ALLOCATIONS OF PROBABILITY: A Theory of Partial Belief

Glenn Ray Shafer

A DISSERTATION

PRESENTED TO THE

FACULTY OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY THE

DEPARTMENT OF

STATISTICS

June, 1973

PREFACE

This essay constitutes Part II of a proposed monograph devoted to the exposition and justification of some of A.P. Dempster's methods of statistical inference. Part I of that monograph exists in draft form and is devoted primarily to a historical and critical account of the Bayesian paradigm for statistical inference. Statistical inference is not taken up directly in the present essay, but the ideas discussed here are directly relevant to a justification of Dempster's methods, some details of which may be found in my essay "A Theory of Statistical Support."

The ideas expounded here are directly inspired by my study of Professor Dempster's work, a study that began when I attended his seminar at Harvard in the spring of 1971. The reader will note that the quantities $Bel(A), P^*(A)$ treated axiomatically here correspond to the quantities $P_*(A), P^*(A)$ derived by Dempster from multivalued mappings. Unfortunately, the exact relationship between the present axiomatization and Dempster's original formulation remains somewhat obscure to me. In particular, I do not know how to express the condition of condensability in terms of multivalued mappings, though the examples that most interested Dempster were condensable.

The present essay does not include a discussion of the theory of integration on probability algebras. Using that theory, one can easily extend to allocations the discussion of several topics that are usually treated for distributions of probability. These include measures of location and dispersion, as well as analogues to entropy. Interestingly enough, the concept of entropy, rather overworked for distributions, breaks into two distinct concepts for allocations. One of these is related to the degree of conflict present in the evidence, while the other is related to the precision and strength of the evidence.

Aside from my obvious debt to Professor Dempster, I am also indebted to my wife Terry and my many other friends, teachers and fellow students who have helped me with these ideas. These include Paul Benacerraf, Thomas Corwin, Robert Epp, Alan Gross, Ian Hacking, Richard Hamming, Richard Jeffrey, Simon Kochen, Rod Montgomery, Edward Nelson, Dana Scott, Gary Simon, John Tukey and Paul Velleman. Peter Bloomfield, Richard Holley and Hale Trotter have been especially generous with their time. And Geoffrey Watson, my supervisor, has provided much needed encouragement.

I must also express my gratitude to the National Science Foundation; I have been supported by one of their graduate fellowships during the past three years.

ABSTRACT

A function $\mathrm{Bel}: \mathcal{Q} \to [0,1]$ on a Boolean algebra of propositions $\mathcal{Q}$ is a belief function if

I. $\mathrm{Bel}(\Lambda_{\mathcal{Q}})=0$, where $\Lambda_{\mathcal{Q}}$ is the impossible proposition.

II. $\mathrm{Bel}(V_{\mathcal{Q}})=1$, where $V_{\mathcal{Q}}$ is the sure proposition.

III. $\mathrm{Bel}(A_1 \vee \cdots \vee A_n) \geq \sum_i \mathrm{Bel}(A_i) - \sum_{i<j} \mathrm{Bel}(A_i \wedge A_j) + - \cdots + (-1)^{n+1}\mathrm{Bel}(A_1 \wedge \cdots \wedge A_n)$ for all $A_1,\ldots,A_n \in \mathcal{Q}$.

The adoption of Bel means the adoption, for each $A \in \mathcal{Q}$, of the quantity Bel(A) as one's degree of belief in the proposition A. If these degrees of belief correspond to the degrees to which the evidence supports the various A, then the quantities $P^*(A)=1-\mathrm{Bel}(\bar{A})$ will correspond to the degrees to which the various A are plausible in light of the evidence.

Axioms I-III are satisfied by probability functions, but they are also satisfied by many functions that are not probability functions. In particular, they are always satisfied by the vacuous belief function, given by $\mathrm{Bel}(V_{\mathcal{Q}})=1$ and $\mathrm{Bel}(A)=0$ for all $A \neq V_{\mathcal{Q}}$.

A pair $(\mathcal{M}, \mu)$ is a probability algebra if $\mathcal{M}$ is a complete Boolean algebra and μ is a positive and completely additive measure with $\mu(V_{\mathcal{M}})=1$. A mapping ρ from a Boolean algebra of propositions $\mathcal{Q}$ into a probability algebra $\mathcal{M}$ is called an allocation of probability if

1. $\rho(\Lambda_{\mathcal{Q}})=\Lambda_{\mathcal{M}}$.
2. $\rho(V_{\mathcal{Q}})=V_{\mathcal{M}}$.
3. $\rho(A \wedge B)=\rho(A) \wedge \rho(B)$ for all $A,B \in \mathcal{Q}$.

As it turns out, $\mu\circ\rho$ is a belief function whenever ρ is an allocation of probability, and any belief function Bel: $\mathcal{A}\rightarrow[0,1]$ can be represented in the form $\mu\circ\rho$ for some allocation ρ on $\mathcal{A}$. Intuitively, the elements of $\mathcal{M}$ are probability masses, or portions of one's total belief, and $\rho(A)$ is that portion of one's belief which one commits to A. Hence the axioms for belief functions correspond to the idea that having a certain degree of belief in a proposition means committing that proportion of one's total belief to it.

An allocation ρ on a power set $\mathcal{P}(\mathcal{S})$ is <u>condensable</u> if $\rho(\bigcap\mathcal{C})=\bigwedge_{C\varepsilon\mathcal{C}}\rho(C)$ for all $\mathcal{C}\subset\mathcal{P}(\mathcal{S})$. This is equivalent to $P^*(A)=\sup\{P^*(B)\mid B\subset A;\ B \text{ is finite}\}$ for all $A\subset\mathcal{S}$. Condensability can be defended as a natural condition for belief functions that are derived from empirical evidence, and it plays an important role in the abstract theory.

A belief function Bel_0 on a subalgebra $\mathcal{A}_0$ of a Boolean algebra $\mathcal{A}$ naturally induces a belief function Bel on $\mathcal{A}$. And belief functions on independent subalgebras of $\mathcal{A}$ can be combined by a natural rule to produce a belief function on $\mathcal{A}$. Study of this rule leads one to distinguish between <u>orthogonality</u> and <u>cognitive independence</u> for two independent subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ with respect to a belief function Bel. Orthogonality means that $\mathrm{Bel}(A\wedge B)=\mathrm{Bel}(A)\mathrm{Bel}(B)$ whenever $A\,\varepsilon\,\mathcal{A}_1$ and $B\,\varepsilon\,\mathcal{A}_2$, while cognitive independence means that $P^*(A\wedge B)=P^*(A)P^*(B)$ whenever $A\,\varepsilon\,\mathcal{A}_1$ and $B\,\varepsilon\,\mathcal{A}_2$.

Dempster's rules of conditioning and combination are techniques for modifying a belief function Bel: $\mathcal{A}\rightarrow[0,1]$ on the basis of new evidence or opinion. The rule of conditioning tells how to modify Bel when one learns that a given proposition $A\,\varepsilon\,\mathcal{A}$ is true. The rule of combination tells how to combine Bel with a new belief function Bel': $\mathcal{A} \to [0,1]$ so that the resulting belief function corresponds to the total evidence--the evidence that would be obtained by pooling the evidence underlying Bel with that underlying Bel'. Both of these rules are most applicable and most easily expressed in the condensable case.

TABLE OF CONTENTS

CHAPTER 1. DEGREES OF BELIEF

This chapter adduces and defends a set of rules governing degrees of belief for propositions in a Boolean algebra $\mathcal{A}$.

Intuitively, a Boolean algebra of propositions $\mathcal{A}$ is simply a collection of propositions which includes the impossible proposition Λ, the sure proposition V, the negation $\bar{A}$ of any of its elements A, and the conjunction $A \wedge B$ and the disjunction $A \vee B$ of any pair A,B of its elements. One writes $A \leq B$ to indicate that A implies B, and one assumes that A=B whenever both $A \leq B$ and $B \leq A$. I will assume that the reader is familiar with the mathematical structure of Boolean algebras and with the rules governing the symbols $\leq, \Lambda, V, -, \bar{}, \wedge$, and $\vee$. If he is not he may wish to consult Chapter 3 below, or he may wish to rely on a simple analogy with the symbols $\subset, \emptyset, \mathcal{S}, \sim, \tilde{}, \cap$ and $\cup$, as they apply to subsets of a set $\mathcal{S}$. ($A \subset B$ means that A is contained in B, $\emptyset$ denotes the empty set, $A \sim B$ is the set of points of A that are not in B, $\tilde{A} = \mathcal{S} \sim A$, $A \cap B$ denotes the intersection of the subsets A and B, and $A \cup B$ denotes the union of the subsets A and B.)

1. Axioms for Degrees of Belief

In Part I of this essay, I discussed at length reasons why the axiom of additivity is not always appropriate as a rule for degrees of belief, and I concluded in particular that it is not appropriate for the problem of statistical support. Nonetheless, I find that I cannot ignore the tremendous intuitive attraction of the classical theory of epistemic probability, and I can understand why many people find this attraction more weighty than any abstract argument. This attraction appears to stem from an intuitive understanding we have of probabilities, which, though it is seldom made entirely explicit, gives many of the rules that the subjective probabilist associates with degrees of belief a compelling, almost self-evident quality. We have an intuitive picture of probabilities, and it is that picture, rather than the formal rule of additivity, that we find hardest to give up.

The axiom of additivity is not, however, the most fundamental part of this intuitive picture. There are other rules that we associate with probabilities as degrees of belief that seem to be more fundamental, and correspondingly more self-evident. A good example is the rule of monotonicity, which states that if one proposition implies a second proposition, then the second proposition deserves at least as great a degree of belief as the first. In this chapter, we will discover that many of these more fundamental rules, as well as the intuitive picture underlying them, can be preserved even though the rule of additivity is dropped.

Let us take a closer look at the rule of monotonicity, for example, and try to understand the intuitive picture that makes it so self-evident. Denoting the degree of belief in A by Bel(A), we can express that rule by saying that

$$\text{if } A \leq B, \text{ then } \mathrm{Bel}(A) \leq \mathrm{Bel}(B).$$

Two corollaries of this rule are that Λ, the impossible proposition, should have the lowest degree of belief, conventionally zero, while V, the sure proposition, should have the highest degree of belief, conventionally one. The rule itself is obviously more than a convention; it is somehow necessary, given our intuitive ideas about how degrees of belief should work.

How can we make these intuitive ideas more explict? One way to bring them out is to examine the intuitive arguments that we might use in support of the rule of monotonicity. The reader is invited to consider what sort of intuitive argument he might offer; I find myself saying something like this: "If A implies B, then whenever A is true, B is true. So whatever belief I associate with A's being true, I must also associate with B's being true; and hence the belief I associate with B will include the belief I associate with A. In other words, the portion of my belief committed to B will include the portion committed to A. And in particular, its measure will be greater."

The fundamental feature of the picture revealed by this argument is that our belief appears in it as a measurable substance, various portions of which are committed to various propositions. This is natural enough an idealization; it merely makes explicit the notion that the relation between a degree of belief and complete belief is like the relation between a part and a whole. A secondary aspect of the picture is a restriction on our freedom in committing portions of this belief to various propositions, namely, the requirement that a portion of belief committed to a given proposition must also be committed to any more inclusive proposition. A further restriction, of course, is that none of our belief may be committed to Λ, while all of it must be committed to V; I accordingly adopt the convention that the total measure of our belief is equal to one.

What other restrictions are natural to this intuitive picture? One that seems natural enough is the requirement that a given portion of belief should not be simultaneously committed to two incompatible propositions. This requirement leads to the rule of superadditivity, which states that the degree of belief in the disjunction of two incompatible propositions should be at least as great as the sum of the degrees of belief for the separate propositions. In symbols:

$$\text{if } A \wedge B = \Lambda, \text{ then } \mathrm{Bel}(A) + \mathrm{Bel}(B) \leq \mathrm{Bel}(A \vee B).$$

In order to justify this rule, one should note that $A \leq A \vee B$ and $B \leq A \vee B$, so that the belief committed to $A \vee B$ must include both the belief committed to A and the belief committed to B. And there can be no overlap; since A and B are incompatible, none of the belief committed to one of them can also be committed to the other. Hence the measure of the belief committed to $A \vee B$ must be at least as great as the sum of the measures of these two separate portions of belief.

Notice that there is nothing in our intuitive picture to require that the inequality in the rule of superadditivity be replaced by equality. Equality would hold, evidently, only if all the belief committed to $A \vee B$ were necessarily committed either to A or to B. This would be a very strong restriction compared with the two restrictions that we have just considered, and we will find that it is not necessary for a coherent theory of degrees of belief.

This is not to say, though, that no further restrictions are appropriate on our freedom to commit our idealized portions of belief to different propositions. One further restriction that seems unavoidable is the requirement that any portion of belief that is committed to both of two propositions should also be committed to their logical conjunction. This may seem like a tautology, but it has a great many consequences.

For a start, we can use it to deduce the rule

$$\mathrm{Bel}(A)+\mathrm{Bel}(B)-\mathrm{Bel}(A\wedge B) \leq \mathrm{Bel}(A\vee B)$$

for all pairs of propositions in the Boolean algebra for which one has degrees of belief. The argument for this rule again depends on the fact that the belief committed to $A\vee B$ must include at least all the belief committed either to A or to B or to both. For the left-hand side represents the measure of this latter belief, obtained by adding the measure of the belief committed to A to the measure of the belief committed to B and subtracting the measure of what is counted twice, namely the belief committed both to A and to B.

A similar inequality will hold for triplets of propositions A, B and C:

$$\mathrm{Bel}(A)+\mathrm{Bel}(B)+\mathrm{Bel}(C)-\mathrm{Bel}(A\wedge B)-\mathrm{Bel}(A\wedge C)-\mathrm{Bel}(B\wedge C)+\mathrm{Bel}(A\vee B\vee C) \leq \mathrm{Bel}(A\vee B\vee C).$$

Here the left-hand side is the measure of all the belief that is committed to at least one of A, B and C. To see that this is so, notice that the quantity $\mathrm{Bel}(A)+\mathrm{Bel}(B)+\mathrm{Bel}(C)$ overstates that measure, for that portion of belief that is committed to both of any two of the propositions is counted twice, while that committed to all three is counted three times. When one subtracts the quantity $\mathrm{Bel}(A\wedge B)+\mathrm{Bel}(A\wedge C)+\mathrm{Bel}(B\wedge C)$, one is subtracting exactly once the measure of the belief committed to exactly two of the propositions, as is appropriate, but one is subtracting three times the measure of the belief committed to all three, and this is once too often. Hence one must finally add $\mathrm{Bel}(A\wedge B\wedge C)$ back again.

A similar inequality can be obtained for any finite collection $A_1,\ldots,A_n$ of propositions by comparing the measure of the belief committed to $A_1 \vee \cdots \vee A_n$ with the measure of all the belief that is committed to at least one of the A_i. That inequality is

$$\sum \mathrm{Bel}(A_i) - \sum \mathrm{Bel}(A_i \wedge A_j) + \sum \mathrm{Bel}(A_i \wedge A_j \wedge A_k) - + \cdots + (-1)^{n+1} \mathrm{Bel}(A_1 \wedge \cdots \wedge A_n) \leq \mathrm{Bel}(A_1 \vee \cdots \vee A_n$$

As we will see, these inequalities, together with the conventions $\mathrm{Bel}(\Lambda)=0$ and $\mathrm{Bel}(V)=1$, provide a satisfactory basis for a general theory of degrees of belief. Hence I will use them for a formal definition.

Definition. A function Bel on a Boolean algebra is a belief function if it takes values between zero and one and satisfies the following three axioms:

(I). $\mathrm{Bel}(\Lambda)=0$.

(II). $\mathrm{Bel}(V)=1$.

(III). If $n \geq 1$ and $A_1, \ldots, A_n$ are elements of the Boolean algebra, then

$$\mathrm{Bel}(A_1 \vee \cdots \vee A_n) \geq \sum \mathrm{Bel}(A_i) - \sum \mathrm{Bel}(A_i \wedge A_j) + - \cdots + (-1)^{n+1} \mathrm{Bel}(A_1 \wedge \cdots \wedge A_n).$$

My claim that this definition provides a satisfactory basis for a general theory of degrees of belief will be supported in two different ways in the following pages. On the one hand, we will see that these axioms more or less exhaust the consequences of our intuitive picture of "portions of belief" and that that intuitive picture is at least as attractive as the more special one usually associated with subjective probabilities. On the other hand, we will see that the axioms are general enough to encompass many systems of degrees of belief that are attractive and useful but fail to satisfy Kolmogorov's axioms. The demonstration that these new axioms are equivalent to the intuitive picture of portions of belief can be left for the next chapter, but it appropriate to illustrate their generality immediately.

2. Four Examples of Belief Functions

In this section I will exhibit four simple examples of belief functions. For the first two examples, I will verify Axiom III directly. For the last two, though, I will leave such a verification until the next chapter, where it will be facilitated by a fuller understanding of the structure of belief functions.

A. The Vacuous Belief Function

The simplest belief function on any Boolean algebra of propositions is the one that assigns degree of belief zero to every proposition except the sure proposition, which must of course have degree of belief one. This belief function corresponds to a complete lack of opinion—one has too little evidence or is too skeptical to assign a positive degree of belief to any proposition in the Boolean algebra except the one that is logically certain. I will call it the vacuous belief function. Axioms I and II obviously hold for this belief function, but how can we establish Axiom III?

First note that if none of the propositions $A_1,\ldots,A_n$ are equal to $\mathcal{V}$, then the right-hand side of the inequality is zero, so that the inequality necessarily holds. Suppose, on the other hand, that some of the A_i are equal to $\mathcal{V}$—say k of them. Then $\binom{k}{2}$ of the propositions $A_i \wedge A_j$ will also be equal to $\mathcal{V}$, $\binom{k}{3}$ of the propositions $A_i \wedge A_j \wedge A_k$, etc. Hence we will have

$$\sum \mathrm{Bel}(A_i)=k=\binom{k}{1},$$

$$\sum \mathrm{Bel}(A_i \wedge A_j)=\binom{k}{2},$$

$$\mathrm{Bel}(A_i \wedge A_j \wedge A_k)=\binom{k}{3},$$

etc., so that the right-hand side of the inequality will be

$$\binom{k}{1}-\binom{k}{2}+\binom{k}{3}-\cdots+(-1)^{k+1}\binom{k}{k}=1-(1-1)^k=1,$$

But $A_1 \vee \cdots \vee A_n$ will also be equal to $\mathcal{V}$, so that the left-hand side, $\mathrm{Bel}(A_1 \vee \cdots \vee A_n)$, will also be equal to 1, and the inequality will hold with equality. Hence Axiom III does indeed hold for the vacuous belief function.

B. Belief Functions on a Four-Element Boolean Algebra

The vacuous belief function is the only possible one on the two-element Boolean algebra $\{\Lambda, \mathcal{V}\}$, but there are more possibilities for a four-element Boolean algebra $\{\Lambda, A, \bar{A}, \mathcal{V}\}$. Suppose, for example that A is the proposition that there is life of Mars. A belief function on this Boolean algebra would then summarize one's degrees with respect to that proposition, both for and against it. Some might profess a complete lack of opinion about the proposition and adopt the vacuous belief function, but others will have some degree of belief either for or against it, or both, even if those degrees of belief are rather weak. One might, for example, profess a degree of belief of 1/10 in A, a degree of belief of 2/10 in $\bar{A}$, and of course a degree of belief 1 in $A \vee \bar{A} = \mathcal{V}$. But will the function Bel with values $\mathrm{Bel}(\Lambda)=0$, $\mathrm{Bel}(A)=1/10$, $\mathrm{Bel}(\bar{A})=2/10$ and $\mathrm{Bel}(\mathcal{V})=1$ satisfy Axiom III?

