper, which dealt with the incidence of tuberculosis in blacks and whites from two cities. First, however, some preliminary observations must be established.

Lemma 3a. $P(E_1E_2\ldots E_n)$

$$= P(E_n|E_1E_2\ldots E_{n+1})P(E_{n-1}|E_1E_2\ldots E_{n-2})\ldots P(E_2|E_1)P(E_1).$$

Proof. By the definition of conditional probability, the right side of the equation may be written

$$\frac{P(E_1E_2\ldots E_n)}{P(E_1E_2\ldots E_{n-1})} \cdot \frac{P(E_1E_2\ldots E_{n-1})}{P(E_1E_2\ldots E_{n-2})} \cdots \frac{P(E_1E_2)}{P(E_1)} P(E_1),$$

which, by successive cancellation equals $P(E_1E_2\ldots E_n)$.

The conditional version of Lemma 3a, where n = 2, is as follows:

Lemma 3b. $P(AC|B) = P(C|B)P(A|BC)$.

Proof. As in the proof of Lemma 3, the right side of the equation may be written

$$\frac{P(BC)}{P(B)} \cdot \frac{P(ABC)}{P(BC)} = \frac{P(ABC)}{P(B)} = P(AC|B).$$

We may define independence of two partitions of a probability space as follows: Let $[A_1,A_2,\ldots,A_n]$ and $[B_1,B_2,\ldots,B_m]$ be partitions of a probability space, with $n \geq 2$ and $k \geq 2$. We write $[A_1,A_2,\ldots,A_n] \perp [B_1,B_2,\ldots,B_m]$ iff $A_i \perp B_j$ $\forall$ i = 1,2,...,n, and j = 1,2,...,m.

Using these results, we can proceed to establish sufficient conditions for blocking Simpson's paradox in the general case.

Proposition 19. If the partitions $[B,B^C]$ and $[C_1,C_2,\ldots,C_n]$ are independent, and, if $P(A|BC_i) \rho P(A|B^CC_i)$ $\forall$ i = 1,...,n, (ρ is $\leq$, $<$, $\geq$, $>$, or =), then $P(A|B) \rho P(A|B^C)$, i.e., Simpson's paradox is blocked.

Proof. Suppose the partitions $[B,B^C]$ and $[C ,\ldots,C_n]$ are independent and that $P(A|BCi) \rho P(A|B^CCi)$ $\forall$ i = 1,2,...,n. Then

$$P(A|B) = \sum_{i=1}^{n} P(AC_i|B)$$

$$P(A|B) = \sum_{i=1}^{n} P(C_i|B)P(A|BC_i) \qquad \text{by Lemma 3b}$$

$$= \sum_{i=1}^{n} P(C_i)P(A|BC_i) \qquad \text{by independence of } [B,B^C] \text{ and } [C_1,C_2,\ldots,Cn]$$

$$\rho \sum_{i=1}^{n} P(C_i)P(A|BC_i)$$

$$= \sum_{i=1}^{n} P(C_i|B^C)P(A|B^CC_i) \qquad \text{again, by independence of } [B,B^C] \text{ and } [C_1,C_2, \ldots,C_n] \text{ and Lemma 3b.}$$

$$= \sum_{i=1}^{n} P(AC_i|B^C)$$

$$= P(A|B^C).$$

We may apply Proposition 19 to Example 3 as follows: Let

C = city A,

C^C = city B,

A = tuberculosis,

A^C = healthy,

B = black, and

B^C = white.

Further, let ρ = >. Then, $[C,C^C] \perp [B,B^C]$ is equivalent to saying race and habitat are independent. In fact, for Example 3, this is not the case, since city A's population is 80% black, while city B's population is 80% white. Since inhabitants of city A are twice as likely to contract TB as are inhabitants of city B, it seems only natural that a higher incidence of the disease should show up in blacks overall.

Before establishing the next condition for blocking Simpson's paradox, we need the following Lemma.

Lemma 4. $P(A|BC) > P(A|B^cC) \iff P(A|BC) > P(A|C)$.

Proof. $P(A|BC) > P(A|C)$

$\iff A \nearrow B$, given C

$\iff A \searrow B^c$, given C

$\iff P(A|C) > P(A|B^cC)$

$\iff P(A|BC) > P(A|B^cC)$.

Proposition 20. If the partitions $[A,A^c]$ and $[C_1,C_2,\ldots C_n]$ are independent, and if $P(A|BC_i)\ \rho\ P(A|B^cC_i)\ \forall i = 1,\ldots,n$, ($\rho$ is $\leq$, $<$,$\geq$,$>$, or $=$)., then $P(A|B)\ \rho\ P(A|B^c)$, i.e., Simpson's paradox is blocked.

Proof. Suppose $P(A|BC_i)\ \rho\ P(A|B^cC_i)$, then by Lemma 4, $P(A|BC_i)\ \rho\ P(A|C_i)$, but by independence of $[A,A^c]$ and $[C_1,\ldots,C_n]$, $P(A|C_i) = P(A)$, so $P(A|BC_i)\ \rho\ P(A)\ \forall i = 1, \ldots n$. Then

$$P(A|B) = \sum_{i=1}^{n} P(AC_i|B)$$

$$= \sum_{i=1}^{n} P(C_i|B)P(A|BC_i) \qquad \text{by Lemma 3b}$$

$$\rho \sum_{i=1}^{n} P(C_i|B)P(A)$$

$$= P(A) \sum P(C_i|B)$$

$$= P(A) \cdot 1 = P(A).$$

Using the same model for application to Example 3 as before, we see that $[C,C^C] \perp [A,A^C]$ says that incidence of TB and habitat are independent. In fact, this is not the case, since people of either race in city A are twice as likely to contract TB as are people living in city B. Since blacks constitute 80% of the population of city A, it seems natural that a higher incidence of TB contraction is evident in them.

It is possible to apply Propositions 19 and 20 to Example 2, which dealt with baseball batting averages. Let

X = the total number of acts of going to bat, with the outcomes of all players on all teams,

C_i = roster position i, $\forall i = 1,\ldots,n$, where n is the number of ordered roster positions on a team,

B = team B,

B^C = team A,

A = hit,

A^C = no hit, and

ρ = >. Then the occurrence of Simpson's paradox in Example 2 is blocked if either $[B,B^C] \perp [C_1,C_2,\ldots,C_n]$, that is, if at bats per team is proportional to at bats per roster position, or $[A,A^C] \perp [C_1,C_2,\ldots,C_n]$, that is, if hits per at bats on the aggregate is proportional to hits per at bats across the partition of different roster positions.

Propositions 19 and 20 also hold with the inequalities reversed.

Proposition 21a. If $\min_i \{P(A|BC_i)\} > \max_i \{P(A|B^cC_i)\}$, and if $P(A|BC_i) > P(A|B^cC_i)$ $\forall i = 1,\ldots,n$, then $P(A|B) > P(A|B^c)$, i.e., Simpson's paradox is blocked.

Proof. Let $\min_i \{P(A|BC_i)\} = \mu$ and

$\max_i \{P(A|B^cC_i)\} = \mu^*$, then by assumption $\mu > \mu^*$.

Then,

$$P(A|B) = \sum_{i=1}^{n} P(AC_i|B)$$

$$= \sum_{i=1}^{n} P(C_i|B)P(A|BC_i) \qquad \text{by Lemma 3b}$$

$$\geq \sum_{i=1}^{n} P(C_i|B)\mu$$

$$= \mu \sum_{i=1}^{n} P(C_i|B) = \mu \cdot 1 = \mu. \quad \text{Also}$$

$$P(A|B^c) = \sum_{i=1}^{n} P(AC_i|B^c)$$

$$= \sum_{i=1}^{n} P(C_i|B^c)P(A|B^cC_i) \qquad \text{by Lemma 3b}$$

$$\leq \sum_{i=1}^{n} P(C_i|B^c)\ \mu^*$$

$$= \mu^* \sum_{i=1}^{n} P(C_i|B^c) = \mu^* \cdot 1 = \mu^*. \quad \text{So}$$

$P(A|B) \geq \mu > \mu^* \geq P(A|B^c)$, and hence $P(A|B) \geq P(A|B^c)$.

Proposition 21b. If $\min_i \{P(A|BC_i)\} > \max_i \{P(A|B^cC_i)\}$, and if $P(A|BC_i) \geq P(A|B^cC_i)$ $\forall i = 1,\ldots,n$, then $P(A|B) \geq P(A|B^c)$, i.e., Simpson's paradox is blocked.

Proof. The proof is similar to the proof of Proposition 21a.

Clearly, Proposition 21a fails to hold in Example 3, since .01 = P (tuberculosis/black in city B) = $\mu < \mu^*$ = P (tuberculosis/white in city A) = .02. Rephrasing Proposition 21a, Simpson's paradox would be blocked in this case if the lowest black TB rate in any city were higher than the highest white TB rate in any city.

Propositions 21a and 21b also hold if the inequalities are reversed and the words min and max are interchanged.

3.2 The Berkeley admissions case. One of the most well-known cases of the occurrence of Simpson's paradox involved admissions rates to graduate programs at the University of California at Berkeley. A chi-square test on aggregate data indicated that there were 277 more men accepted than would be expected if acceptance were independent of sex. This discovery in 1973 was particularly troublesome in light of the increased awareness at the time of sex discrimination. Nickel, Hammel, and O'Connell [1] set out to discover the basis of this apparent bias, and came up with some counterintuitive results. At the beginning of their study, they made two simple assumptions:

i) males and females are "equivalently" qualified from field to field, and

ii) sex ratios of applicants to fields are not correlated with any other admission factors.