It is not difficult to show that it does, as does any function $\mathrm{Bel}: \{\Lambda, A, \bar{A}, \mathcal{V}\} \rightarrow [0,1]$ that satisfies $\mathrm{Bel}(\Lambda)=0$, $\mathrm{Bel}(\mathcal{V})=1$ and $\mathrm{Bel}(A)+\mathrm{Bel}(\bar{A}) \leq 1$. Suppose, indeed, that $A_1, \ldots, A_n$ are all propositions from $\{\Lambda, A, \bar{A}, \mathcal{V}\}$. Then let a be the number of the n propositions that are equal to Λ, b the number that are equal to A, c the number that are equal to $\bar{A}$, and d the number equal to $\mathcal{V}$; $a+b+c+d=n$. Then

$$\sum \mathrm{Bel}(A_i) = \mathrm{Bel}(A)\binom{b}{1} + \mathrm{Bel}(\bar{A})\binom{c}{1} + \binom{d}{1},$$

$$\sum \mathrm{Bel}(A_i \wedge A_j) = \mathrm{Bel}(A)\left[\binom{b}{2}+\binom{b}{1}\binom{d}{1}\right] + \mathrm{Bel}(\overline{A})\left[\binom{c}{2}+\binom{c}{1}\binom{d}{1}\right] + \binom{d}{2},$$

$$\sum \mathrm{Bel}(A_i \wedge A_j \wedge A_k) = \mathrm{Bel}(A)\left[\binom{b}{3}+\binom{b}{2}\binom{d}{1}+\binom{b}{1}\binom{d}{2}\right] + \mathrm{Bel}(\overline{A})\left[\binom{c}{3}+\binom{c}{2}\binom{d}{1}+\binom{c}{1}\binom{d}{2}\right] + \binom{d}{3},$$

etc.; and

$$\sum \mathrm{Bel}(A_i) - \sum \mathrm{Bel}(A_i \wedge A_j) + - \cdots + (-1)^{n+1}\mathrm{Bel}(A_1 \wedge \cdots \wedge A_n)$$
$$= \mathrm{Bel}(A)\left[\binom{b}{1}-\binom{b}{2}+-\cdots+(-1)^{b+1}\binom{b}{b}\right]\left[\binom{d}{0}-\binom{d}{1}+-\cdots+(-1)^{d}\binom{d}{d}\right]$$
$$+ \mathrm{Bel}(\overline{A})\left[\binom{c}{1}-\binom{c}{2}+-\cdots+(-1)^{c+1}\binom{c}{c}\right]\left[\binom{d}{0}-\binom{d}{1}+-\cdots+(-1)^{d}\binom{d}{d}\right]$$
$$+ \left[\binom{d}{1}-\binom{d}{2}+-\cdots+(-1)^{d+1}\binom{d}{d}\right].$$

This last expression is equal to

	1	if $d > 0$,
	$\mathrm{Bel}(A)+\mathrm{Bel}(\overline{A})$	if $d=0$, $c > 0$, and $b > 0$,
	$\mathrm{Bel}(A)$	if $d=0$, $c > 0$, and $b=0$,
	$\mathrm{Bel}(\overline{A})$	if $d=0$, $c=0$, and $b > 0$,
and	0	if $d=0$, $c=0$, and $b=0$.

But $\mathrm{Bel}(A_1 \vee \cdots \vee A_n)$ will be equal to

	1	if $d > 0$,
	1	if $d=0$, $c > 0$, and $b > 0$,
	$\mathrm{Bel}(A)$	if $d=0$, $c > 0$, and $b=0$,
	$\mathrm{Bel}(\overline{A})$	if $d=0$, $c=0$, and $b > 0$,
and	0	if $d=0$, $c=0$, and $b=0$.

Hence Axiom III will indeed hold provided $\mathrm{Bel}(A)+\mathrm{Bel}(\overline{A}) \leq 1$. And it will not hold if $\mathrm{Bel}(A)+\mathrm{Bel}(\overline{A}) > 1$. Hence a function $\mathrm{Bel}: \{\Lambda, A, \overline{A}, V\} \to [0,1]$ is a belief function if and only if it satisfies $\mathrm{Bel}(\Lambda)=0$, $\mathrm{Bel}(V)=1$ and $\mathrm{Bel}(A)+\mathrm{Bel}(\overline{A}) \leq 1$.

One consequence of this is that our notion of a belief function is general enough to accommodate degrees of belief arising from James

Bernoulli's notion of a "pure argument." (See Bernoulli's Artis Conjectandi, pp. 218-220; or Part I of this essay.) Indeed, Bernoulli obtained "probabilities" of β/α and zero, respectively, for a thing and its opposite when β out of $\alpha=\beta+\gamma$ cases proved the thing but the other γ cases proved nothing. If we translate "probability" into "degree of belief" and "thing and its opposite" into "proposition and its negation," this becomes $\mathrm{Bel}(A)=\beta/\alpha$ and $\mathrm{Bel}(\overline{A})=0$.

C. The Senate Example

A more picturesque example of a belief function involves the first meeting of the United States Senate in 1789. At the time of that meeting, eleven States had ratified the Constitution. Of these eleven, five chose Federalists to fill both of their Senate seats, four chose Democratic-Republicans, and two, Connecticut and Pennsylvania, chose both a Federalist and a Democratic-Republican. The overall split was thus twelve to ten in favor of the Federalists. The first order of business for the Senate was to select a temporary presiding officer who would have the honor of counting the ballots that elected George Washington as the first President of the United States. I do not know how that presiding officer was in fact selected, but let us imagine that in order to avoid State rivalry for such a historical honor, it was done by lot rather than by vote. Imagine, indeed, the following procedure: a soldier is employed to choose at random the name of a State and then select as he pleases one of the two Senators from that State. Examining this situation before the selection is made and having no knowledge about the preferences of the soldier, what reasonable degree of belief might I accord the proposition, say, that a Democratic-Republican will be chosen?

The phrase "at random" may raise questions in some minds, but for my purposes it suffices to suppose that the selection of the State is to be carried out in such a way that I am willing to accord a degree of belief of 1/11 to the proposition that any particular State will be chosen. On the other hand, by saying that the soldier selects one of the two Senators from the resulting State "as he pleases," and adding that I have no knowledge of his preferences, I mean to convey the notion that I have no positive degree of belief that he will choose one or the other.

The algebra of all the propositions about who will chosen corresponds in a natural way to the field of all subsets of the set of the twenty-two

Langdon	(D)	Wingate	(D)
Few	(D)	Gunn	(D)
Lee	(D)	Grayson	(D)
Izard	(D)	Butler	(D)
Johnson	(D)	Ellsworth	(F)
Maclay	(D)	Morris	(F)
Strong	(F)	Dalton	(F)
Paterson	(F)	Elmer	(F)
Bassett	(F)	Read	(F)
Carroll	(F)	Henry	(F)
King	(F)	Schuyler	(F)

The 22 Senators

New Hampshire	(D,D)
Georgia	(D,D)
Virginia	(D,D)
South Carolina	(D,D)
Connecticut	(D,F)
Pennsylvania	(D,F)
Massachusetts	(F,F)
New Jersey	(F,F)
Delaware	(F,F)
Maryland	(F,F)
New York	(F,F)

The 11 States

D	D
D	D
D	D
D	D
D	F
D	F
F	F
F	F
F	F
F	F
F	F

The 2 Parties

Figure 1. The Senate Problem

Senators. For example, the proposition that either Senator Carroll or Senator King will be chosen corresponds to the subset {Carroll, King}. The situation is illustrated by Figure 1. In the first panel of that figure, the set of Senators is shown; the second panel represents the same set, partitioned only to the extent of dividing the States; while in the third panel the set is partitioned between Democratic-Republican and Federalist Senators.

My degree of belief that a Democratic-Republican will be chosen seems to be 4/11, for I have that degree of belief that New Hampshire, Georgia, Virginia or South Carolina will be chosen, in which case the soldier cannot help choosing a Democratic-Republican. I cannot add any of the belief committed to Connecticut or Pennsylvania to this, for I do not claim any positive degree of belief that the soldier will choose the Democratic-Republican rather than the Federalist in the event that one of those States is chosen. Similarly, my degree of belief that a Federalist will be chosen is 5/11. And in general my degree of belief Bel(A) that the Senator chosen will be in any given subset A of Senators will be k/11, where k is the number of States both of whose Senators are in A.

D. The Kansas Example

This final example is distinguished by the fact that the belief function is defined on an infinite Boolean algebra of propositions. Let us suppose that a military base is to be located somewhere in the State of Kansas, and that its exact location is to be determined as follows: One of the members of Congress from Kansas will be chosen at random, and he will be allowed to locate the base anywhere he pleases within the region he represents. Consider the Boolean algebra of all propositions of the form "The base will be located within R," where R is any region (or subset) of Kansas. What degrees of belief should one have for such propositions?

Well, there are seven Kansans in Congress; the five Representatives represent the districts shown in Figure 2, while each of the two Senators represent the State as a whole. Intuitively, our total belief must be divided into seven equal pieces, one corresponding to each of the seven politicians; and the degree of belief for the proposition "The base will be located in R" will be equal to k/7, where k is the number of districts which lie entirely within R. In particular, that degree of belief cannot exceed 5/7 unless R is the whole State, in which case the proposition is the sure proposition.

The collection $\mathcal{Q}$ of all propositions of the form "The base will be located in R," where R is a subset of Kansas, is indeed a Boolean algebra. And it is infinite, for there are an infinite number of subsets of Kansas.

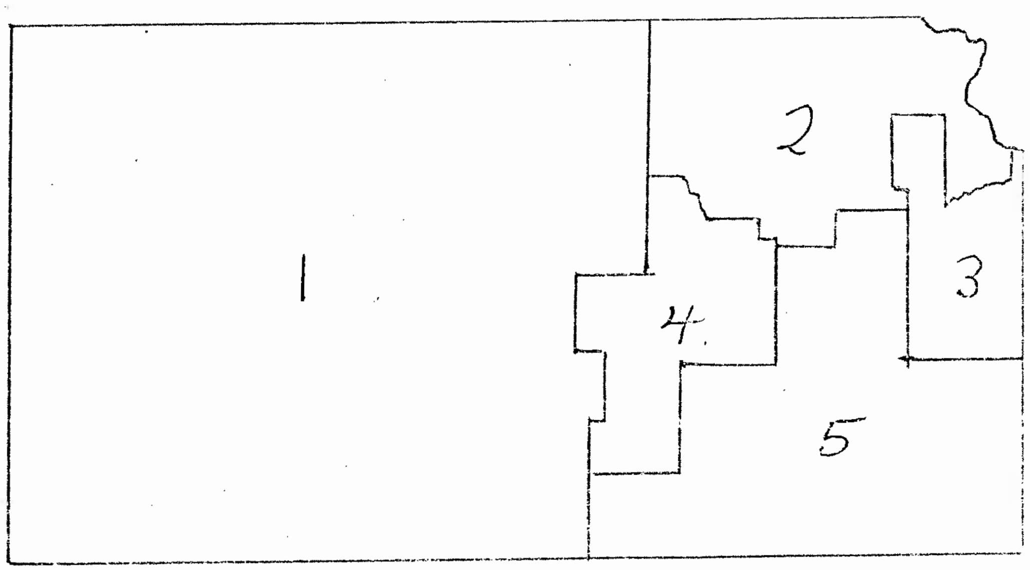

Figure 2. The Five Congressional Districts

And as we will see in the next chapter, the function

$$\text{Bel}: \mathcal{A} \rightarrow [0,1]: \text{"The base will be located in R."} \rightsquigarrow k/7,$$

where k is the number of districts lying entirely in R, is a belief function.

3. Upper Probabilities

The most striking feature of the preceding examples of degrees of belief is of course their failure to obey the rule of additivity, a failure that is most conspicuous in the case of a proposition and its negation. In practical terms, this failure of additivity means that one's degree of belief in a proposition does not necessarily determine one's degree of belief in its negation, so that the two quantities constitute distinct items of information. If degrees of belief were to follow the rule of additivity, then one's degree of belief $P(A)$ in a proposition A would determine one's degree of belief $P(\bar{A})$ in its negation through the relation $P(A)+P(\bar{A})=1$, or $P(\bar{A})=1-P(A)$; and once we knew someone's degree of belief in a proposition, we would learn nothing new if we were to be told his degree of belief in its negation. But the degrees of belief we have been studying do not work this way; we often have $\text{Bel}(A)+\text{Bel}(\bar{A}) < 1$, and knowledge of $\text{Bel}(A)$ does not guarantee knowledge of $\text{Bel}(\bar{A})$.

Another way of putting the matter is to say that a small value for $\text{Bel}(A)$ does not necessarily imply a large value for $\text{Bel}(\bar{A})$. Since $\text{Bel}(\bar{A})$ measures one's degree of belief in $\bar{A}$, or one's degree of disbelief in A, this assertion means, in English, that a low degree of belief does not necessarily imply a high degree of disbelief. In other words, we must distinguish between mere lack of belief and actual disbelief. Such a distinction is often left undrawn in everyday language: "I don't believe it" usually means "I believe the opposite." But it is a valuable distinction, and one that is usually made by careful thinkers. As an illustration of the distinction, consider again the proposition A="There is life on Mars," and its negation $\bar{A}$="There is no life on Mars." Suppose I know little about Mars, in particular have no reason to believe A, and accordingly have no belief in it whatever. Does this mean that I disbelieve A, i.e., that I have a strong belief in $\bar{A}$? I think not; it seems to me that an agnostic view is possible: I might entirely lack any belief either in A or $\bar{A}$. Or at less of an extreme, I might have no belief in A but only a mild belief in $\bar{A}$. For example, I might put $\mathrm{Bel}(A)=0$ and $\mathrm{Bel}(\bar{A})=\frac{1}{2}$.

A felicitous synonym for disbelief, as something susceptible of degree, is <u>doubt</u>, and this is the term I will employ in the sequel: one's degree of belief in $\bar{A}$ will be called one's degree of doubt for A. In this vocabulary, the assertion that both $\mathrm{Bel}(A)$ and $\mathrm{Bel}(\bar{A})$ might be small becomes the assertion that one might lack both belief and doubt for something. In many situations, one's degree of doubt for a proposition is more important than one's degree of belief in it. A low degree of doubt, for example, while not necessarily implying that one strongly believes a proposition, does indicate that one finds it plausible.

More generally, the extent to which one finds a proposition plausible is always inversely related to one's degree of doubt for it: the more one doubts it the less one finds it plausible. This fact leads us to think of the quantity $1-\mathrm{Bel}(\bar{A})$ as a measure of the extent to which one finds A plausible. As it turns out, this quantity will play an important role in our theory, and it will be convenient to have a name for it. Following A. P. Dempster, I will call $1-\mathrm{Bel}(\bar{A})$ the <u>upper probability</u> of A, and denote it by $P^*(A)$.

The function $P^*: \mathcal{A} \to [0,1] : A \rightsquigarrow 1-\mathrm{Bel}(\bar{A})$ will be called the upper probability function associated with Bel. It obviously conveys exactly the same information as Bel does, for Bel can be recovered from P^* through the relation $\mathrm{Bel}(A)=1-P^*(\bar{A})$. What are the rules for P^* that correspond to our rules for Bel? This question is answered by the following definition and theorem:

<u>Definition</u>. A function $P^*: \mathcal{A} \to [0,1]$ on a Boolean algebra $\mathcal{A}$ is an <u>upper probability function</u> if

(1). $P^*(\Lambda)=0$.

(2). $P^*(V)=1$.

(3). If $n \geq 1$ and $A_1,\ldots,A_n$ are elements of $\mathcal{A}$, then

$$P^*(A_1\wedge\cdots\wedge A_n) \leq \sum P^*(A_i) - \sum P^*(A_i \vee A_j) + - \cdots + (-1)^{n+1} P^*(A_1 \vee \cdots \vee A_n).$$

<u>Theorem</u>. A mapping $P^*: \mathcal{A} \to [0,1]$ is an upper probability function if and only if the mapping $\mathrm{Bel}: \mathcal{A} \to [0,1]$ defined by $\mathrm{Bel}(A)=1-P^*(\bar{A})$ is a belief function.

<u>Proof</u>: The only non-trivial part of the proof is the demonstration that the inequality (3) for P^* is equivalent to the third axiom for belief functions for Bel. Substituting $1-\mathrm{Bel}(\bar{A})$ for $P^*(A)$ in (3) gives

$$\begin{aligned} 1-\mathrm{Bel}(\bar{A}_1 \vee \cdots \vee \bar{A}_n) &\leq \sum \left[1-\mathrm{Bel}(\bar{A}_i)\right] - \sum \left[1-\mathrm{Bel}(\bar{A}_i \wedge \bar{A}_j)\right] + - \cdots + (-1)^{n+1} \left[1-\mathrm{Bel}(\bar{A}_1 \wedge \cdots \wedge \bar{A}_n)\right] \\ &= \left[\binom{n}{1} - \binom{n}{2} + - \cdots + (-1)^{n+1}\binom{n}{n}\right] \\ &\quad - \left[\sum \mathrm{Bel}(\bar{A}_i) - \sum \mathrm{Bel}(\bar{A}_i \wedge \bar{A}_j) + - \cdots + (-1)^{n+1} \mathrm{Bel}(\bar{A}_1 \wedge \cdots \wedge \bar{A}_n)\right]. \end{aligned}$$

Since $\binom{n}{1}-\binom{n}{2}+-\cdots+(-1)^{n+1}\binom{n}{n}=\binom{n}{0}-(1-1)^n=1$, this is equivalent to

$$Bel(\bar{A}_1\vee\cdots\vee\bar{A}_n)\geq\sum Bel(\bar{A}_i)-\sum Bel(\bar{A}_i\wedge\bar{A}_j)+-\cdots+(-1)^{n+1}Bel(\bar{A}_1\wedge\cdots\wedge\bar{A}_n).$$

Since (3) is equivalent to this last inequality for all $\bar{A}_1,\ldots,\bar{A}_n$ in $\mathcal{Q}$, it is also equivalent to it for all $A_1,\ldots,A_n\in\mathcal{Q}$, for every proposition in a Boolean algebra is the negation of some other proposition in it. But this gives us precisely the third axiom for belief functions.

Rule (3) for upper probability functions can be written in another way which is also useful.

<u>Theorem.</u> Suppose f is a real-valued function on a Boolean algebra . Then f is an upper probability function if and only if

(i). $f(\Lambda)=0$.

(ii). $f(V)=1$.

(iii). If $n\geq 1$ and $B,A_1,\ldots,A_n$ are elements of $\mathcal{Q}$, then

$$f(B)-\sum f(B\vee A_i)+\sum f(B\vee A_i\vee A_j)-+\cdots+(-1)^n f(B\vee A_1\vee\cdots\vee A_n)\leq 0.$$

<u>Proof:</u> Suppose indeed that f is an upper probability function. Then applying rule (3) to $B\vee A_1,\ldots,B\vee A_n$ yields

$$f(B\vee(\wedge A_i))\leq\sum f(B\vee A_i)-\sum f(B\vee A_i\vee A_j)+-\cdots+(-1)^{n+1}f(B\vee A_1\vee\cdots\vee A_n).$$

Since upper probability functions always take non-negative values, rule (3) implies in particular that f is monotone. Hence $f(B)\leq f(B\vee(\wedge A_i))$, and (iii) follows.

To see that (iii) implies rule (3) for upper probability functions, set $B=A_1\wedge\cdots\wedge A_n$ and transfer all the terms on the left-hand side of (iii) except the first to the right-hand side.

Finally, setting $n=1$ and choosing B and A_1 so that $B\leq A_1$, (iii) becomes $f(B)\leq f(A_1)$. Hence if f obeys (i)–(iii) it will take values between zero and one. Hence f [illegible]

The quantity

$$f(B)-\sum f(B\vee A_i)+\sum f(B\vee A_i\vee A_j)-+\cdots+(-1)^n f(B\vee A_1\vee\cdots\vee A_n)$$

can be written somewhat more compactly as

$$\sum_{J\subset\{1,\ldots,n\}} (-1)^{\text{card } J}\, f(B\vee(\bigvee_{i\varepsilon J} A_i)),$$

where card J, or the <u>cardinality</u> of J, is the number of elements in J, and $\bigvee_{i\varepsilon\emptyset} A_i$ is understood to be equal to Λ. It is sometimes denoted $\nabla_n^f(B;A_1,\ldots,A_n)$ and called the <u>n^{th} successive difference</u> of f(B) with respect to $A_1,\ldots,A_n$. This terminology derives from the fact that the quantities $\nabla_n^f(B;A_1,\ldots,A_n)$ can be specified recursively by the relations

$$\nabla_1^f(B;A_1)=f(B)-f(B\vee A_1)$$

and

$$\nabla_{n+1}^f(B;A_1,\ldots,A_{n+1})=\nabla_n^f(B;A_1,\ldots,A_n)-\nabla_n^f(B\vee A_{n+1};A_1,\ldots,A_n).$$

(See Choquet, p. 169.)

4. The Logical and Subjective Vocabularies

The theory that we have been developing in this chapter is overtly subjective. It is a theory of belief, and it deals with the degrees to which we believe and doubt propositions, not with the degrees to which they deserve belief or doubt. But the subjective notions of degree of belief and upper probability are obviously parallel to the logical notions of degree of support and degree of plausibility, developed in Part I of this essay. That parallelism, as exhibited in Table 1, connects belief with support and upper probability with plausibility.

The notions of support and plausibility are not subjective, for they refer to the objective relation between a proposition and the evidence for and against it; we take it as an objective, if sometimes elusive fact that the evidence does or does not support a proposition to a given degree, or that it does or does not leave it plausible to a given degree. But these logical quantities, if they are known, obviously determine the degrees of belief and the upper probabilities that we ought to have: we ought to believe a proposition to the extent that the evidence supports it, and our upper probability for a proposition, or the degree to which we find the proposition plausible, ought to equal its actual degree of plausibility.

My ultimate interest in this essay lies on the logical side of the ledger in Table 1; I want to measure degrees of support for statistical hypotheses. Why, then, am I constructing a subjective theory? The answer, of course, is that I will eventually want to impose on support functions the rules and structure that is being developed here for belief functions. Such an imposition will be partially justified by the general argument that degrees of support correspond to reasonable degrees of belief given the evidence and hence should obey rules that are appropriate for degrees of belief.

Subjective		Logical	
Degree of Belief	$\mathrm{Bel}(A)$	Degree of Support	$S(A)$
Degree of Doubt	$\mathrm{Bel}(\bar{A})$	Degree of Dubiety	$S(\bar{A})$
Upper Probability	$1-\mathrm{Bel}(\bar{A})$	Degree of Plausibility	$1-S(\bar{A})$

Table 1. The Two Vocabularies

5. Probabilities as Degrees of Belief

The examples of Bernoulli and Lambert would provide some historical justification for the claim that degrees of belief satisfying our axioms for belief functions deserve to be called "probabilities," and I am tempted to make such a claim. But it is doubtful that such a claim would be accepted. Since the time of Laplace probabilities have had to be additive, and it seems likely that they will remain under that constraint for a good while.

A probability function on a Boolean algebra $\mathcal{A}$, then, is still a function $P: \mathcal{A} \to [0,1]$ that obeys the rules

(1). $P(\Lambda)=0$,

(2). $P(V)=1$,

and (3). $P(A)+P(B)=P(A \vee B)$ whenever $A, B \in \mathcal{A}$ and $A \wedge B = \Lambda$.

Since belief functions do not need to obey (3), they need not be probability functions. On the other hand, a probability function always qualifies as a belief function. To prove that this is so, it is only necessary to show that a probability function always obeys the inequalities in Axiom III for belief functions. Actually, a probability function always satisfies those inequalities with equality.

Theorem. If $P: \mathcal{A} \to [0,1]$ is a probability function and $A_1, \ldots, A_n$ are any elements of $\mathcal{A}$, then

$$P(A_1 \vee \cdots \vee A_n) = \sum P(A_i) - \sum P(A_i \wedge A_j) + - \cdots + (-1)^{n+1} P(A_1 \wedge \cdots \wedge A_n).$$

Proof: Consider the set of all vectors $x=(x_1, \ldots, x_n)$ such that each x_i is equal to either zero or one. For each such x, set $A_x = B_1 \wedge \cdots \wedge B_n$, where $B_i = A_i$ if $x_i = 1$ and $B_i = \overline{A}_i$ if $x_i = 0$. Then all the A_x are in $\mathcal{Q}$, and they are pairwise incompatible. Further, $A_1 \vee \cdots \vee A_n = \vee\{A_x \mid \text{some } x_i = 1\}$ and $A_{i_1} \wedge \cdots \wedge A_{i_k} = \vee\{A_x \mid x_{i_1} = \cdots = x_{i_k} = 1\}$. Thus

$$P(A_1 \vee \cdots \vee A_n) = \sum\{P(A_x) \mid \text{some } x_i = 1\},$$

and

$$\sum P(A_i) - \sum P(A_i \wedge A_j) + - \cdots + (-1)^{n+1} P(A_1 \wedge \cdots \wedge A_n)$$
$$= \sum\left(\sum\{P(A_x) \mid x_i = 1\}\right) - \sum\left(\sum\{P(A_x) \mid x_i = x_j = 1\}\right) + - \cdots + (-1)^{n+1} P(A_1 \wedge \cdots \wedge A_n).$$

The number of times that any particular x appears in this summation is evidently determined by the number of ones in x. In fact, if x contains r ones, then A_x will occur $\binom{r}{1} - \binom{r}{2} + - \cdots + (-1)^{r+1}\binom{r}{r}$ times. But this is equal to 1 unless r=0. Hence $P(A_x)$ occurs the same number of times in this formula as in the formula for $P(A_1 \vee \cdots \vee A_n)$, and the equation is correct. ▨

Obviously, a belief function is a probability function if and only if it obeys the axiom of additivity. More interesting, in light of the discussion in Part I, is the fact that a belief function is a probability function if and only if it obeys the special case of the axiom of additivity that relates the belief in a proposition to the belief in its negation. The proof of the non-trivial part of this assertion is given below:

<u>Theorem.</u> If a belief function Bel on a Boolean algebra $\mathcal{Q}$ satisfies $\mathrm{Bel}(A) + \mathrm{Bel}(\overline{A}) = 1$ for all $A \in \mathcal{Q}$, then Bel is a probability function.

<u>Proof:</u> We need to show that $\mathrm{Bel}(A \vee B) = \mathrm{Bel}(A) + \mathrm{Bel}(B)$ for any two elements $A, B \in \mathcal{Q}$ such that $A \wedge B = \Lambda$. But we already know, by the rule of superadditivity, that $\mathrm{Bel}(A \vee B) \geq \mathrm{Bel}(A) + \mathrm{Bel}(B)$, so we need

only show that $Bel(A \vee B) \leq Bel(A)+Bel(B)$. But if we apply Axiom III to $\overline{A}$ and $\overline{B}$, we obtain

$$Bel(\overline{A})+Bel(\overline{B}) \leq Bel(\overline{A}\vee\overline{B})+Bel(\overline{A}\wedge\overline{B}),$$

and substituting $1-Bel(\overline{X})$ for $Bel(X)$ in each term of this inequality yields

$$1-Bel(A)+1-Bel(B) \leq 1-Bel(A\wedge B)+1-Bel(A\vee B).$$

Since $Bel(A\wedge B)=0$, this becomes $Bel(A\vee B) \leq Bel(A)+Bel(B)$. ▨

So a belief function Bel is a probability function if and only if it satisfies $Bel(A)=1-Bel(\overline{A})$ for all A. But in general the upper probability function associated with a belief function Bel is given by $P^*(A)=1-Bel(\overline{A})$. So a belief function is a probability function if and only if it is identical to its upper probability function.

It is worth noting that the rule

$$P(A_1\vee\cdots\vee A_n)=\sum P(A_i)-\sum P(A_i\wedge A_j)+-\cdots+(-1)^{n+1}P(A_1\wedge\cdots\wedge A_n)$$

for probability functions also implies the rule

$$P(A_1\wedge\cdots\wedge A_n)=\sum P(A_i)-\sum P(A_i\vee A_j)+-\cdots+(-1)^{n+1}P(A_1\vee\cdots\vee A_n).$$

To derive the second equation from the first, one need only replace each A_i by $\overline{A}_i$. Hence probability functions also satisfy rule (3) for upper probability functions with equality.

From the point of view of our theory, then, a probability function is a special kind of belief function, and a "subjective probability" is a special kind of degree of belief. Indeed, it might be called a two-sided degree of belief, for it supplies a degree of belief both for a proposition and for the negation of that proposition.

6. Discounting Belief Functions

It often happens that we obtain our opinions and beliefs on a topic from someone else in whose judgment we have a reasonable degree of confidence. In most cases, of course, we will not have absolute confidence in this other person's opinions and hence will wish to discount those opinions at least slightly before adopting them as our own. This process of discounting can be represented quite simply in the theory of belief functions.

Suppose, indeed, that the other person's belief function is $\mathrm{Bel}_0 : \mathcal{A} \to [0,1]$, and that one's degree of confidence in the other person's judgment is α, which is some number between zero and one. Then the natural thing to do is to adopt the quantity $\alpha \cdot \mathrm{Bel}_0(A)$ as one's degree of belief in any proposition $A \in \mathcal{A}$ that is not the sure proposition $\mathcal{V}$. Formally, one would adopt the belief function $\mathrm{Bel}: \mathcal{A} \to [0,1]$ defined by $\mathrm{Bel}(\mathcal{V})=1$ and $\mathrm{Bel}(A)= \alpha \cdot \mathrm{Bel}_0(A)$ for all $A \neq \mathcal{V}$. It is easily verified that the function Bel defined in this way is indeed a belief function.

The process of discounting a belief function is a special case of the process of taking a linear mixture of two or more belief functions. If Bel_1 and Bel_2 are two different belief functions on the same Boolean algebra of propositions $\mathcal{A}$, and if α is a number between zero and one, then the function $\mathrm{Bel}: \mathcal{A} \to [0,1]$ defined by $\mathrm{Bel}(A)= \alpha\, \mathrm{Bel}_1(A)+(1-\alpha)\mathrm{Bel}_2(A)$ for all $A \in \mathcal{A}$ will be a belief function; it is said to be a linear mixture of Bel_1 and Bel_2. It is evident that discounting a belief function Bel_0 by the factor α is the same as taking a linear mixture of Bel_1 and the vacuous belief function, using coefficients α and $(1-\alpha)$.

When a belief function is passed from person to person, being discounted each time, the degrees of belief accorded to the non-sure propositions constantly decrease. Hence the notion of discounting belief functions can be used to represent the diminishing credence that we lend to hearsay or to any tradition of testimony as its source becomes more remote.

These ideas are hardly novel. In fact, they were quite common in the eighteenth century discussions of the probability of testimony, which were much concerned with the bothersome idea that the probability of the scriptures diminishes with time. By and large, though, the notion of discounting "probabilities" did not survive into the nineteenth century. Its failure to survive can be attributed to its conflict with the rule of additivity for probabilities; for once additivity is assumed, the diminishing probability of the tradition comes to imply an increasing probability for the denial of the tradition--and this seems less reasonable.

7. A Counterexample

In Part I, I strongly criticized the attempt by some students of subjective probability to insist that "rational" degrees of belief ought to obey the rule of additivity. In fact, I questioned the very idea that abstract considerations could lead to rules that were absolutely obligatory for all reasonable systems of degrees of belief. But what about the rules that I have offered in this chapter? Are there reasonable systems of degrees of belief that would violate them?

There are, and it is easy to construct examples. One general method for constructing such examples is provided by the notion of an aleatory law. An aleatory law P on a set $\mathcal{X}$ is a function $P: \mathcal{P}(\mathcal{X}) \rightarrow [0,1]$, where $\mathcal{P}(\mathcal{X})$ is the set of all subsets of a set $\mathcal{X}$, and $P(A)$, for each $A \subset \mathcal{X}$, is taken to be the chance or objective probability that the outcome of a certain experiment or process will be in A. It is a commonplace that if we were really certain that some process were governed by an aleatory law then we would be justified in adopting as our degree of belief in the occurrence of a given event the chance assigned to that event by the aleatory law. The set $\mathcal{P}(\mathcal{X})$ can be interpreted, of course, as a Boolean algebra, and the resulting system of degrees of belief would be a probability function and hence a belief function. More generally, though, we might contemplate the situation, however fictional, in which we are absolutely certain that the process is governed by one of a given collection $\{P_\theta\}_{\theta \in \Theta}$ of aleatory laws. In such a case we might be justified in adopting as our degree of belief in the given event the infimum of the chances assigned that event by the various laws P_θ. More precisely, if the aleatory laws were on an observation space $\mathcal{X}$, we might define $B: \mathcal{P}(\mathcal{X}) \rightarrow [0,1]$ by $B(A) = \inf_{\theta \in \Theta} P_\theta(A)$. Such a function B will in general not be a probability function. And while it will satisfy $B(\emptyset)=0$ and $B(\mathcal{X})=1$, it will not satisfy Axiom III for belief functions unless the class of aleatory laws $\{P_\theta\}_{\theta \in \Theta}$ is chosen with particular care.

Dempster has given the following example where the function B does not satisfy Axiom III. Letting $\mathcal{X}$ consist of the four possibilities $\{bb, bw, wb, ww\}$, we contemplate the three aleatory laws given by

$$P_1(bb)=\tfrac{1}{4},\quad P_1(bw)=\tfrac{1}{4},\quad P_1(wb)=\tfrac{1}{4},\quad P_1(ww)=\tfrac{1}{4};$$
$$P_2(bb)=\tfrac{1}{2},\quad P_2(bw)=0,\quad P_2(wb)=\tfrac{1}{2},\quad P_2(ww)=0;$$
$$P_3(bb)=0,\quad P_3(bw)=\tfrac{1}{2},\quad P_3(wb)=0,\quad P_3(ww)=\tfrac{1}{2}.$$

We could imagine this situation arising if $\mathcal{X}$ consisted of all the possible results from drawing balls successively from two urns, the first of which was known to contain one black and one white ball, and the second of which might contain either one of each color or else two of the same color. The aleatory law P_1 would then correspond to the case where the second urn contained one black and one white ball, P_2 would correspond to the case where it contained two black balls, and P_3 to the case where it contained two white balls. Setting $A_1=\{bb, bw\}$ and $A_2=\{bb, ww\}$, we obtain $B(A_1)=B(A_2)=B(A_1 \vee A_2)=\frac{1}{2}$ and $B(A_1 \wedge A_2)=0$; and this violates the requirement that $B(A_1)+B(A_2)-B(A_1 \wedge A_2)$ should not exceed $B(A_1 \vee A_2)$.

8. Axiom III

In a sense, the third axiom for belief functions includes an infinite number of axioms, one for each natural number n. One might hope at first that it should be unnecessary to have so many axioms; perhaps the first few would imply the others. Unfortunately, though, it is necessary to state the axiom for an infinite number of different integers; for while the truth of the axiom for a given value of n implies its truth for smaller values, it does not imply its truth for larger values. This section is devoted to establishing these facts.

Theorem. Suppose $\mathcal{A}$ is a Boolean algebra, n is a natural number, and $B: \mathcal{A} \to [0,1]$ satisfies

(i). $B(\Lambda)=0$,

(ii). $B(V)=1$,

and (iii). $B(A_1 \vee \cdots \vee A_n) \geq \sum B(A_i) - \sum B(A_i \wedge A_j) + \cdots + (-1)^{n+1} B(A_1 \wedge \cdots \wedge A_n)$

for all sequences $A_1, \ldots, A_n$ of n elements of $\mathcal{A}$. Then

$$B(A_1\vee\cdots\vee A_{n-1})\geq\sum B(A_i)-\sum B(A_i\wedge A_j)+-\cdots+(-1)^n B(A_1\wedge\cdots\wedge A_{n-1})$$

for all sequences $A_1,\ldots,A_{n-1}$ of n-1 elements of $\mathcal{A}$.

<u>Proof:</u> Suppose $A_1,\ldots,A_{n-1}$ are elements of $\mathcal{A}$, and set $A_n=\Lambda$.

Then by (iii),

$$B(A_1\vee\cdots\vee A_{n-1})=B(A_1\vee\cdots\vee A_n)\geq\sum_{i\leq n} B(A_i)-\sum_{i<j\leq n}B(A_i\wedge A_j)+-\cdots+(-1)^{n+1}B(A_1\wedge\cdots\wedge A_n)$$
$$=\sum_{i\leq n-1} B(A_i)-\sum_{i<j\leq n-1}B(A_i\wedge A_j)+-\cdots+(-1)^n B(A_1\wedge\cdots\wedge A_{n-1}).$$

<u>Theorem.</u> Let n be a natural number. Then there exists a Boolean algebra $\mathcal{A}$ and a function $B:\mathcal{A}\to[0,1]$ such that

(i). $B(\Lambda)=0$,

(ii). $B(V)=1$,

(iii). $B(S_1\vee\cdots\vee S_n)\geq\sum B(S_i)-\sum B(S_i\wedge S_j)+-\cdots+(-1)^{n+1}B(S_1\wedge\cdots\wedge S_n)$

for all sequences $S_1,\ldots,S_n$ of n elements of $\mathcal{A}$, and yet

(iv). $B(A_1\vee\cdots\vee A_{n+1})<\sum B(A_i)-\sum B(A_i\wedge A_j)+-\cdots+(-1)^{n+2}B(A_1\wedge\cdots\wedge A_{n+1})$

for some sequence $A_1,\ldots,A_{n+1}$ of n+1 elements of $\mathcal{A}$.

The rest of this section is devoted to an example establishing this theorem. Set $\mathcal{S}$ equal to a set of n+2 elements:

$$\mathcal{S}=\{a_1,\ldots,a_{n+2}\},$$

and set $\mathcal{A}=\mathcal{P}(\mathcal{S})$, the set of all subsets of $\mathcal{S}$ interpreted as a Boolean algebra. Define $B:\mathcal{P}(\mathcal{S})\to[0,1]$ by setting B(A) equal to

	1	if $A=\mathcal{S}$,
	2/(n+1)	if A includes a_1 and n of $\{a_2,\ldots,a_{n+2}\}$,
	1/(n+1)	if A includes a_1 but fewer than n of $\{a_2,\ldots,a_{n+2}\}$,
and	0	if A does not include a_1.

Since $V=\mathcal{S}$ and $\Lambda=\emptyset$, conditions (i) and (ii) of the theorem are true for this example. The other two are also true, but more difficult to demonstrate.

First, let us establish (iv). To this end, note that a subset A of $\mathcal{S}$ satisfies $B(A)=2/(n+1)$ if and only if $A=\mathcal{S}-\{a_i\}$ for some i between 2 and n+1. Hence there are exactly n+1 distinct subsets of $\mathcal{S}$ that have a value of B equal to $2/(n+1)$. Denote these by $A_1,\ldots,A_{n+1}$. Then $B(A_1\vee\cdots\vee A_{n+1})=1$, while

$$\begin{aligned}
&\sum B(A_i)-\sum B(A_i\wedge A_j)+\cdots+(-1)^{n+2}B(A_1\wedge\cdots\wedge A_{n+1})\\
&=\binom{n+1}{1}(2/(n+1))-\binom{n+1}{2}(1/(n+1))+\binom{n+1}{3}(1/(n+1))-+\cdots+(-1)^{n+2}\binom{n+1}{n+1}(1/(n+1))\\
&=1+(1/(n+1))\left[\binom{n+1}{1}-\binom{n+1}{2}+-\cdots+(-1)^{n+2}\binom{n+1}{n+1}\right]\\
&=1+(1/(n+1)).
\end{aligned}$$

Hence (iv) is satisfied by the sequence $A_1,\ldots,A_{n+1}$.

Now let us establish (iii). Actually, we will establish that whenever $1\leq k\leq n$ and $S_1,\ldots,S_k$ are subsets of $\mathcal{S}$,

$$B(S_1\vee\cdots\vee S_k)\geq\sum B(S_i)-\sum B(S_i\wedge S_j)+-\cdots+(-1)^{k+1}B(S_1\wedge\cdots\wedge S_k). \qquad (1)$$

<u>Case 1</u>. $B(S_i)=2/(n+1)$ for $i=1,\ldots,k$. Let $A_1,\ldots,A_{n+1}$ be as above, and for each j, $j=1,\ldots,n+1$, let k_j be the number of the S_i equal to A_j. Then $k=k_1+\cdots+k_{n+1}$. And

$$B(S_i)=(2/(n+1))k=(1/(n+1))\left[\binom{k}{1}+\sum\binom{k_j}{1}\right],$$

$$B(S_i\wedge S_j)=(2/(n+1))\sum\binom{k_j}{2}+(1/(n+1))\left[\binom{k}{2}-\sum\binom{k_j}{2}\right]=(1/(n+1))\left[\binom{k}{2}+\sum\binom{k_j}{2}\right],$$

etc. Hence

$$\begin{aligned}
&\sum B(S_i)-\sum B(S_i\wedge S_j)+-\cdots+(-1)^{k+1}B(S_1\wedge\cdots\wedge S_k)\\
&=(1/(n+1))\left[\binom{k}{1}-\binom{k}{2}+-\cdots+(-1)^{k+1}\binom{k}{k}\right]+(1/(n+1))\left[\sum\binom{k_j}{1}-\sum\binom{k_j}{2}+-\cdots+(-1)^{k+1}\sum\binom{k_j}{k}\right.\\
&=(r+1)/(n+1),
\end{aligned}$$

where r is the number of j for which $k_j>0$, and $1\leq r\leq n$. If there is only one such j, then the above becomes $2/(n+1)$, which will be equal to $B(S_1\vee\cdots\vee S_k)=B(A_j)$. If there is more than one, then $B(S_1\vee\cdots\vee S_k)=1$, but $(r+1)/(n+1)$ will still not exceed 1, so (1) will still hold.