The researchers investigated 85 individual departments, eliminating 16 departments which either had no female applicants, or accepted all female applicants who applied. Of the 85 remaining departments, there were only 4 departments in which the indicated bias would occur by chance less than 5% of the time. These 4 accounted for a shortfall of 26 females. Surprisingly, the corresponding opposite bias accounted for a shortfall of 64 males. So when the individual departments are investigated, there actually appeared to be a slight bias against men. While there were 277 too few females accepted overall, not a single individual department could be forced to accept responsibility.

The paper suggests several possible reasons that the paradox would occur, including:

i) Some departments are much harder to enter than others for either sex, in fact, "The proportion of women applicants tends to be high in departments that are hard to get into and low in those that are easy to get into."

ii) The first assumption was incorrect. Men and women are not equally qualified for various reasons, regardless of how desirable that situation may be.

The writers conclude that in order to determine the extent of bias, we must be careful to base our results on both aggregate data and individual department data.

While Bickel, Hamel, and O'Connell offer means of interpreting occurrences of Simpson's paradox, we can use Propositions 17-19 to establish conditions which would prevent the paradox form happening. This possibility raises philosophical questions which are not easily answered. Guarantee of the global independence of sex and admission status would require either uniformizing the admission status across all departments, or changing the percentage of female applicants so that it equals the percentage of females accepted. Either of these alternatives requires substantial social modeling, and such attempts seem doomed to failure.

3.3 The logical status of statistical paradoxes.

We must be cautious when we attempt to discern the meaning of statistical paradoxes, such as Smipson's paradox, or the intransivity of the dependence relationships. These are not paradoxes in the same sense that strictly semantical paradoxes are. One example of such a paradox is the statement, "This sentence is false." Such a statement is considered a paradox because it is self-contradictory. Assumption of the statement's truth leads to the conclusion that it is false, and vice versa. This statement, and those like it, have no logical basis, and are sometimes dismissed by logicians as being not well grounded.

The statistical paradoxes are something entirely different. They are arrived at as the result of logical propositions. They are not contradictory, even though they seem to be so. They are paradoxes

in the sense that they go against expectation. It is possible to provide an explanation for the forces at work when the paradoxical occurrence is observed.

In the case of Simpson's paradox, where two seemingly contradictory results are indicated from studying the data in different ways, the truth is not that either one of the conclusions is incorrect, but that they provide alternate points of view of the same thing. It is with this understanding that seemingly contradictory results can both be declared to be truthful. Our mistake is one of misunderstanding the nature of the problem. As the examples in this paper serve to indicate, it is simply incorrect to indiscriminately draw conclusions based on guessing the results of combining partial information. Careful statistical analysis will always provide more reliable results than decisions based on intuition.

The following question may then be asked. If Simpson's paradox furnishes seemingly contradictory results, which of the two approaches is more correct, the one based on aggregate or segregated data? The answer is largely situational. If we reexamine Example 3, in which blacks on the whole contract TB 50% more than whites, a case can be made that an individual might prefer to be white, but blacks in city A are as unlikely as any person in the study to catch the disease, and less likely to do so than aggregated whites. Perhaps the best solution is to let each individual decide which interpretation of the data is most meaningful for him.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Bickel, B. J., Hammel, E. A., O'Connell, J. W., "Sex Bias in Graduate Admissions: Data from Berkeley," Science, 187 (February 1975), 398-404.

2. Blyth, C. R., "On Simpson's Paradox and the Sure-Thing Principle," Journal of the American Statistical Association, 67 (June 1972), 364-366.

3. Falk, Bar-Hillel, M., "Probablistic Dependence Between Events," Mathematics Magazine, 53 (1980), 106-107.

4. Gardner, M., "On the Fabric of Inductive Logic and Some Probability Paradoxes," Scientific American, 234 (1976), 119-122.

5. Lempert, R. O., "Modeling Relevance," Michigan Law Review, 75 (April 1977), 1021-1057.

6. Martin, E., "Simpson's Paradox Resolved: A Reply to Hintzman," Psychological Review, 88, No. (4), 372-374.

7. Ross, S., A First Course In Probability Theory, New York: Macmillan Publishing Company, 1984.

8. Smipson, E. H., "The Interpretation of Interaction in Contingency Tables," Journal of the Royal Statistical Society, Series B, 13, No. 2 (1951), 238-241.

9. Tribe, A. B., "Trial By Mathematics: Precision and Ritual in the Legal Process," Harvard Law Review, 84, No. 5 (1971), 1329-1332.

10. Wagner, C. H., "Simpson's Paradox in Real Life," American Statistician, 36, No. 1 (1982), 46-47.

VITA

David Eugene Vinson was born in Nashville, Tennessee, on July 16, 1958. He graduated from McGavock High School in Nashville in June, 1976. After spending one semester at Middle Tennessee State University, he entered The University of Tennessee, Knoxville in September, 1977, and in June 1982 received a Bachelor of Science degree in Mathematics. In September, 1982, he accepted a teaching assistantship at The University of Tennessee, Knoxville, and began study toward a Master's degree. This degree was awarded in December, 1985.

The author is a nominee member of the American Mathematical Society. Mr. Vinson plans to pursue employment opportunities in the field of college mathematics education.