<u>Case 2</u>. Some of the S_i do not have $B(S_i)=2/(n+1)$. Let s be the number of the i, i=1,...,k, for which $B(S_i)\neq 2/(n+1)$. Then let us establish the inequality (1) by induction on s. The case s=0 was established in the preceding paragraph. So suppose $s \geq 1$, and suppose (1) holds for all smaller values of s. We may also assume that k is one of the i for which $B(S_i)\neq 2/(n+1)$. And the right-hand side of (1) becomes

$$\sum B(S_i)-\sum B(S_i\wedge S_j)+-\cdots+(-1)^{k+1}B(S_1\wedge\cdots\wedge S_k)$$

$$=B_1+B_2,$$

where

$$B_1=\sum_{i\leq k-1} B(S_i)-\sum_{i<j\leq k-1} B(S_i\wedge S_j)+-\cdots+(-1)^k B(S_1\wedge\cdots\wedge S_{k-1}),$$

and

$$B_2=B(S_k)-(\sum_{i\leq k-1} B(S_k\wedge S_i)-\sum_{i<j\leq k-1} B(S_k\wedge S_i\wedge S_j)+-\cdots+(-1)^k B(S_k\wedge S_1\wedge\cdots\wedge S_{k-1})).$$

By the inductive hypothesis, $B(S_1\vee\cdots\vee S_{k-1})\geq B_1$. Now consider separately the cases where $B(S_k)=1$ and where $B(S_k)$ is less than $2/(n+1)$. In the first case, $B(S_1\vee\cdots\vee S_k)=1$ and $B_2=1-B_1$, so $B(S_1\vee\cdots\vee S_k)=1=B_1+B_2$, and (1) holds. In the second case, $B_2=0$, and $B(S_1\vee\cdots\vee S_k)\geq B(S_1\vee\cdots\vee S_{k-1})\geq B_1=B_1+B_2$, and again (1) holds. This completes the demonstration.

▨

9. <u>Bibliographic Notes</u>

Axiom III for belief functions was derived by A. P. Dempster for "lower probabilities induced by a multivariate mapping" in his 1967 paper in the <u>Annals of Mathematical Statistics</u>. Earler, Gustave Choquet had used Axiom III to define a "monotone set function of order ∞." (See pp. 169-71 of his "Theory of Capacities.") To my knowledge, though, no one has previously adduced these inequalities as intuitively appealing rules for degrees of belief.

The example involving the first United States Senate takes some liberties with history. Actually, only twelve of the twenty-two Senators were present on April 6, 1789, when the Senate elected John Langdon of New Hampshire as its President pro tempore. (See De Pauw, p. 8.) Furthermore, the division into the two parties was not clearly established at that time, so that the affiliations I have imputed to the various Senators are open to dispute. They are based on the votes of July 18, 1789, on the bill establishing a Department of Foreign Affairs, and the votes of August 4, 1789, on the bill establishing a Department of War. (De Pauw, pp. 86-7 and 104-6.)

For more information on the "non-additive probabilities" obtained by James Bernoulli and Johann Heinrich Lambert, the reader may consult pp. 218-220 of Bernoulli's Artis Conjectandi and pp. 318-421 of Volume 2 of Lambert's Neues Organon. Bernoulli's and Lambert's ideas are discussed in detail in Part I of this essay. References to the eighteenth century discussion of how the probability of testimony diminishes with its transmission can be found in Todhunter's history, pp. 54, 462 and 500. The matter was also discussed by Diderot in the article "Probabilité" of his famous Encyclopédie.

The example reproduced in section 7 was given on pp. 51-3 of Dempster's The Theory of Statistical Inference.

CHAPTER 2. ALLOCATIONS OF PROBABILITY

This chapter develops explicitly the intuitive picture underlying the axioms for belief functions. This results in the mathematical notion of an allocation of probability, and in the theorem that every belief function can be represented by an allocation of probability.

1. Constraint Relations

I used the term "portions of belief" in the preceding chapter so as to emphasize the differences between the theory developed there and probability theory, but it is evident that the intuition involved is really quite close to the intuition of students of subjective probability, who are accustomed to thinking of their probability as a measurable substance that can be divided into various pieces and distributed over a Boolean algebra of propositions. Indeed, it is in the method of distribution rather than in the nature of the abstracted probability that the differences will be found between the theory of belief functions and the more special theory of probability functions. Hence I find it entirely appropriate to follow the probabilist in using the word probability in place of the word belief when I am thinking of belief as something admitting of degree, and in the rest of this essay I will speak of pieces of probability or of probability masses rather than of portions of belief.

In this vocabulary, the intuitive picture developed in the preceding chapter involves the division of our probability into various probability masses, each of which may or may not be associated with or committed to a given proposition. There are of course restrictions on our freedom to commit probability masses to propositions; when I was adducing the rules for belief functions I mentioned the following ones:

(I) No probability mass may be committed to Λ.

(II) Every probability mass must be committed to V.

(III) If $A_1 \leq A_2$, then any probability mass committed to A_1 must also be committed to A_2.

(IV) Any probability mass that is committed to both A_1 and A_2 must also be committed to $A_1 \wedge A_2$.

This list is not exhaustive, though; we can easily extend it if we think a little more about the relations among our probability masses.

Our various probability masses are not conceived of in isolation; they are all pieces of the same fixed quantity of idealized substance representing our probability, and hence they can bear various relations to each other. For example, one probability mass may be part of another. Or one may consist precisely of the overlap between a pair of others, or perhaps of all the probability that is in either one or the other of a pair of others. I will write $M_1 \leq M_2$ to indicate that M_1 is part of M_2, or is contained in M_2; and I will denote by $M_1 \vee M_2$ the "union" of M_1 and M_2, or the probability mass consisting of all the probability that is in either M_1 or M_2.

Once we have established the ideas of containment and union for probability masses, the following additional rules impose themselves on the relation of "commitment" between probability masses and propositions:

(V) If the probability mass M_1 is committed to A and $M_2 \leq M_1$, then M_2 is also committed to A.

(VI) If the probability masses M_1 and M_2 are both committed to A, then the probability mass $M_1 \vee M_2$ is committed to A.

Evidently, the collection of our probability masses is beginning to acquire the same formal structure possessed by the collections of propositions we have dealt with; it is beginning to resemble a Boolean algebra.

In order to develop this structure further, let us denote the collection of our probability masses by the letter $\mathcal{M}$. We already have the relation "$\leq$" of containment which holds between some pairs of elements of $\mathcal{M}$; and for any two elements M_1, M_2, we have an element $M_1 \vee M_2$ which is their union. Intuitively, we should also have for each pair M_1, M_2 a probability mass $M_1 \wedge M_2 \epsilon \mathcal{M}$ representing their overlap or "intersection." And for each element $M \epsilon \mathcal{M}$ there ought to be an element $\overline{M} \epsilon \mathcal{M}$ which consists precisely of the probability that is not in M. There are difficulties, though, with the symbols "$\wedge$" and "$-$". The difficulty with writing $M_1 \wedge M_2$ is that M_1 and M_2 might be "disjoint" -- they might fail to overlap. In such a case there would be no probability mass for $M_1 \wedge M_2$ to denote. Similarly, if $M \epsilon \mathcal{M}$ is the probability mass consisting of all our probability, then there will be no probability left over to constitute the probability mass $\overline{M}$. Both of these problems can be met by the invention of a "null" probability mass, thought of as consisting of no probability at all. If we denote this null probability mass by Λ or $\Lambda_{\mathcal{M}}$ and denote the probability mass consisting of all our probability by V or $V_{\mathcal{M}}$, then we will be able to set $M_1 \wedge M_2 = \Lambda_{\mathcal{M}}$ whenever M_1 and M_2 do not overlap, and we will be able to set $\overline{V_{\mathcal{M}}} = \Lambda_{\mathcal{M}}$. It will also be convenient to establish the convention that $\Lambda_{\mathcal{M}} \leq M$ for all $M \epsilon \mathcal{M}$.

Our collection $\mathcal{M}$ of probability masses is now endowed with all symbols we have used for Boolean algebras of propositions. It has a relation "$\leq$", operations "$\wedge$", "$\vee$" and "$\overline{\ \ }$", and distinguished elements "Λ" and "V". Furthermore, these symbols have all the properties that we have been accustomed to in Boolean algebras of propositions. For example, $\Lambda \leq M \leq V$ for all $M \in \mathcal{M}$; and for any $M \in \mathcal{M}$, $M \wedge \overline{M} = \Lambda$ and $M \vee \overline{M} = V$. In the following pages I will call $\mathcal{M}$ a "Boolean algebra of probability masses" or a "probability algebra," and I will use these symbols and their properties freely.

In assuming that our probability is represented mathematically as a Boolean algebra $\mathcal{M}$, I am again taking for granted that the structure of Boolean algebras is intuitively clear. The reader who is dissatisfied with this intuitive approach may wish to turn to the first two sections of Chapter 3, where Boolean algebras are defined and studied abstractly.

We are dealing, then, with a Boolean algebra of probability masses $\mathcal{M}$ and a Boolean algebra of propositions $\mathcal{Q}$, and we have six rules that govern the relation of "commitment" between a probability mass $M \in \mathcal{M}$ and a proposition $A \in \mathcal{Q}$. If we write "M ct A" to signify that M is committed to A, these rules can be listed more neatly as follows:

(1) (a) If M ct A_1 and $A_1 \leq A_2$, then M ct A_2

(b) If M ct A_1 and M ct A_2, then M ct $A_1 \wedge A_2$.

(c) M ct $V_{\mathcal{Q}}$ for all $M \in \mathcal{M}$.

(2) (a) If M_1 ct A and $M_2 \leq M_1$, then M_2 ct A.

(b) If M_1 ct A and M_2 ct A, then $M_1 \vee M_2$ ct A.

(c) $\Lambda_{\mathcal{M}}$ ct A for all $A \in \mathcal{Q}$.

(3) If M ct $\Lambda_{\mathcal{Q}}$ then $M = \Lambda_{\mathcal{M}}$.

The last rule above, rule (3), has been slightly modified from its form as rule (I) in the first list; instead of saying that no probability mass can be committed to $\Lambda_{\mathfrak{A}}$, I now say that only the null probability mass can be so committed. And I have added a new rule, (2c), which says that the null probability mass is committed to any proposition. This is a harmless convention, and it rounds out the mathematical picture.

Both for reasons of euphony and for intuitive reasons that will emerge later, I will usually read "M ct A" as "M is constrained to A" rather than as "M is committed to A." And I will call a binary relation "ct" between a Boolean algebra of probability masses $\mathcal{M}$ and a Boolean algebra $\mathfrak{A}$ a <u>constraint relation</u> if it satisfies the three conditions just listed.

Thus far, I have argued that our collection $\mathcal{M}$ of probability masses should have the structure of a Boolean algebra, but it also has a further structure: every probability mass $M \in \mathcal{M}$ has a <u>measure</u>. We need, evidently, a function $\mu: \mathcal{M} \rightarrow [0, 1]$ that assigns to each element M its measure $\mu(M)$.

<u>Definition</u>. If $\mathcal{M}$ is a Boolean algebra, then a function $\mu: \mathcal{M} \rightarrow [0, 1]$ is a <u>measure</u> if

(1) $\mu(\Lambda_{\mathcal{M}}) = 0$,

(2) $\mu(V_{\mathcal{M}}) = 1$,

and (3) $\mu(M_1) + \mu(M_2) = \mu(M_1 \vee M_2)$ whenever $M_1, M_2 \in \mathcal{M}$ and $M_1 \wedge M_2 = \Lambda_{\mathcal{M}}$.

If $\mathcal{M}$ is a Boolean algebra and $\mu: \mathcal{M} \rightarrow [0, 1]$ is a measure, then the pair $(\mathcal{M}, \mu)$ is a <u>Boolean algebra of probability masses</u>, or a <u>measure algebra</u>.

Conditions (1) - (3) in this definition should be intuitively evident. Formally, they are the same conditions as those used in the previous chapter to define a "probability function" on a Boolean algebra. Hence a measure will have all the same properties as a probability function.

It may occur to the reader that the preceding definition of a measure algebra does not capture all the properties that we might intuitively ascribe to the idealized substance that represents our probability. The definition does not exclude, for example, the possibility that a probability mass M not equal to $\Lambda_{\mathcal{M}}$ might have $\mu(M) = 0$; yet intuitively a probability mass M ought always to have positive measure unless it contains no probability at all and hence is equal to $\Lambda_{\mathcal{M}}$. Another inadequacy of the present definition is the lack of any requirement of "additivity" for the measures of infinite disjoint collections of probability masses. Later we will find that we can impose further conditions on measure algebras so as to correct these inadequacies. The present definition, though, will serve us well in this chapter.

2. Allocations of Probability

The mathematical notion of a constraint relation still does not quite do full justice to the intuitive picture that I used to derive the axioms for degrees of belief in Chapter 1. For in that derivation I spoke repeatedly of the "total portion of belief associated with a given proposition." In the present vocabulary, this would be the total probability mass constrained to the proposition; and it is not clear how this "total probability mass" can be identified in terms of the constraint relation.

Intuitively, the "total probability mass" constrained to proposition $A \epsilon \mathcal{A}$ would be a probability mass $M \epsilon \mathcal{M}$ with the properties (i) M ct A and (ii) if $M' \epsilon \mathcal{M}$, then M' ct A if and only if $M' \leq M$. But unfortunately, nothing in our mathematical definition of a constraint relation requires the existence of such a probability mass M for each proposition A.

We need, then, to insist that such a probability mass $M \epsilon \mathcal{M}$ should exist for each $A \epsilon \mathcal{A}$. The natural way to do this is to postulate the existence of a mapping $\rho : \mathcal{A} \rightarrow \mathcal{M}$ that assigns the appropriate M to each A. The constraint relation ct can then be defined in terms of the mapping ρ.

What properties should the mapping ρ have? As it turns out, the essential properties of ρ are those determined by the facts that (i) No probability mass except $\Lambda_{\mathcal{M}}$ is constrained to $\Lambda_{\mathcal{A}}$, (ii) All the probability (i.e., $V_{\mathcal{M}}$) is constrained to $V_{\mathcal{A}}$, and (iii) The total probability mass constrained to $A_1 \wedge A_2$ consists precisely of the intersection of the total probability mass constrained to A_1 and the total probability mass constrained to A_2.

<u>Definition</u>. A mapping $\rho : \mathcal{A} \rightarrow \mathcal{M}$ from a Boolean algebra of propositions $\mathcal{A}$ to a Boolean algebra of probability masses $\mathcal{M}$ is an <u>allocation of probability</u> if ρ satisfies the following three conditions:

(i) $\rho(\Lambda_{\mathcal{A}}) = \Lambda_{\mathcal{M}}$,

(ii) $\rho(V_{\mathcal{A}}) = V_{\mathcal{M}}$,

(iii) $\rho(A_1 \wedge A_2) = \rho(A_1) \wedge \rho(A_2)$ for all $A_1, A_2 \epsilon \mathcal{A}$.

Since $\rho(A)$ is the total probability mass constrained to A, a given element $M \epsilon \mathcal{M}$ should be constrained to A if and only if $M \leq \rho(A)$. It is easily demonstrated that a binary relation defined in this way actually is a constraint relation. I will refer to it as the constraint relation given or specified by ρ.

Theorem. Suppose $\rho:\mathcal{A} \to \mathcal{M}$ is an allocation of probability. Then the binary relation "ct" between $\mathcal{M}$ and $\mathcal{A}$ defined by "M ct A if and only if $M \subseteq \rho(A)$" is a constraint relation.

Proof: It is necessary to establish conditions (1), (2) and (3) in the definition of a constraint relation. Condition (2) is immediate; and the others are implied by the three conditions in the definition of an allocation: (i) implies (3), (ii) implies (1c) and (iii) implies (1a) and (1b).

It should be reiterated that not every constraint relation is specified by an allocation. But when there does exist an allocation specifying a given constraint relation, that allocation is unique.

Since $\rho(A)$ represents the total portion of our probability that is associated with the proposition A, its measure $\mu(\rho(A))$ ought to be our degree of belief in A. Hence the function $\mu \circ \rho:\mathcal{A} \to [1, 0]$ gives our degree of belief in the various propositions in $\mathcal{A}$. Will $\mu \circ \rho$ always be a belief function? It certainly ought to be, for the notion of an allocation of probability is a mathematical abstraction of the very intuitive picture that I used in deriving the axioms for belief functions.

Theorem. Suppose $(\mathcal{M}, \mu)$ is a Boolean algebra of probability masses, $\mathcal{A}$ is a Boolean algebra of propositions, and $\rho:\mathcal{A} \to \mathcal{M}$ is an allocation of probability. Then $\mu \circ \rho$ is a belief function on $\mathcal{A}$.

Proof: (i) $(\mu\circ\rho)(\Lambda_{\mathcal{Q}}) = \mu(\Lambda_{\mathcal{M}}) = 0.$

(ii) $(\mu\circ\rho)(V_{\mathcal{Q}}) = \mu(V_{\mathcal{M}}) = 1.$

(iii) Mathematically, the function μ qualifies as a probability function on $\mathcal{M}$. Hence, according to section 5 of the preceding chapter, μ itself satisfies the inequalities for belief functions with equality. And it is a simple consequence of the definition of an allocation (cf. Chapter 3, section 3) that $\rho(A_1) \leq \rho(A_2)$ whenever $A_1 \leq A_2$. Similarly, $\mu(M_1) \leq \mu(M_2)$ whenever $M_1 \leq M_2$. Hence for any elements $A_1, \ldots, A_n \epsilon \mathcal{Q}$, $\rho(A_1) \vee \ldots \vee \rho(A_n) \leq \rho(A_1 \vee \ldots \vee A_n)$, and

$$\mu\circ\rho(A_1 \vee \ldots \vee A_n) \geq \mu(\rho(A_1) \vee \ldots \vee \rho(A_n))$$

$$= \sum_i \mu(\rho(A_i)) - \sum_{i<j} \mu(\rho(A_i)\wedge\rho(A_j)) + \ldots + (-1)^{n+1}\mu(\rho(A_1)\wedge\ldots\wedge\rho(A_n))$$

$$= \sum_i \mu\circ\rho(A_i) - \sum_{i<j} \mu\circ\rho(A_i\wedge A_j) + \ldots + (-1)^{n+1}\mu\circ\rho(A_1\wedge\ldots\wedge A_n).$$

▨

Since it provides a mathematical representation for the intuitive picture underlying belief functions, the notion of an allocation of probability is the mathematical core of the theory of partial belief presented in this essay. In the bulk of this theory, the notion of an allocation will in fact be taken as basic. This seems to me to be appropriate, but it throws into question the adequacy of our axioms for degrees of belief. For it might be that some functions satisfying those axioms could not be represented by an allocation of probability. In fact, the axioms are adequate, and there are no such functions. In other words, if $\mathcal{Q}$ is a

Boolean algebra and Bel: $\mathcal{A} \to [0, 1]$ is a belief function, then there must exist a Boolean algebra of probability masses $(\mathcal{M}, \mu)$ and an allocation of probability $\rho: \mathcal{A} \to \mathcal{M}$ such that Bel $= \mu \circ \rho$. Most of the rest of this chapter is devoted to the proof of this fact.

3. Four Examples of Allocations

The simplest way to prove that a function Bel on a Boolean algebra $\mathcal{A}$ is a belief function is usually to construct an allocation that represents it. In this section I will provide such constructions for the examples of belief functions that were given in Chapter 1.

A. The Vacuous Belief Function

Recall that the vacuous belief function on a Boolean algebra $\mathcal{A}$ is given by

$$\mathrm{Bel}(A) = \begin{cases} 0 & \text{if } A \neq V_{\mathcal{A}} \\ 1 & \text{if } A = V_{\mathcal{A}}. \end{cases}$$

In order to represent this belief function, we construct a two-element measure algebra $\mathcal{M} = \{\Lambda_{\mathcal{M}}, V_{\mathcal{M}}\}$, with $\mu(\Lambda_{\mathcal{M}}) = 0$ and $\mu(V_{\mathcal{M}}) = 1$. We then define an allocation $\rho: \mathcal{A} \to \mathcal{M}$ by

$$\rho(A) = \begin{cases} \Lambda_{\mathcal{M}} & \text{if } A \neq V_{\mathcal{A}} \\ V_{\mathcal{M}} & \text{if } A = V_{\mathcal{A}}. \end{cases}$$

It is easily verified that $(\mathcal{M}, \mu)$ is a measure algebra, that ρ is an allocation of probability, and that Bel $= \mu \circ \rho$. The construction can be described intuitively, of course, by saying that all one's probability is committed to $\mathcal{V}_{\mathcal{Q}}$, while none of one's probability is committed to any other proposition in $\mathcal{Q}$.

B. Belief Functions on a Four-Element Boolean Algebra

In Chapter 1, we saw that when $\mathcal{Q} = \{\Lambda, A, \overline{A}, \mathcal{V}\}$, any function $Bel: \mathcal{Q} \rightarrow [0, 1]$ satisfying $Bel(\Lambda) = 0$, $Bel(\mathcal{V}) = 1$ and $Bel(A) + Bel(\overline{A}) \leq 1$ is a belief function. In order to represent such a belief function, we require in general an eight-element measure algebra $\mathcal{M}$.

Suppose, indeed, that $Bel(A) = \alpha_1$ and $Bel(\overline{A}) = \alpha_2$; $\alpha_1 + \alpha_2 \leq 1$. Then we can construct $\mathcal{M}$ by postulating first that $\mathcal{M}$ contains disjoint probability mass M_1, M_2 and M_3 with measures α_1, α_2 and $1 - \alpha_1 - \alpha_2$ respectively, and then including all the unions of pairs of these three. More explicitly, say that $\mathcal{M}$ consists of:

$$
\begin{array}{lll}
M_0 = \Lambda_{\mathcal{M}} & \text{with } \mu(M_0) = 0, \\
M_1 & \text{with } \mu(M_1) = \alpha_1, \\
M_2 & \text{with } \mu(M_2) = \alpha_2, \\
M_3 & \text{with } \mu(M_3) = 1 - \alpha_1 - \alpha_2, \\
M_4 = M_1 \vee M_2 & \text{with } \mu(M_4) = \alpha_1 + \alpha_2 \\
M_5 = M_1 \vee M_3 & \text{with } \mu(M_5) = 1 - \alpha_2 \\
M_6 = M_2 \vee M_3 & \text{with } \mu(M_6) = 1 - \alpha_1 \\
M_7 = M_1 \vee M_2 \vee M_3 = \mathcal{V}_{\mathcal{M}} & \text{with } \mu(M_7) = 1.
\end{array}
$$

Intuitively, $\mathcal{M}$ consists of all the probability masses that can be constructed from the three "basic probability masses," M_1, M_2 and M_3.

The allocation $\rho: \mathcal{Q} \rightarrow \mathcal{M}$ is given, of course, by $\rho(\Lambda_{\mathcal{Q}}) = \Lambda_{\mathcal{M}}$, $\rho(A) = M_1$, $\rho(\overline{A}) = M_2$, and $\rho(\mathcal{V}_{\mathcal{Q}}) = \mathcal{V}_{\mathcal{M}}$. Evidently, $Bel = \mu \circ \rho$.

C. The Senate Example

The measure algebra in this example is easy to describe intuitively: there are eleven disjoint basic probability masses, each with measure 1/11. It would be a bit tedious, though, to enumerate all the probability masses, for there are $2^{11} = 2,048$ of them.

Suppose we number the States shown in Figure 1 of Chapter 1 in the order they are shown there -- New Hampshire being number 1 and New York being number 11. Then we can suppose that our i'th basic probability mass, M_i, corresponds to the i'th State. The allocation $\rho: \mathcal{A} \rightarrow \mathcal{M}$ can then be described by saying that it maps the proposition A = "The Senator chosen will be in the subset A of the twenty-two Senators" into the probability mass formed by the union of all the basic probability masses corresponding to States both of whose Senators are in A. If there are k such states, the measure of $\rho(A)$ will be k/11.

D. The Kansas Example

For this example, we need six basic probability masses and $2^6 = 128$ probability masses in $\mathcal{M}$ altogether. Five of the basic probability masses, say, $M_1, \ldots, M_5$, have measure 1/7, while a sixth, say, M_6, has measure 2/7. The allocation $\rho: \mathcal{A} \rightarrow \mathcal{M}$ can be described by saying that $\rho(\mathrm{V}_{\mathcal{A}}) = \mathrm{V}_{\mathcal{M}} = M_1 \vee \ldots \vee M_6$, whereas if R is a <u>proper</u> subset of Kansas, ρ maps the proposition "The base will be located in R" into the probability mass consisting of the union of those probability masses M_i (i between one and five) such that the i'th Congressional district lies within R.

4. The Allowment of Probability

Let us pause to describe the upper probability function P^* in terms of the allocation ρ. Whereas Bel(A) can be understood as the measure of the total probability mass that is constrained to A, $P^*(A)$ can be understood as the measure of the total probability mass that is not constrained <u>away from</u> A. For $\rho(\overline{A})$ is the total probability mass that is constrained to $\overline{A}$, i.e., away from A; and its complement $\overline{\rho(\overline{A})}$ is therefore the total probability mass that is not constrained away from A. And $\mu(\overline{\rho(\overline{A})}) = 1 - \mu(\rho(\overline{A})) = 1 - \mathrm{Bel}(\overline{A}) = P^*(A)$.

Let $\zeta: \mathfrak{A} \to \mathfrak{M}$ be the mapping defined by $\zeta(A) = \overline{\rho(\overline{A})}$. Then $P^* = \mu \circ \zeta$. In the sequel we will often be interested in the upper probabilities of propositions and hence in the mapping ζ. Since $\zeta(A)$ can be described as the total probability mass that can be allowed to A, I will call ζ an <u>allowment</u>.

<u>Definition.</u> Suppose $\rho: \mathfrak{A} \to \mathfrak{M}$ is an allocation of probability. Then the mapping $\zeta: \mathfrak{A} \to \mathfrak{M} : A \rightsquigarrow \overline{\rho(\overline{A})}$ will be called the <u>allowment of probability corresponding to</u> ρ.

<u>Theorem.</u> Suppose $\zeta: \mathfrak{A} \to \mathfrak{M}$ is an allowment of probability. Then

(i) $\zeta(\Lambda_{\mathfrak{A}}) = \Lambda_{\mathfrak{M}}$.

(ii) $\zeta(V_{\mathfrak{A}}) = V_{\mathfrak{M}}$.

(iii) $\zeta(A_1 \vee A_2) = \zeta(A_1) \vee \zeta(A_2)$ for all $A_1, A_2 \in \mathfrak{A}$.

<u>Proof.</u> (i) $\zeta(\Lambda_{\mathfrak{A}}) = \overline{\rho(\overline{\Lambda_{\mathfrak{A}}})} = \overline{\rho(V_{\mathfrak{A}})} = \overline{\top_{\mathfrak{M}}} = \bot_{\mathfrak{M}}$.

(ii) $\zeta(\top_{\mathfrak{A}}) = \overline{\rho(\overline{\top_{\mathfrak{A}}})} = \overline{\rho(\bot_{\mathfrak{A}})} = \overline{\bot_{\mathfrak{M}}} = \top_{\mathfrak{M}}$.

(iii) $\zeta(A_1 \vee A_2) = \overline{\rho(\overline{A_1 \vee A_2})} = \overline{\rho(\overline{A_1} \wedge \overline{A_2})} = \overline{\rho(\overline{A_1}) \wedge \rho(\overline{A_2})}$

$= \overline{\rho(\overline{A_1})} \vee \overline{\rho(\overline{A_2})} = \zeta(A_1) \vee \zeta(A_2)$. ▨

5. Some Simple Consequences of the Axioms

Our present task is to justify the claim that the axioms for degrees of belief actually force conformity with the intuitive picture involving allocations of probability. Our first step will be to explore some of the immediate consequences of those axioms.

First, let us verify that a belief function Bel: $\mathcal{Q} \to [0, 1]$ does indeed obey the rule of monotonicity, i. e., that it satisfies $\mathrm{Bel}(A) \leq \mathrm{Bel}(B)$ whenever $A, B \in \mathcal{Q}$ and $A \leq B$. To do so, we need only substitute A and B-A for A_1 and A_2 in axiom III for n = 2, obtaining

$$\mathrm{Bel}\,(A \vee (B-A)) \geq \mathrm{Bel}\,(A) + \mathrm{Bel}\,(B-A) - \mathrm{Bel}\,(A \wedge (B-A)),$$

or
$$\mathrm{Bel}(B) \geq \mathrm{Bel}(A) + \mathrm{Bel}(B-A).$$

Secondly, let us investigate in detail the quantities

$$\beta(A_1, \ldots, A_n) = \sum_i \mathrm{Bel}(A_i) - \sum_{i<j} \mathrm{Bel}(A_i \wedge A_j) + \ldots + (-1)^{n+1} \mathrm{Bel}(A_1 \wedge \ldots \wedge A_n)$$

for various collections $\{A_1, \ldots, A_n\}$ of elements of $\mathcal{Q}$. Obviously, $\beta(A_1, \ldots, A_n)$ depends only on the collection $\{A_1, \ldots, A_n\}$, and not on the order of the A_i. According to our intuitive interpretation, $\beta(A_1, \ldots, A_n)$ should measure the total probability that is constrained to at least one of the A_i, and one can easily adduce many conditions that the quantities $\beta(A_1, \ldots, A_n)$ should satisfy if they are to conform to this intuitive interpretation. For example, they will have to satisfy

$$\beta(A_1, \ldots, A_n) \leq \beta(A_1, \ldots, A_{n+1}) \qquad (1)$$

for all collections $\{A_1, \ldots, A_{n+1}\} \subset \mathcal{Q}$.

Actually, (1) is easily deduced from the formula

$$\beta(A_1, \ldots, A_{n+1}) = \beta(A_1, \ldots, A_n) + \beta(A_{n+1}) - \beta(A_1 \wedge A_{n+1}, \ldots, A_n \wedge A_{n+1}), \tag{2}$$

which in turn follows from a simple calculation:

$$\begin{aligned}
\beta(A_1, \ldots, A_{n+1}) &= \sum_i \mathrm{Bel}(A_i) - \sum_{i<j} \mathrm{Bel}(A_i \wedge A_j) + \sum_{i<j<k} \mathrm{Bel}(A_i \wedge A_j \wedge A_k) - + \ldots \\
&= \Big(\sum_{i \le n} \mathrm{Bel}(A_i) - \sum_{i<j\le n} \mathrm{Bel}(A_i \wedge A_j) + \sum_{i<j<k\le n} \mathrm{Bel}(A_i \wedge A_j \wedge A_k) - + \ldots \Big) \\
&\quad + \Big(\mathrm{Bel}(A_{n+1}) - \sum_{i \le n} \mathrm{Bel}(A_i \wedge A_{n+1}) + \sum_{i<j\le n} \mathrm{Bel}(A_i \wedge A_j \wedge A_{n+1}) - + \ldots \Big) \\
&= \beta(A_1, \ldots, A_n) + \beta(A_{n+1}) - \beta(A_1 \wedge A_{n+1}, \ldots, A_n \wedge A_{n+1}).
\end{aligned}$$

To deduce (1) from (2), we need only use the rule of monotonicity and axiom III to conclude that

$$\beta(A_{n+1}) = \mathrm{Bel}(A_{n+1}) \geq \mathrm{Bel}((A_1 \wedge A_{n+1}) \vee \ldots \vee (A_n \wedge A_{n+1})) \geq \beta(A_1 \wedge A_{n+1}, \ldots, A_n \wedge A_{n+1}).$$

Of course, formula (2) itself has a simple intuitive interpretation; it says that the measure of the probability constrained to one of the A_i, $i = 1, \ldots, n+1$, is equal to the measure of the probability constrained to one of the first n A_i plus the measure of the probability constrained to A_{n+1}, less the measure of that probability which is constrained to both A_{n+1} and one of the first n A_i and thus is counted twice.

If the element A_{n+1} were actually a subelement of one of the elements $A_1, \ldots, A_n$, say, $A_{n+1} \leq A_n$, then any probability constrained to A_{n+1} would already be constrained to A_n, and it would seem that equality should hold in (1). This is obviously true for $n = 1$, for if $A_2 \leq A_1$, then

$$\beta(A_1, A_2) = \mathrm{Bel}(A_1) + \mathrm{Bel}(A_2) - \mathrm{Bel}(A_1 \wedge A_2)$$

$$= \mathrm{Bel}(A_1) + \mathrm{Bel}(A_2) - \mathrm{Bel}(A_2)$$

$$= \beta(A_1).$$

And it follows for larger values of n by induction: if it is true for $n \leq k-1$, and $A_1, \ldots, A_{k+1} \in \mathcal{U}$ and $A_{k+1} \leq A_k$, then $A_i \wedge A_{k+1} \leq A_k \wedge A_{k+1}$ for $i = 1, \ldots, k$, and

$$\beta(A_{k+1}) = \beta(A_k \wedge A_{k+1}) = \beta(A_{k-1} \wedge A_{k+1}, A_k \wedge A_{k+1})$$

$$= \beta(A_{k-2} \wedge A_{k+1}, A_{k-1} \wedge A_{k+1}, A_k \wedge A_{k+1})$$

$$= \ldots = \beta(A_1 \wedge A_{k+1}, \ldots, A_k \wedge A_{k+1}),$$

and from (2) it follows that

$$\beta(A_1, \ldots, A_{k+1}) = \beta(A_1, \ldots, A_k).$$

It follows from (1) that whenever $\{A_1, \ldots, A_n\} \subset \{B_1, \ldots, B_m\} \subset \mathcal{Q}$, $\beta(A_1, \ldots, A_n) \leq \beta(B_1, \ldots, B_m)$. Actually this inequality will hold even when $\{A_1, \ldots, A_n\}$ is not contained in $\{B_1, \ldots, B_m\}$ provided that for each $A_i \in \{A_1, \ldots, A_n\}$ there is a $B_j \in \{B_1, \ldots, B_m\}$ such that $A_i \leq B_j$. For if the A_i are subelements of the B_j in this fashion, then it follows from the preceding paragraph that $\beta(B_1, \ldots, B_m) = \beta(B_1, \ldots, B_m, A_1, \ldots, A_n)$, and since $\{A_1, \ldots, A_n\} \subset \{B_1, \ldots, B_m, A_1, \ldots, A_n\}$, $\beta(A_1, \ldots, A_n) \leq \beta(B_1, \ldots, B_m, A_1, \ldots, A_n)$.

If two collections $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_m\}$ are related in the fashion just described, i.e., if for each A_i there is a B_j such that

$A_i \leq B_j$, then it is convenient to say that $\{B_1, \ldots, B_m\}$ <u>majorizes</u> $\{A_1, \ldots, A_n\}$. In this vocabulary, the assertion of the preceding paragraph is simply that $\beta(A_1, \ldots, A_n) \leq \beta(B_1, \ldots, B_m)$ whenever $\{A_1, \ldots, A_n\}$ is majorized by $\{B_1, \ldots, B_m\}$. Similarly, $\beta(A_1, \ldots, A_n) = \beta(B_1, \ldots, B_m)$ whenever $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_m\}$ majorize each other.

The following proposition may strike the reader as a bit too technical to provide any further insight into belief functions, but it will be useful to us later.

<u>Theorem</u>. Suppose $\{A_1, \ldots, A_n\}$, $\{B_1, \ldots, B_m\}$ and $\{C_1, \ldots, C_k\}$ are finite subsets of $\mathcal{Q}$. Suppose further that $\{B_1, \ldots, B_m\}$ majorizes $\{A_1, \ldots, A_n\}$, and that $\{A_1, \ldots, A_n\}$ majorizes $\{B_1 \wedge C_i, \ldots, B_m \wedge C_i\}$ for each i, i = 1, ..., k. Then

$$\beta(B_1, \ldots, B_m) - \beta(A_1, \ldots, A_n) = \beta(B_1, \ldots, B_m, C_1, \ldots, C_k) - \beta(A_1, \ldots, A_n, C_1, \ldots, C_k).$$

<u>Proof</u>. It suffices to prove the proposition for k = 1, i.e., for the case where $\{C_1, \ldots, C_k\} = \{C\}$. By (2),

$$\beta(B_1, \ldots, B_m, C) = \beta(B_1, \ldots, B_m) + \beta(C) - \beta(B_1 \wedge C, \ldots, B_m \wedge C),$$

$$\beta(A_1, \ldots, A_n, C) = \beta(A_1, \ldots, A_n) + \beta(C) - \beta(A_1 \wedge C, \ldots, A_n \wedge C).$$

Subtraction of the second equation from the first gives the desired result provided that

$$\beta(B_1 \wedge C, \ldots, B_m \wedge C) = \beta(A_1 \wedge C, \ldots, A_n \wedge C).$$

But this equation does hold, for it follows from the hypotheses of the theorem that $\{A_1 \wedge C, \ldots, A_n \wedge C\}$ and $\{B_1 \wedge C, \ldots, B_m \wedge C\}$ majorize each other.

▨

6. The Representation Theorem: Finite Case

Theorem. Suppose $\mathcal{Q}$ is a finite Boolean algebra and Bel: $\mathcal{Q} \rightarrow [0, 1]$ is a belief function. Then there exists a Boolean algebra of probability masses $(\mathcal{M}, \mu)$ and an allocation of probability $\rho: \mathcal{Q} \rightarrow \mathcal{M}$ such that $\mathrm{Bel} = \mu \circ \rho$.

In order to prove this theorem, I will construct the measure algebra $\mathcal{M}$ as a field of subsets. (See Chapter 3, section 6) More precisely, I will take $\mathcal{M}$ to be the field of all subsets of $\mathcal{J} = \mathcal{Q} - \{\bot_{\mathcal{Q}}\}$, and define a constraint relation between $\mathcal{M}$ and $\mathcal{Q}$ by saying that M is constrained to A if and only if $A' \leq A$ for each $A' \in M$. This is indeed a constraint relation, and it is given by the allocation $\rho: \mathcal{Q} \rightarrow \mathcal{M}: A \rightsquigarrow \{A' \mid A' \leq A,\ A' \neq \bot_{\mathcal{Q}}\}$.

In order to define the measure μ on $\mathcal{M}$, first define the basic probability number m_A for each $A \in \mathcal{J}$ by

$$m_A = \mathrm{Bel}(A) - \beta(A_1, \ldots, A_n),$$

where $A_1, \ldots, A_n$ are all the proper subelements of A, and β is defined as in section 5. The function μ is then defined by

$$\mu(M) = \sum_{A \in M} m_A .$$

Since the quantities m_A are non-negative, μ is evidently non-negative and additive; in order to show that μ is a measure on $\mathcal{M}$, it therefore suffices to show that $\mu(\mathcal{J}) = 1$. This, however, is merely a special case of the relation $\mathrm{Bel}(A) = \mu(\rho(A))$, which we need to establish in general.

In order to verify that $\mathrm{Bel}(A) = \mu(\rho(A))$, it is convenient to appeal to the fact that $\mathcal{Q}$ is isomorphic to the field of all subsets of the set $\mathcal{S}$ of atomic propositions of $\mathcal{Q}$. (See Chapter 3, section 6.) Thinking of each element A of $\mathcal{A}$ as a subset of $\mathcal{S}$, let c(A) denote its cardinality, and set $\mathrm{Par}(A) = (-1)^{c(A)}$. In other words, the parity of A is taken to be +1 if A has an even number of elements and -1 if A has an odd number of elements. Considering a fixed non-zero element A of $\mathcal{A}$, denote as before by $A_1, \ldots, A_n$ the proper subelements of A. Now, in general $n = 2^{c(A)} - 1$. Exactly c(A) of the elements $A_1, \ldots, A_n$, on the other hand, will obey $c(A_i) = c(A) - 1$; if we suppose that these are the first c(A), then $\{A_1, \ldots, A_n\}$ is majorized by $\{A_1, \ldots, A_{c(A)}\}$. Hence $\beta(A_1, \ldots, A_n) = \beta(A_1, \ldots, A_{c(A)})$. Now

$$\beta(A_1, \ldots, A_{c(A)}) = \sum_{i<c(A)} \mathrm{Bel}(A_i) - \sum_{i<j\leq c(A)} \mathrm{Bel}(A_i \cap A_j) + - \ldots + \\ + (-1)^{c(A)+1} \mathrm{Bel}(A_1 \cap \ldots \cap A_{c(A)})$$

and it is easily seen that for each i, i = 1, ..., n $\mathrm{Bel}(A_i)$ occurs exactly once in the right-hand side of this equation, with sign equal to $\mathrm{Par}(A - A_i) = \mathrm{Par}(A) \cdot \mathrm{Par}(A_i)$. Hence

$$\beta(A_1, \ldots, A_{c(A)}) = -\mathrm{Par}(A) \sum_{A' < A,\, A' \neq A} \mathrm{Bel}(A')\,\mathrm{Par}(A'),$$

and

$$m_A = \mathrm{Bel}(A) - \beta(A_1, \ldots, A_{c(A)}) = \mathrm{Par}(A) \sum_{A' \leq A} \mathrm{Bel}(A')\,\mathrm{Par}(A').$$

With this expression for m_A, it is easy to verify that

$\mathrm{Bel}(A) = \mu(\rho(A))$: Setting $m_\Lambda = 0$, we can write

$$\begin{aligned} \mu(\rho(A)) &= \mu(\{A' \mid A' \subset A,\ A' \neq \Lambda\}) \\ &= \sum_{A' \leq A,\ A' \neq \Lambda} m_{A'} = \sum_{A' \leq A} m_{A'} \\ &= \sum_{A' \leq A} \mathrm{Par}(A') \Big(\sum_{A'' \leq A'} \mathrm{Bel}(A'')\,\mathrm{Par}(A'') \Big) \end{aligned}$$

$$= \sum_{A''\lneq A} \mathrm{Bel}(A'')\ \mathrm{Par}(A'')\ \Big(\sum_{A''\lneq A'\lneq A} \mathrm{Par}(A')\Big).$$

But

$$\sum_{A''\lneq A'\lneq A} \mathrm{Par}(A') = (1-1)^{c(A'')-c(A)}\ \mathrm{Par}(A) = \begin{cases} 0 & \text{if } A \neq A'' \\ \mathrm{Par}(A) & \text{if } A = A'' \end{cases}$$

Hence

$$\mu(\rho(A)) = \mathrm{Bel}(A).$$

7. Measures on Semifields of Subsets

In order to prove our representation theorem in the general case, we need to know how to extend a measure from a semifield to a field of subsets. The exposition in this section is adapted from Kolmogorov and Fomin, pp. 17-22.

Definition. A non-empty collection $\mathcal{E}$ of subsets of a non-empty set $\mathcal{S}$ is called a semifield of subsets of $\mathcal{S}$ if it satisfies the following conditions:

(i) $\mathcal{E}$ contains the empty set $\emptyset$.

(ii) $\mathcal{E}$ contains the set $\mathcal{S}$ itself.

(iii) If $A, B \in \mathcal{E}$, then $A \cap B \in \mathcal{E}$.

(iv) If A and $A_1 \subset A$ are both elements of $\mathcal{E}$, then

$$A = \bigcup_{i=1}^{n} A_i,$$

where the sets A_i are pairwise disjoint elements of $\mathcal{E}$, and the first of the sets A_i is the given set A_1.

The following example will make condition (iv) more intuitively accessible: Let $\mathcal{S}$ be a rectangle in the plane whose sides are parallel to the coordinate axes and let $\mathcal{E}$ consist of the empty set $\emptyset$ together with all the rectangles that are contained in $\mathcal{S}$ and whose sides are also parallel to the coordinate axes. Then $\mathcal{E}$ will be a semifield of subsets of $\mathcal{S}$.

Suppose that $\mathcal{E}$ is a semifield of subsets of $\mathcal{S}$, and denote by $\mathcal{F}$ the collection of subsets of $\mathcal{S}$ of the form

$$A = \bigcup_{i=1}^{n} A_i,$$

where n is a positive integer, and the A_i are pairwise disjoint elements of $\mathcal{E}$. Then it is easily shown that $\mathcal{F}$ is a field, i.e., that $\mathcal{F}$ is closed under union, intersection and complementation. In order to show that $\mathcal{F}$ is closed under intersection, for example, note that if $A, B \in \mathcal{F}$, then $A = \bigcup_{i=1}^{n} A_i$ for some pairwise disjoint elements $A_1, \ldots, A_n$ of $\mathcal{E}$ and $B = \bigcup_{j=1}^{m} B_j$ for some pairwise disjoint elements $B_1, \ldots, B_m$ of $\mathcal{E}$. So

$$A \cap B = \left(\bigcup_{i=1}^{n} A_i\right) \cap \left(\bigcup_{j=1}^{m} B_j\right) = \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} (A_i \cap B_j).$$

But the $A_i \cap B_j$ are certainly pairwise disjoint and are all in $\mathcal{E}$, by clause (iii) of the definition of a semifield. Hence $A \cap B$ is an element of $\mathcal{F}$.

On the other hand, it is evident that any field of subsets of $\mathcal{S}$ that contains $\mathcal{E}$ must contain $\mathcal{F}$; hence $\mathcal{F}$ must be the smallest field of subsets of $\mathcal{S}$ containing $\mathcal{E}$, which is sometimes called the field of subsets of $\mathcal{S}$ generated by $\mathcal{E}$. We are led, therefore, to the following theorem.

Theorem. If $\mathcal{E}$ is a semifield of subsets of $\mathcal{S}$, then the field of subsets of $\mathcal{S}$ generated by $\mathcal{E}$ consists of those subsets of $\mathcal{S}$ that admit of a disjoint partition into elements of $\mathcal{E}$.

We are now prepared to attack the problem of extending a measure on $\mathcal{E}$.

Definition. A function $\mu: \to [0,\infty]$ on a semifield $\mathcal{E}$ of subsets of a set $\mathcal{S}$ is a measure if whenever

$$A = \bigcup_{i=1}^{n} A_i$$

is a finite partition of A, and A, A_1, ..., $A_n \in \mathcal{E}$,

$$\mu(A) = \sum_{i=1}^{n} \mu(A_i).$$

It is easily seen that if μ is a measure, then $\mu(\phi) = 0$. Hence if $\mathcal{E}$ is actually a field and $\mu(\mathcal{S}) = 1$, then the measure μ satisfies the usual rules: $\mu(\phi) = 0$, $\mu(\mathcal{S}) = 1$ and $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$ whenever $S_1 \cap S_2 = \phi$.

Theorem. If $\mathcal{E}$ is a semifield of subsets of $\mathcal{S}$, and $\mu: \mathcal{E} \to [0,\infty]$ is a measure, then μ has a unique extension to a measure on the field $\mathcal{F}$ generated by $\mathcal{E}$.

Proof: According to the preceding theorem, any element $A \in \mathcal{F}$ admits of a finite partition $A = \bigcup_i A_i$ into elements of $\mathcal{E}$. Define a function $\nu: \mathcal{F} \to [0,\infty]$ by

$$\nu(A) = \sum_i \mu(A_i).$$

In order to see that the value $\nu(A)$ is independent of the partition, notice that if $A = \bigcup_j B_j$ is another partition of A into elements of $\mathcal{E}$, then since $A_i = \bigcup_j (A_i \cap B_j)$ and $B_j = \bigcup_i (A_i \cap B_j)$ are partitions of the elements A_i and B_j into pairwise disjoint elements of $\mathcal{E}$,

$$\sum_i \mu(A_i) = \sum_i \mu\left(\bigcup_j (A_i \cap B_j)\right) = \sum_i \sum_j \mu(A_i \cap B_j)$$

$$= \sum_j \mu\left(\bigcup_i (A_i \cap B_j)\right) = \sum_j \mu(B_j).$$

The additivity of ν for elements of $\mathcal{F}$ is evident, so ν is indeed a measure on $\mathcal{F}$. To see that ν is the unique measure on $\mathcal{F}$ that extends μ, notice that if ν' were another measure on $\mathcal{F}$ that agreed with μ on $\mathcal{E}$, then ν' would have to satisfy

$$\nu'(A) = \sum_i \nu'(A_i) = \sum_i \nu(A_i) = \nu(A)$$

for any element A of $\mathcal{F}$ admitting a finite partition into elements A_i of $\mathcal{E}$. ▨

8. The Representation Theorem: General Case

Theorem. Suppose $\mathrm{Bel}: \mathcal{Q} \to [0, 1]$ is a belief function. Then there exists a measure algebra $(\mathcal{M}, \mu)$ and an allocation $\rho: \mathcal{Q} \to \mathcal{M}$ such that $\mathrm{Bel} = \mu \circ \rho$.

The rest of this section is devoted to the proof of this theorem. The corresponding theorem in the finite case was proven by constructing $\mathcal{M}$ as the field of all subsets of the set of non-zero elements of $\mathcal{Q}$. In the present proof, we will have to content ourselves with a smaller field of subsets of that set.

Let $\mathcal{J} = \mathcal{Q} - \{\Lambda_{\mathcal{Q}}\}$, and for each non-empty finite subset $\mathcal{C}$ of $\mathcal{Q}$ set

$$R(\mathcal{C}) = \{A \mid A \in \mathcal{J};\ A \leq C \text{ for some } C \in \mathcal{C}\} \subset \mathcal{J},$$

and set

$$\beta(\mathcal{C}) = \beta(A_1, \ldots, A_n),$$

where $A_1, \ldots, A_n$ are the elements of $\mathcal{C}$ and $\beta(A_1, \ldots, A_n)$ is defined as in section 5. If $\mathcal{C} = \phi$, set $R(\mathcal{C}) = \phi$ and $\beta(\mathcal{C}) = 0$. Set

$$\mathcal{R} = \{R(\mathcal{C}) \mid \mathcal{C} \subset \mathcal{Q};\ \mathcal{C} \text{ finite}\}.$$

There is a natural way to map $\mathcal{Q}$ into $\mathcal{R}$; one simply maps A to $R(\{A\})$. The strategy of this proof will be to develop this mapping into an allocation by extending $\mathcal{R}$ to a field of subsets of $\mathcal{J}$ and using the quantities $\beta(\mathcal{C})$ to define a measure on that field.

Throughout this proof, the letters $\mathcal{C}, \mathcal{D}, \mathcal{G}$ and $\mathcal{H}$ will always denote finite subsets of $\mathcal{Q}$, and the letters A, C and D will always denote elements of $\mathcal{Q}$. $\mathcal{C}_A$ will denote the finite subset of $\mathcal{Q}$ given by

$$\mathcal{C}_A = \{A \wedge C \mid C \in \mathcal{C}\},$$

and $\mathcal{C} \boxtimes \mathcal{D}$ will denote the finite subset of $\mathcal{Q}$ given by

$$\mathcal{C} \boxtimes \mathcal{D} = \{C \wedge D \mid C \in \mathcal{C}, D \in \mathcal{D}\}.$$

Evidently, $\mathcal{C} \boxtimes \mathcal{D} = \mathcal{D} \boxtimes \mathcal{C}$, and $\mathcal{C} \boxtimes \{A\} = \mathcal{C}_A$. Notice also that a distributive law holds for $\boxtimes$ and $\cup$:

$$(\mathcal{C} \cup \mathcal{D}) \boxtimes \mathcal{G} = (\mathcal{C} \boxtimes \mathcal{G}) \cup (\mathcal{D} \boxtimes \mathcal{G}).$$

Let us say that $\mathcal{C}$ majorizes $\mathcal{D}$ if for each non-zero $D \in \mathcal{D}$ there exists an element $C \in \mathcal{C}$ such that $D \leq C$. I will use the notation "$\mathcal{D} \alpha \mathcal{C}$" to indicate that $\mathcal{C}$ majorizes $\mathcal{D}$. The following facts follow from section 5:

(1) If $\mathcal{D} \alpha \mathcal{C}$, then $0 \leq \beta(\mathcal{D}) \leq \beta(\mathcal{C}) \leq 1$.

(2) $\beta(\mathcal{C} \cup \{A\}) = \beta(\mathcal{C}) + \beta(\{A\}) - \beta(\mathcal{C}_A)$.

(3) If $\mathcal{D} \alpha \mathcal{C}$ and $\mathcal{C}_A \alpha \mathcal{D}$ for each $A \in \mathcal{G}$, then

$$\beta(\mathcal{C} \cup \mathcal{G}) - \beta(\mathcal{D} \cup \mathcal{G}) = \beta(\mathcal{C}) - \beta(\mathcal{D}).$$

Now it is obvious that $R(\mathcal{D}) \subset R(\mathcal{C})$ if and only if $\mathcal{D} \alpha \mathcal{C}$. This implies in particular that if $R(\mathcal{D}) = R(\mathcal{C})$, then $\beta(\mathcal{D}) = \beta(\mathcal{C})$. So we can define a mapping

$$b_o : \mathcal{R} \longrightarrow [0, 1]$$

by setting

$$b_o(R(\mathcal{C})) = \beta(\mathcal{C}).$$

Now the collection $\mathcal{R}$ of subsets of $\mathcal{J}$ is closed under the operations of union and intersection. As a matter of fact,

$$R(\mathcal{C}) \cup R(\mathcal{D}) = R(\mathcal{C} \cup \mathcal{D})$$

and

$$R(\mathcal{C}) \cap R(\mathcal{D}) = R(\mathcal{C} \boxtimes \mathcal{D}).$$

Notice that these relations imply in particular that

$R(\mathcal{C}) - R(\mathcal{D}) = R(\mathcal{C} \cup \mathcal{D}) - R(\mathcal{D}) = R(\mathcal{C}) - R(\mathcal{C} \boxtimes \mathcal{D})$ for all finite subsets $\mathcal{C}, \mathcal{D}$ of $\mathcal{Q}$.

Our first step in enlarging $\mathcal{R}$ will be to include all differences. Set:

$$\mathcal{E} = \{R_1 - R_2 \mid R_1, R_2 \in \mathcal{R}\} = \{R(\mathcal{C}) - R(\mathcal{D}) \mid \mathcal{C}, \mathcal{D} \text{ are finite subsets of } \mathcal{Q}\}.$$

Notice that if $E = R(\mathcal{C}) - R(\mathcal{D})$ is in $\mathcal{E}$, then E can also be expressed in the form $E = R(\mathcal{C} \cup \mathcal{D}) - R(\mathcal{D})$. Hence every element of $\mathcal{E}$ is of the form $R_1 - R_2$ where $R_2 \subset R_1$, i.e., of the form $R(\mathcal{C}) - R(\mathcal{D})$, where $\mathcal{D} \alpha \mathcal{C}$.

Lemma 1. $\mathcal{E}$ is a semifield of subsets of $\mathcal{T}$.

Proof: (i) $\phi = R(\phi) - R(\phi)$ is in $\mathcal{E}$.

(ii) $\mathcal{T} = R(\{\nabla\}) - R(\phi)$ is in $\mathcal{E}$.

(iii) Suppose $E_1 = R(\mathcal{C}_1) - R(\mathcal{A}_1)$ and $E_2 = R(\mathcal{C}_2) - R(\mathcal{A}_2)$ are in $\mathcal{E}$.

Then

$$\begin{aligned} E_1 \cap E_2 &= (R(\mathcal{C}_1) \cap \overline{R(\mathcal{A}_1)}) \cap (R(\mathcal{C}_2) \cap \overline{R(\mathcal{A}_2)}) \\ &= (R(\mathcal{C}_1) \cap R(\mathcal{C}_2)) \cap \overline{R(\mathcal{A}_1) \cup R(\mathcal{A}_2)} \\ &= R(\mathcal{C}_1 \boxtimes \mathcal{C}_2) \cap \overline{R(\mathcal{A}_1 \cup \mathcal{A}_2)} \\ &= R(\mathcal{C}_1 \boxtimes \mathcal{C}_2) - R(\mathcal{A}_1 \cup \mathcal{A}_2) \end{aligned}$$

is in $\mathcal{E}$.

(iv) Suppose $E_1 = R(\mathcal{C}_1) - R(\mathcal{A}_1)$ and $E_2 = R(\mathcal{C}_2) - R(\mathcal{A}_2)$ are in $\mathcal{E}$, and $E_1 \subset E_2$. Then one may assume that $R(\mathcal{A}_1) \subset R(\mathcal{C}_1)$, so that

$$\overline{E_1} = \overline{R(\mathcal{C}_1) \cap \overline{R(\mathcal{A}_1)}} = \overline{R(\mathcal{C}_1)} \cup R(\mathcal{A}_1)$$

will be a disjoint partition of $\overline{E}_1$. Then

$$E_2 - E_1 = E_2 \cap \overline{E}_1 = (E_2 \cap \overline{R(\mathcal{C}_1)}) \cup (E_2 \cap R(\mathcal{A}_1))$$

will also be a disjoint partition. But

$$\begin{aligned} E_2 \cap \overline{R(\mathcal{C}_1)} &= R(\mathcal{C}_2) \cap \overline{R(\mathcal{A}_2)} \cap \overline{R(\mathcal{C}_1)} \\ &= R(\mathcal{C}_2) \cap \overline{R(\mathcal{A}_2 \cup \mathcal{C}_1)} \end{aligned}$$

is in $\mathcal{E}$, and

$$\begin{aligned} E_2 \cap R(\mathcal{A}_1) &= R(\mathcal{C}_2) \cap \overline{R(\mathcal{A}_2)} \cap R(\mathcal{A}_1) \\ &= R(\mathcal{C}_2 \boxtimes \mathcal{A}_1) \cap \overline{R(\mathcal{A}_2)} \end{aligned}$$

is in $\mathscr{E}$, so we have expressed $E_2 - E_1$ as a disjoint partition of elements of $\mathscr{E}$, as required. ▨

Lemma 2. If $\mathscr{A} \alpha \mathscr{C}$ and $(R(\mathscr{C}) - R(\mathscr{A})) \cap R(\mathscr{G}) = \phi$, then

$$\text{(i)} \quad R(\mathscr{C} \cup \mathscr{G}) - R(\mathscr{A} \cup \mathscr{G}) = R(\mathscr{C}) - R(\mathscr{A})$$

$$\text{and (ii)} \quad \beta(\mathscr{C} \cup \mathscr{G}) - \beta(\mathscr{A} \cup \mathscr{G}) = \beta(\mathscr{C}) - \beta(\mathscr{A}).$$

Proof. From the hypothesis it immediately follows that

$$(R(\mathscr{C}) \cup R(\mathscr{G})) - (R(\mathscr{A}) \cup R(\mathscr{G})) = R(\mathscr{C}) - R(\mathscr{A}),$$

whence (i). Consider now any elements $C \in \mathscr{C}$ and $A \in \mathscr{G}$. If $C \wedge A \neq \Lambda$, then $C \wedge A$ is in both $R(\mathscr{C})$ and $R(\mathscr{G})$, and hence must be in $R(\mathscr{A})$; hence there must exist an element $D \in \mathscr{A}$ such that $C \wedge A \leq D$. If $C \wedge A = \Lambda$, on the other hand, then $C \wedge A \leq D$ for any $D \in \mathscr{A}$. In any case, $\mathscr{C}_A \alpha \mathscr{A}$ for all $A \in \mathscr{G}$, and (ii) follows from (3) above. ▨

Lemma 3. If $\mathscr{A} \alpha \mathscr{C}$, $\mathscr{H} \alpha \mathscr{G}$, and

$$R(\mathscr{C}) - R(\mathscr{A}) = R(\mathscr{G}) - R(\mathscr{H}),$$

then
$$\beta(\mathscr{C}) - \beta(\mathscr{A}) = \beta(\mathscr{G}) - \beta(\mathscr{H}).$$

Proof: Since $R(\mathscr{H}) \cap (R(\mathscr{G}) - R(\mathscr{H})) = \phi$, the hypothesis of the lemma implies that $R(\mathscr{H}) \cap (R(\mathscr{C}) - R(\mathscr{A})) = \phi$. By Lemma 2,

$$\text{(i)} \quad R(\mathscr{C}) - R(\mathscr{A}) = R(\mathscr{C} \cup \mathscr{H}) - R(\mathscr{A} \cup \mathscr{H})$$

and
$$\text{(ii)} \quad \beta(\mathscr{C}) - \beta(\mathscr{A}) = \beta(\mathscr{C} \cup \mathscr{H}) - \beta(\mathscr{A} \cup \mathscr{H}).$$

Symmetrically,

$$\text{(iii)} \quad R(\mathscr{G}) - R(\mathscr{H}) = R(\mathscr{G} \cup \mathscr{A}) - R(\mathscr{H} \cup \mathscr{A})$$

and
$$\text{(iv)} \quad \beta(\mathscr{G}) - \beta(\mathscr{H}) = \beta(\mathscr{G} \cup \mathscr{A}) - \beta(\mathscr{H} \cup \mathscr{A}).$$

It follows from (i), (iii) and the hypothesis of the lemma that

$$R(\mathscr{G} \cup \mathscr{A}) - R(\mathscr{H} \cup \mathscr{A}) = R(\mathscr{C} \cup \mathscr{H}) - R(\mathscr{A} \cup \mathscr{H}),$$

whence $R(\mathscr{G} \cup \mathscr{A}) = R(\mathscr{C} \cup \mathscr{H})$. Hence $\beta(\mathscr{G} \cup \mathscr{A}) = \beta(\mathscr{C} \cup \mathscr{H})$. From (ii) and (iv) it then follows that

$$\beta(\mathcal{C}) - \beta(\mathcal{D}) = \beta(\mathcal{G}) - \beta(\mathcal{H}).$$

▨

<u>Lemma 4</u>. If $\mathcal{D} \,\alpha\, \mathcal{C}$ and $(R(\mathcal{C}) - R(\mathcal{D})) \subset R(\mathcal{G})$, then

(i) $R(\mathcal{C} \boxtimes \mathcal{G}) - R(\mathcal{D} \boxtimes \mathcal{G}) = R(\mathcal{C}) - R(\mathcal{D})$

and (ii) $\beta(\mathcal{C} \boxtimes \mathcal{G}) - \beta(\mathcal{D} \boxtimes \mathcal{G}) = \beta(\mathcal{C}) - \beta(\mathcal{D})$.

<u>Proof</u>: From the hypothesis it immediately follows that

$$(R(\mathcal{C}) \cap R(\mathcal{G})) - (R(\mathcal{C}) \cap R(\mathcal{D})) = R(\mathcal{C}) - R(\mathcal{D}),$$

whence (i). The second relation then follows by lemma 3. ▨

Every element of $\mathcal{E}$ can be written in the form $R(\mathcal{C}) - R(\mathcal{D})$, with $\mathcal{D} \,\alpha\, \mathcal{C}$; and according to lemma 3, $\beta(\mathcal{C}) - \beta(\mathcal{D})$ does not depend on the choice of $\mathcal{C}$ and $\mathcal{D}$. Hence a function $b: \mathcal{E} \to [0, 1]$ may be defined by setting $b(E) = \beta(\mathcal{C}) - \beta(\mathcal{D})$ when $E = R(\mathcal{C}) - R(\mathcal{D})$ and $\mathcal{D} \,\alpha\, \mathcal{C}$. The function b is obviously an extension of the function $b_o : \mathcal{R} \to [0, 1]$.

Now let $\mathcal{M}$ be the field of subsets of $\mathcal{J}'$ generated by $\mathcal{E}$, and define a mapping $\rho: \mathcal{A} \to \mathcal{M}$ by $\rho(A) = R(\{A\})$. It is easily verified that

(i) $\rho(\mathcal{V}) = R(\{\mathcal{V}\}) = \mathcal{J}'$.

(ii) $\rho(\Lambda) = R(\{\Lambda\}) = \phi$.

(iii) $\rho(A_1 \cap A_2) = R(\{A_1 \cap A_2\}) = R(\{A_1\} \boxtimes \{A_2\}) = R(\{A_1\}) \cap R(\{A_2\})$
$= \rho(A_1) \cap \rho(A_2)$.

Hence ρ is a non-singular allocation. Furthermore, for each $A \in \mathcal{A}$, $Bel(A) = \beta(\{A\}) = b(R(\{A\})) = b(\rho(A))$. Hence if b could be extended to a measure μ on $\mathcal{M}$, then Bel would be the belief function induced by the allocation ρ into the probability algebra $(\mathcal{M}, \mu)$, and the proof would be complete. But we learned in section 5 that b can be extended to a probability function μ on $\mathcal{M}$ provided that b is a measure on $\mathcal{E}$. Hence our only remaining task is to show that b is a measure.

In order to show that b is a measure, one must show that if $E = E_1 \cup \ldots \cup E_n$ is a disjoint partition of E and if E, $E_1, \ldots, E_n \in \mathcal{E}$, then $b(E) = \sum_{i=1}^{n} b(E_i)$. In order to carry out such a demonstration, let us fix E and the E_i and express them in the form

$$E = R(\mathcal{C}) - R(\mathcal{D})$$

and

$$E_i = R(\mathcal{C}_i) - R(\mathcal{D}_i)$$

for i = 1, ..., n. We may assume that $\mathcal{D} \alpha \mathcal{C}$ and $\mathcal{D}_i \alpha \mathcal{C}_i$ for i = 1, ..., n. We may also assume that $\mathcal{C}_i \alpha \mathcal{C}$ for i = 1, ..., n, for if this did not hold then the $\mathcal{C}_i$ and the $\mathcal{D}_i$ could be replaced by the sets $\mathcal{C}_i \boxtimes \mathcal{C}$ and $\mathcal{D}_i \boxtimes \mathcal{D}$, respectively.

Set

$$\mathcal{W} = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_n \cup \mathcal{D}_1 \cup \ldots \cup \mathcal{D}_n \cup \mathcal{C} \cup \mathcal{D},$$

and set

$$\mathcal{V} = \{A_1 \wedge \ldots \wedge A_n \mid n \geq 1 \text{ and } A_1, \ldots, A_n \in \mathcal{W}\}.$$

Then $\mathcal{V}$ is finite, $\mathcal{W} \subset \mathcal{V}$, $\mathcal{V} \alpha \mathcal{C}$, and $\mathcal{V}$ is closed under conjunctions, i.e., if A and A' are in $\mathcal{V}$, then $A \wedge A'$ is in $\mathcal{V}$. The partial ordering that $\mathcal{V}$ inherits from $\mathcal{A}$ (see Chapter 3, section 1) can be extended to a total ordering on $\mathcal{V}$, i.e., the elements of $\mathcal{V}$ can be indexed $V_1, \ldots, V_r$ so that if $V_s \leq V_t$, then $s \leq t$. Suppose this is done, and set $\mathcal{V}_s = \{V_t \mid t \leq s\}$ for s = 1, ..., r. Set $V_o = \phi$.

For each s, s = 1, ..., r, set

$$R_s = R(\mathcal{D} \cup \mathcal{V}_s) - R(\mathcal{D} \cup \mathcal{V}_{s-1})$$

and

$$\beta_s = \beta(\mathcal{D} \cup \mathcal{V}_s) - \beta(\mathcal{D} \cup \mathcal{V}_{s-1}).$$

Notice that $R_s \subset E$ for s = 1, ..., r; for $\mathcal{D} \cup \mathcal{V}_s \alpha \mathcal{C}$ and $\mathcal{D} \alpha \mathcal{D} \cup \mathcal{V}_{s-1}$.

Lemma 5. For each s, $s = 1, \ldots, r$, there exists an integer k between 1 and n such that $R_s \subset E_k$.

Proof: Since $R(\mathcal{A} \cup \mathcal{V}_s) = R(\mathcal{A} \cup \mathcal{V}_{s-1}) \cup R(\{V_s\})$,

$$R_s = R(\{V_s\}) - R(\mathcal{A} \cup \mathcal{V}_{s-1}).$$

If $R_s = \phi$, then the conclusion of the lemma follows trivially, so it may be supposed that $R_s \neq \phi$. In this case, V_s must be in R_s and hence in E. Let K be the integer for which $V_s \in E_K$. Then there must exist an element $C \in \mathcal{C}_K$ such that $V_x \leq C$, but V_s cannot satisfy $V_s \leq V$ for any $V \in \mathcal{D}_K$. We must now show that any other element $A \in R_s$ must also be in E_K. But if $A \in R_s$, then A satisfies $A \leq V_s \leq C$ and fails to satisfy $A \leq V_t$ for any other element V_t such that $t < s$. Since $C \in \mathcal{C}_K$, $A \in R(\mathcal{C}_K)$, and it suffices to show that $A \notin R(\mathcal{D}_K)$ -- i.e., that $A \leq D$ does not hold for any $D \in \mathcal{D}_K$.

Let us suppose that $A \leq D$ does hold for a given $D \in \mathcal{D}_K$ and derive a contradiction. Indeed, if $A \leq D$, then $A \leq V_s \wedge D$. But since $V_s \in E_K$, $V_s \leq D$ does not hold, and hence $V_s \wedge D$ must be a proper sub-element of V_s and is therefore equal to V_t for some $t < s$. Since we have $A \leq V_t$ for some t such that $t < s$, this is our contradiction. $\boxed{///}$

It follows from lemma 5 that the set $\{1, \ldots, r\}$ can be partitioned into n disjoint sets $N_1, \ldots, N_n$ such that $R_s \subset E_i$ if $s \in N_i$.

If $s \in N_i$, then $R_s \subset E_i$ and $R(\mathcal{D}_i) \cap R_s = \phi$ and $R_s \subset R(\mathcal{C}_i)$, so successive applications of lemmas 2 and 4 result in

$$R_s = R((\mathcal{A} \cup \mathcal{D}_i \cup \mathcal{V}_s) \boxtimes \mathcal{C}_i) - R((\mathcal{A} \cup \mathcal{D}_i \cup \mathcal{V}_{s-1}) \boxtimes \mathcal{C}_i)$$

and

$$\beta_s = \beta((\mathcal{A} \cup \mathcal{D}_i \cup \mathcal{V}_s) \boxtimes \mathcal{C}_i) - \beta((\mathcal{A} \cup \mathcal{D}_i \cup \mathcal{V}_{s-1}) \boxtimes \mathcal{C}_i).$$

If, on the other hand, $s \notin N_i$, then either $R_s = \phi$ or R_s is not contained in E_i. In either case,

$$\begin{aligned}
\phi &= R_s \cap E_i \\
&= (R(\mathcal{A} \cup \mathcal{V}_s) - R(\mathcal{A} \cup \mathcal{V}_{s-1})) \cap (R(\mathcal{C}_i) - R(\mathcal{A}_i)) \\
&= R((\mathcal{A} \cup \mathcal{V}_s) \boxtimes \mathcal{C}_i) - R(\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_{s-1}) \\
&= R((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_s) \boxtimes \mathcal{C}_i) - R((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_{s-1}) \boxtimes \mathcal{C}_i).
\end{aligned}$$

Hence the quantity

$$\beta((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_s) \boxtimes \mathcal{C}_i) - \beta((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_{s-1}) \boxtimes \mathcal{C}_i)$$

is equal to β_s if $s \in N_i$ and zero if $s \notin N_i$. Consequently,

$$\begin{aligned}
\sum_{s \in N_i} \beta_s &= \sum_{s=1}^{r} (\beta((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_s) \boxtimes \mathcal{C}_i) - \beta((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_{s-1}) \boxtimes \mathcal{C}_i)) \\
&= \beta((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_r) \boxtimes \mathcal{C}_i) - \beta((\mathcal{A} \cup \mathcal{A}_i \cup \mathcal{V}_o) \boxtimes \mathcal{C}_i) \\
&= \beta(\mathcal{C}_i) - \beta(\mathcal{A}_i) = b(E_i).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\sum_{s=1}^{r} \beta_s &= \sum_{s=1}^{r} (\beta(\mathcal{A} \cup \mathcal{V}_s) - \beta(\mathcal{A} \cup \mathcal{V}_{s-1})) \\
&= \beta(\mathcal{A} \cup \mathcal{V}_r) - \beta(\mathcal{A}) = \beta(\mathcal{C}) - \beta(\mathcal{A}) = b(E);
\end{aligned}$$

hence

$$\sum_{i=1}^{r} b(E_i) = \sum_{i=1}^{r} \sum_{s \in N_i} \beta_s = \sum_{s=1}^{r} \beta_s = b(E),$$

and b is indeed a measure on $\mathcal{E}$. This completes the proof of the theorem.

9. The Constraint Mapping

Recall that an allocation of probability $\rho: \mathcal{A} \to \mathcal{M}$ is said to specify the constraint relation "ct" between $\mathcal{M}$ and $\mathcal{A}$ whenever " M ct A " is equivalent to $M \leq \rho(A)$. Obviously, ρ and ct are two different ways of conveying exactly the same information, but our attention is concentrated on ρ whenever we attend to a particular proposition $A \in \mathcal{A}$ and ask about the probability that is constrained to A. For $\rho(A)$ is the "largest" probability mass constrained to A, in the sense that the probability masses constrained to A are precisely those which are subelements of $\rho(A)$.

But suppose we fix our attention on a particular probability mass $M \in \mathcal{M}$ and contemplate all the elements of $\mathcal{A}$ to which M is constrained. Then there may or may not be a "smallest" element $\lambda(M)$ among these. In other words, there may or may not be an element $\lambda(M) \in \mathcal{A}$ such that M is constrained to any given element $A \in \mathcal{A}$ if and only if $\lambda(M) \leq A$. If there is such an element $\lambda(M) \in \mathcal{A}$ for each $M \in \mathcal{M}$, then I will call the mapping $\lambda: \mathcal{M} \to \mathcal{A}: M \mapsto \lambda(M)$ the constraint mapping for ρ and ct ; for the mapping λ will specify the "tightest" constraint for each probability mass.

The following definition lists the properties of constraint mappings.

Definition. Suppose $\mathcal{A}$ is a Boolean algebra of propositions and $\mathcal{M}$ is a Boolean algebra of probability masses. Then a mapping $\lambda: \mathcal{M} \to \mathcal{A}$ is a constraint mapping if

(i) $\lambda(\Lambda_{\mathcal{M}}) = \Lambda_{\mathcal{A}}$.

(ii) If $\lambda(M) = \Lambda_{\mathcal{A}}$, then $M = \Lambda_{\mathcal{M}}$.

(iii) $\lambda(M_1 \vee M_2) = \lambda(M_1) \vee \lambda(M_2)$.

One immediate consequence of rule (iii) in this definition is that a constraint mapping λ is monotone, i.e. $\lambda(M_1) \leq \lambda(M_2)$ whenever $M_1 \leq M_2$.

(See Chapter 3, section 3 below.) This formal definition for constraint mappings is justified by the two following propositions.

<u>Theorem.</u> Suppose ct is a constraint relation between a Boolean algebra of probability masses $\mathcal{M}$ and a Boolean algebra of propositions $\mathcal{A}$. And suppose $\lambda : \mathcal{M} \to \mathcal{A}$ is a mapping such that M ct A if and only if $\lambda(M) \leq A$. Then λ is a constraint mapping.

<u>Proof:</u> (i) By rule (2c) for constraint relations, $\Lambda_{\mathcal{M}}$ ct $\Lambda_{\mathcal{A}}$. Hence we must have $\lambda(\Lambda_{\mathcal{M}}) \leq \Lambda_{\mathcal{A}}$, or $\lambda(\Lambda_{\mathcal{M}}) = \Lambda_{\mathcal{A}}$.

(ii) If $\lambda(M) = \Lambda_{\mathcal{A}}$, the M ct $\Lambda_{\mathcal{A}}$. Hence, by rule (3) for constraint relations, $M = \Lambda_{\mathcal{M}}$.

(iii) By hypothesis, M_1 and M_2 are both constrained to a proposition A if and only if $\lambda(M_1) \vee \lambda(M_2) \leq A$. But from rules (2a) and (2b) for constraint relations, M_1 and M_2 are both constrained to A if and only if $M_1 \vee M_2$ is constrained to A. Hence $M_1 \vee M_2$ is constrained to a proposition A if and only if $\lambda(M_1) \vee \lambda(M_2) \leq A$. Hence $\lambda(M_1 \vee M_2) = \lambda(M_1) \vee \lambda(M_2)$. ▨

<u>Theorem.</u> Suppose $\lambda : \mathcal{M} \to \mathcal{A}$ is a constraint mapping. Then the binary relation "ct" between $\mathcal{M}$ and $\mathcal{A}$ defined by "M ct A if and only if $\lambda(M) \leq A$" is a constraint relation. (I will call this the constraint relation given by λ.)

<u>Proof:</u> It is necessary to establish conditions (1), (2) and (3) in the definition of a constraint relation. Condition (1) is immediate; and the others are implied by the three conditions in the definition of a constraint mapping: (i) implies (2c), (ii) implies (3), (iii) implies (2b), and the monotonicity of λ implies (2a). ▨

It should be reiterated that constraint mappings do not always exist, even when an allocation of probability does. In other words, if a constraint relation ct between $\mathcal{M}$ and $\mathcal{A}$ is given directly or by means of an allocation, $\rho: \mathcal{A} \to \mathcal{M}$, then there does not necessarily exist a constraint mapping $\lambda: \mathcal{M} \to \mathcal{A}$ which gives ct. If such a constraint mapping λ does exist, though, it is necessarily unique.

<u>Theorem</u>. Suppose $\mathcal{A}$ is a finite Boolean algebra of propositions, $\mathcal{M}$ is a Boolean algebra of probability masses, and ct is a constraint relation between $\mathcal{M}$ and $\mathcal{A}$. Then a constraint mapping λ exists for ct.

<u>Proof</u>: We can define λ as follows. For each $M \varepsilon \mathcal{M}$, let $A_1, \ldots, A_n$ be all the elements of $\mathcal{A}$ to which M is constrained -- by rule (1c) for constraint relations there is at least one of these, and since $\mathcal{A}$ is finite there can only be a finite number of them. Let $\lambda(M)$ equal to $A_1 \wedge \ldots \wedge A_n$. It then follows from rule (1b) for constraint relations that M ct $\lambda(M)$; hence for any $A \varepsilon \mathcal{A}$, M ct A if and only if $\lambda(M) \leq A$. It follows from the first theorem in this section that λ is a constraint mapping. ▨

In the preceding section we began with an arbitrary belief function on an arbitrary Boolean algebra of propositions $\mathcal{A}$ and constructed an allocation $\rho: \mathcal{A} \to \mathcal{M}$ that gave that belief function. It is natural to ask whether a constraint mapping $\lambda: \mathcal{M} \to \mathcal{A}$ exists forthe allocation ρ so constructed.

The answer is that a constraint mapping does exist. Indeed, if $M \varepsilon \mathcal{M}$, then M is the union of a finite number of disjoint subsets of $\mathcal{J}$ of the form $R(\mathcal{C}_i) - R(\mathcal{A}_i)$ with $\mathcal{A}_i \propto \mathcal{C}_i$. Suppose, indeed, the

$M = \bigcup_{i=1}^{n} (R(\mathcal{C}_i) - R(\mathcal{D}_i))$, where $\mathcal{D}_i \alpha \mathcal{C}_i$ and the $R(\mathcal{C}_i) - R(\mathcal{D}_i)$ are disjoint. For each i, i = 1, ..., n, set $\mathcal{C}'_i$ equal to the subset of $\mathcal{C}_i$ consisting of elements not majorized by $\mathcal{D}_i$; i.e., set

$$\mathcal{C}'_i = \mathcal{C}_i - \{C | C \in \mathcal{C}_i \text{ and } C \leq D \text{ for some } D \in \mathcal{D}_i\}.$$

Then

$$R(\mathcal{C}_i) - R(\mathcal{D}_i) = R(\mathcal{C}'_i) - R(\mathcal{D}_i).$$

Set $\mathcal{C} = \bigcup_{i=1}^{n} \mathcal{C}'_i$. Then $\mathcal{C}$ is a finite subset of $\mathcal{A}$, and $\mathcal{C} \subset M$. Set $\lambda(M)$ equal to the disjunction of all the elements of $\mathcal{C}$. In other words, index the elements of $\mathcal{C}$, say $C_1, \ldots, C_k$, and set $\lambda(M) = C_1 \vee \ldots \vee C_K$. If $\mathcal{C} = \phi$ set $\lambda(M) = \Lambda_{\mathcal{A}}$. Then $\{\lambda(M)\}$ majorizes $\mathcal{C}$, and

$$\rho(\lambda(M)) = R(\{C_1 \vee \ldots \vee C_k\}) \supset R(\mathcal{C}) = R(\mathcal{C}_1) \cup \ldots \cup R(\mathcal{C}_n) \supset M.$$

Hence M ct $\lambda(M)$, and M ct A for all A such that $\lambda(M) \leq A$. On the other hand, if $A \in \mathcal{A}$ and M ct A, then $\mathcal{C} \subset M \subset \rho(A) = R(\{A\})$. This implies that $\{A\}$ majorizes $\mathcal{C}$, whence $\lambda(M) \leq A$. Hence an element A of $\mathcal{A}$ satisfies M ct A if and only if $\lambda(M) \leq A$. Thus the mapping $\lambda: \mathcal{M} \to \mathcal{A}: M \rightsquigarrow \lambda(M)$ is indeed a constraint mapping corresponding to the constraint relation between $\mathcal{M}$ and $\mathcal{A}$.

10. Toward a Better Representation of our Probability

We have now arrived at the conclusion that any belief function on a Boolean algebra $\mathcal{A}$ can be represented by an allocation $\rho: \mathcal{A} \to \mathcal{M}$ that maps $\mathcal{A}$ into a "measure algebra" $\mathcal{M}$. But as I remarked in section 1, our formal definition of a measure algebra falls somewhat short of imposing all the properties that we might want our idealized "probability" to have. The following three properties are the most important of the additional properties that we might want to require of $(\mathcal{M}, \mu)$:

(i) Positivity: If $M \epsilon \mathcal{M}$ and $M \neq \Lambda_{\mathcal{M}}$, then M ought to have measure: $\mu(M) > 0$.

(ii) Completeness: If $\{M_\gamma\}_{\gamma \epsilon \Gamma}$ is any collection of elements of $\mathcal{M}$, then there ought to be an element of $\mathcal{M}$ representing their union and another representing their intersection.

(iii) Complete Additivity: Suppose $\{M_\gamma\}_{\gamma \epsilon \Gamma}$ is a disjoint collection of elements of $\mathcal{M}$. In other words, suppose $M_\gamma \wedge M_{\gamma'} = \Lambda_{\mathcal{M}}$ for all distincts pairs, γ, γ' in Γ. Then the measure of their union ought to be equal to $\sum_\gamma \mu(M_\gamma)$.

These three properties may seem too strong for us to expect that our measure algebra $\mathcal{M}$ should have them. But in fact we can always arrange that $\mathcal{M}$ should have them.

Unfortunately, though, the demonstration of this fact can hardly be carried out without a more thorough knowledge of the mathematics of Boolean algebras. Hence we must turn to an examination of the theory of lattices and Boolean algebras, an examination that is already long overdue.

CHAPTER 3. THE THEORY OF BOOLEAN ALGEBRAS

This chapter is intended as a brief and sketchy introduction to the abstract theory of Boolean algebras. Almost all its vocabulary, assertions and theorems are standard in that theory. For a more thorough study of the subject, the reader may wish to consult Garrett Birkhoff's Lattice Theory, Roman Sikorski's Boolean Algebras, or Paul Halmos' Lectures on Boolean Algebras.

1. Partially Ordered Sets

A binary relation between two sets $\mathcal{A}$ and $\mathcal{B}$ is a subset r of the Cartesian product $\mathcal{A} \times \mathcal{B}$. If $(A, B) \in r$, then one says that the binary relation r holds between A and B, and one writes "A r B." A binary relation "$\leq$" between a set $\mathcal{A}$ and itself is called a partial ordering if

(1) $A \leq A$ for all $A \in \mathcal{A}$.

(2) If $A \leq B$ and $B \leq C$, then $A \leq C$.

(3) If $A \leq B$ and $B \leq A$, then $A = B$.

If a non-empty set has associated with it a partial ordering, then it is called a partially ordered set. If A and B are in a partially ordered set and $A \leq B$, then A is said to minorize B or to be a subelement of B, while B is said to majorize A. If $A \leq B$, and $A \neq B$, then one writes $A < B$ and says that A is a proper subelement of B.

Examples of partially ordered sets abound in mathematics. For example, any set $\mathcal{A}$ of sets becomes a partially ordered set if it is endowed with the partial ordering

$$\leq \; = \{(A,B) \mid A, B \in \mathcal{A}\,;\; A \subset B\},$$

i.e., if it is partially ordered by set inclusion. Other examples are provided by the usual "less than or equal" orderings of numbers.

A partial ordering $\leq$ on a set $\mathcal{A}$ is a <u>total ordering</u> if for every pair $A, B \in \mathcal{A}$, either $A \leq B$ or $B \leq A$.

<u>Theorem</u>. A partial ordering on a finite set can always be extended to a total ordering. More explicitly, if $\leq_o$ is a partial ordering on a finite set $\mathcal{A}$, then there is a total ordering $\leq$ on $\mathcal{A}$ such that $\leq_o \subset \leq$.

<u>Proof</u>: Suppose $\leq_o$ is not a total ordering on $\mathcal{A}$. Then let A, B be elements of $\mathcal{A}$ such that neither $A \leq B$ nor $B \leq A$. Then set

$$\leq_1 \; = \; \leq_o \cup \{(C,D) \mid C \in \mathcal{A}\,;\; D \in \mathcal{A}\,;\; C \leq_o A;\; B \leq_o D\}.$$

It is easily verified that $\leq_1$ is a partial ordering on $\mathcal{A}$ and that $A \leq_1 B$.

Hence a partial ordering can always be extended to make a given "non-comparable" pair comparable. Since $\mathcal{A}$ is finite, one can have only a finite number, say n, of non-comparable pairs; hence the theorem follows by induction. ▨

If $\mathcal{C}$ is a subset of a partially ordered set $\mathcal{A}$, then there can be at most one element $C \in \mathcal{C}$ such that for all $B \in \mathcal{C}$, $C \leq B$. For if there were two such elements C_1 and C_2 in $\mathcal{C}$, then they would satisfy $C_1 \leq C_2$ and $C_2 \leq C_1$ and hence, by (3), $C_1 = C_2$. If such an element does exist it is, quite naturally, called the <u>least</u> element of $\mathcal{C}$. Similarly, $\mathcal{C}$ may or may not have a <u>greatest</u> element -- i.e., an element $C \in \mathcal{C}$ such that for all $B \in \mathcal{C}$, $B \leq C$; but if there is such an element, it is unique. If the partially ordered set $\mathcal{Q}$ itself has a least element, that element is called the <u>zero</u> of $\mathcal{Q}$, and denoted $\perp_{\mathcal{Q}}$ or $\perp$; if it has a greatest element, that element is called the <u>unit</u> of $\mathcal{Q}$ and denoted $\mathrm{V}_{\mathcal{Q}}$ or V.

If $\mathcal{C}$ is a subset of a partially ordered set $\mathcal{Q}$, then an element $A \in \mathcal{Q}$ is called a <u>lower bound</u> of $\mathcal{C}$ if $A \leq C$ for all $C \in \mathcal{C}$, and an element $A \in \mathcal{Q}$ is called an <u>upper bound</u> of $\mathcal{C}$ if $C \leq A$ for all $C \in \mathcal{C}$. It is possible for a given proper subset $\mathcal{C}$ of $\mathcal{Q}$ to have many lower and upper bounds in $\mathcal{Q}$, but $\mathcal{Q}$ itself can have at most one lower bound and one upper bound. For the lower bound of $\mathcal{Q}$, if it exists, is its zero; and the upper bound, if it exists, is its unit. The zero and unit of $\mathcal{Q}$ are sometimes called the <u>universal bounds</u> of $\mathcal{Q}$.

Let $\mathcal{C}$ be a given subset of a partially ordered set $\mathcal{Q}$, and let $\mathcal{L} \subset \mathcal{Q}$ be the collection of all the lower bounds of $\mathcal{C}$. The set $\mathcal{L}$ may or may not be empty, and if it is not empty, then it may or may not have a greatest element. If $\mathcal{L}$ is non-empty, and does have a greatest element, then that element is called, quite naturally, the <u>greatest lower bound</u> of $\mathcal{C}$; it is also called the <u>meet</u> of the elements of $\mathcal{C}$. Similarly, if the collection $\mathcal{U}$ of all upper bounds of $\mathcal{C}$ is non-empty and has a least element, then that element is called the <u>least upper bound</u> of $\mathcal{C}$, or the <u>join</u> of the elements of $\mathcal{C}$.

The notions of meet and join are of central importance in lattice theory, and it may be worthwhile to repeat their definitions in a less verbal way, replacing the set $\mathcal{C}$ with an indexed collection $\{A_\gamma\}_{\gamma\in\Gamma}$ of elements of $\mathcal{Q}$: The <u>meet</u> of a collection $\{A_\gamma\}_{\gamma\in\Gamma}$ is the element $A \in \mathcal{Q}$, unique if it exists, such that $A \leq A_\gamma$ for all $\gamma\in\Gamma$ and $B \leq A$ if $B \in \mathcal{Q}$ is any other element satisfying $B \leq A_\gamma$ for all $\gamma\in\Gamma$. The <u>join</u> of a collection $\{A_\gamma\}_{\gamma\in\Gamma}$ is the element $A \in \mathcal{A}$, unique if it exists, such that $A_\gamma \leq A$ for all $\gamma \in \Gamma$ and $A \leq B$ if $B \in \mathcal{A}$ is any other element satisfying $A_\gamma \leq B$ for all $\gamma \in \Gamma$.

It should be borne in mind that the notions of meet and join are relative to a fixed partially ordered set $\mathcal{A}$. For it is possible that a subset $\mathcal{C}$ of a partially ordered set $\mathcal{A}$ might also be a subset of a different partially ordered set $\mathcal{B}$; in such a case, $\mathcal{C}$ might have, say, one meet in $\mathcal{A}$ and a different one in $\mathcal{B}$ -- or perhaps a meet in $\mathcal{A}$ and no meet at all in $\mathcal{B}$.

The symbol "$\wedge$" usually is used to denote a meet: the meet of $\mathcal{C} \subset \mathcal{A}$ is denoted by $\wedge \mathcal{C}$, the meet of a collection $\{A_\gamma\}_{\gamma \in \Gamma}$ of elements of $\mathcal{A}$ is denoted by $\bigwedge_\gamma A_\gamma$ or $\wedge A_\gamma$, and the meet of a pair of elements A and B of $\mathcal{A}$ is denoted by $A \wedge B$. The symbol "$\vee$" is used analogously for joins; one writes $\vee \mathcal{C}$, $\bigvee_\gamma A_\gamma$ or $\vee A_\gamma$, and $A \vee B$. The similarity between this notation and the notation for intersection and union in the theory of sets is justified by the fact that if $\mathcal{A}$ is a collection of subsets of a given set and $\mathcal{A}$ is closed under the operations of union and intersection, then $\mathcal{A}$ is partially ordered by set inclusion and every collection $\{A_\gamma\}_{\gamma \in \Gamma}$ of elements of has a meet and a join, which are given by the intersection and union respectively.

A partially ordered set $\mathcal{A}$ is called a <u>meet-semilattice</u> if every pair of elements in $\mathcal{A}$ have a meet in $\mathcal{A}$. Similarly, it is called a <u>join-semilattice</u> if every pair of elements has a join, and simply a <u>lattice</u> if every pair of elements has both a meet and a join.

It is easily deduced that meets and joins exist for all <u>finite</u> collections of elements in a lattice. They need not exist, however, for infinite collections. A lattice for which they all do exist is said to be <u>complete</u>. A finite lattice is necessarily complete. Actually, the existence of meets for all collections of elements in a lattice implies the existence of joins for all collections and hence implies completeness; similarly, the existence of joins for all collections implies completeness.

If meets and joins exist for all countable collections of elements in a lattice, then the lattice is said to be σ-complete. Of course, a complete lattice is σ-complete. The existence either of meets for all countable collections or of joins for all countable collections is sufficient to assure σ-completeness.

Notice that a complete lattice necessarily has universal bounds, for the meet of all the elements of the lattice will be the zero, and the join of all the elements of the lattice will be the unit.

If a partially ordered set has only one element, then that element will be both the zero and the unit, but if the set has more than one element, then the zero and unit must be distinct if they both exist. It is easily seen that if $\mathcal{\Lambda}$ is the zero of a partially ordered set $\mathcal{A}$ and $A \in \mathcal{A}$, then $A \wedge \Lambda = \Lambda$ and $A \vee \Lambda = A$. Similarly if $\mathcal{V}$ if the unit of $\mathcal{A}$ and $A \in \mathcal{A}$, then $A \wedge \mathcal{V} = A$ and $A \vee \mathcal{V} = \mathcal{V}$.

If A, B and C are elements of a lattice and $B \leq C$, then $B \vee A \leq C \vee A$ and $B \wedge A \leq C \wedge A$.

A lattice is distributive if all triplets of elements A, B, C in the lattice satisfy

$$(1) \quad A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$$

and

$$(2) \quad A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C).$$

Actually, either of rules (1) and (2) implies the other. They also imply various infinite distributive laws. Among them:

$$A \wedge (\bigvee_{\gamma} A_{\gamma}) = \bigvee_{\gamma} (A \wedge A_{\gamma})$$

and

$$(\bigvee_{\alpha} A_{\alpha}) \wedge (\bigvee_{\beta} B_{\beta}) = \bigvee_{\alpha, \beta} (A_{\alpha} \wedge B_{\beta}).$$

These equations are to be interpreted in the sense that if the left side exists, then so does the right, and the two are equal.

If A and B are elements of a lattice with zero and unit and $A \wedge B = \Lambda$ and $A \vee B = \mathrm{V}$, then B is called a complement of A. Complements in a distributive lattice are unique if they exist, the unique complement of an element A being denoted by $\overline{A}$. A distributive lattice with distinct zero and unit that includes complements for all its elements is called a Boolean algebra.

2. Boolean Algebras

The definition of a Boolean algebra is based on the whole series of concepts and definitions set forth in the preceding section. It is possible, though, to translate the definition into a list of conditions that a set $\mathcal{Q}$ of objects must satisfy in order to qualify as a Boolean algebra:

(1) Existence of a partial ordering: $\mathcal{Q}$ must have an ordering that obeys the rules for partial orderings.

(2) Existence of a zero: $\mathcal{Q}$ must have an element that minorizes all the other elements. (Such an element is necessarily unique and is denoted Λ.)

(3) Existence of a unit: $\mathcal{Q}$ must have an element that majorizes all the other elements. (Such an element is necessarily unique and is denoted V.)

(4) Non-identity of the zero and unit: Λ and V must be distinct.

(Equivalently, $\mathcal{A}$ must have at least two elements.)

(5) Existence of meets: For every pair of elements A and B in $\mathcal{A}$, there is a greatest element among those that minorize them both. (This element is denoted by $A \wedge B$.)

(6) Existence of joins: For every pair of elements A and B in $\mathcal{A}$, there is a least element among those that majorize them both. (This element is denoted $A \vee B$.)

(7) Distributivity: For any triplet of elements A, B and C in $\mathcal{A}$, $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$ and $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$.

(8) Existence of complements: For every element A there is an element B such that $A \wedge B = \perp$ and $A \vee B = \top$. (Such an element B is necessarily unique and is denoted $\overline{A}$.)

This list of conditions should enable us to decide whether our "Boolean algebras of propositions" and "Boolean algebras of probability masses" really pass muster to qualify as Boolean algebras in the mathematical sense.

Consider first a "Boolean algebra of propositions." I have been using this term to refer to any non-empty collection of propositions that includes the negation of each of its elements and the conjunction and disjunction of each pair of its elements. Such a collection does indeed satisfy the eight conditions listed above when it is partially ordered by implication -- i.e., when one proposition is said to minorize another if and only if it implies the other. The third rule for partial orderings then corresponds to the fact that propositions are held to be identical when they are logically equivalent. The zero and the unit are the impossible and sure propositions, respectively; the meet and join of two propositions are their conjunction and disjunction, respectively; and the complement of a proposition is its negation. The only one of the eight conditions that might cause any head-scratching is the requirement of distributivity, and careful thought will show it to be satisfied. So a "Boolean algebra of propositions" is indeed a Boolean algebra.

Since a Boolean algebra of propositions $\mathcal{A}$ contains the conjunction and disjunction of any pair of its elements, it also contains the conjunction and disjunction of any finite number $A_1, \ldots, A_n$ of its elements -- and of course the conjunction of the elements $A_1, \ldots, A_n$ will be their meet in $\mathcal{A}$ and their disjunction will be their join in $\mathcal{A}$. It should be noted, however, that $\mathcal{A}$ need not contain propositions corresponding to the logical conjunction or disjunction of any given infinite collection of its elements. If $\mathcal{A}$ does contain a proposition corresponding to the conjunction, say, of an infinite collection $\{A_\gamma\}_{\gamma\epsilon\Gamma}$ of its elements, then that proposition would be the meet of the collection $\{A_\gamma\}_{\gamma\epsilon\Gamma}$. But if $\mathcal{A}$ does not contain such a proposition, then the collection might not even have a meet -- and if it does have a meet, that meet might not be the conjunction of the elements $\{A_\gamma\}_{\gamma\epsilon\Gamma}$. In short, finite meets and joins can always be interpreted as conjunctions and disjunctions, respectively, but infinite ones cannot always be.

How about "Boolean algebras of probability masses"? Do they qualify as Boolean algebras in the mathematical sense? As it stands now, our notion of a Boolean algebra of probability masses is based merely on the intuitive idea that probability masses are pieces of an idealized substanced called our "probability" -- an idealized substance that may not even consist of points. But it is evident that this intuitive idea readily fits with the eight conditions for Boolean algebras. The partial ordering is provided by setting $A \leq B$ whenever the probability mass A is part of the larger probability mass B. The unit, of course, is the entire probability mass. The existence of a zero is not so obvious: one might not at first contemplate a single probability mass that is part of all the others. But it is possible to invent a "null" probability mass and make it part of all the others by convention. The meet and join of two probability masses correspond intuitively to their "intersection" and "union"; while the complement of a given probability mass consists of precisely what is left over. The distributive laws are also intuitively valid.

There are a great many relations that are always satisfied by meets, joins and complements in a Boolean algebra, and I have taken several of them for granted in my discussions of Boolean algebras of propositions and Boolean algebras of probability masses.

Notice, for example, that for any two elements A and B of a Boolean algebra, $B \leq A$ if and only if $B \wedge \overline{A} = \curlywedge$. For if $B \leq A$, then $B \wedge \overline{A} \leq A \wedge \overline{A} = \curlywedge$. And if $B \wedge \overline{A} = \curlywedge$, then

$$(B \wedge \overline{A}) \vee (B \wedge A) = \curlywedge \vee (B \wedge A) = B \wedge A.$$

But by the first distributive law, the left-hand side of this equation is equal to $B \wedge (A \vee \overline{A}) = B \wedge \Upsilon = B$. So $B = B \wedge A$, and hence $B \leq A$. If $A \wedge B = \curlywedge$, then A and B are said to be <u>disjoint</u>; hence the preceding fact can be expressed by saying that $B \leq A$ if and only if B and $\overline{A}$ are disjoint. The quantity $B \wedge \overline{A}$ is often written as B-A.

If A, B, C are elements of a Boolean algebra, A and B are disjoint and $C = A \vee B$, then the expression "$A \vee B$" is called a <u>disjoint partition</u>

of C. Notice that for any two elements A, B in a Boolean algebra, $(A \wedge B) \vee (A - B)$ is a disjoint partition of A. Notice also that if $A \vee B$ is a disjoint partition of C and $A = C$, then $B = \Lambda$.

For any pair of elements A and B in a Boolean algebra,

$$\overline{A \vee B} = \overline{A} \wedge \overline{B} \qquad \text{and} \qquad \overline{A \wedge B} = \overline{A} \vee \overline{B}.$$

These identities, and the analogous ones for meets and joins of any finite number of elements, are known as de Morgan's laws.

Like a lattice, a Boolean algebra is called <u>complete</u> if it contains meets and joins for all its subsets, even infinite ones. Complete Boolean algebras obey the infinite version of De Morgan's laws, to wit:

$$\overline{\bigvee_{\gamma} A_{\gamma}} = \bigwedge_{\gamma} \overline{A}_{\gamma} \quad \text{and} \quad \overline{\bigwedge_{\gamma} A_{\gamma}} = \bigvee_{\gamma} \overline{A}_{\gamma}$$

We will sometimes be interested in the weaker condition of σ-completeness. A Boolean algebra is called <u>σ-complete</u>, of course, if it contains meets and joins for all <u>countable</u> collections of its elements.

3. The Mappings of Lattice Theory

Like any algebraic theory, the theory of lattices and Boolean algebras gives a prominent role to mappings that preserve the structure of its objects. In this section, we will learn the names of some of these mappings.

The simplest requirement in this context is that a mapping should preserve the structure of a semi-lattice. If $\mathcal{A}$ and $\mathcal{B}$ are meet-semi-lattices, for example, then a mapping $\theta: \mathcal{A} \to \mathcal{B}$ that obeys

(1) $\theta(A_1 \wedge A_2) = \theta(A_1) \wedge \theta(A_2)$ for all A_1, A_2 in $\mathcal{A}$

is called a meet-morphism. Similarly, if $\mathcal{A}$ and $\mathcal{B}$ are join-semilattices and a mapping $\theta: \mathcal{A} \to \mathcal{B}$ obeys

$$(2)\ \theta(A_1 \vee A_2) = \theta(A_1) \vee \theta(A_2) \text{ for all } A_1, A_2 \text{ in } \mathcal{A},$$

then θ is called a <u>join -morphism</u>. We saw examples of meet-morphisms in the preceding chapter: an allocation of probability is a meet-morphism, while a constraint mapping or an allowment is a join -morphism.

Meet-morphisms and join-morphisms are both order-preserving, or <u>isotone</u>; in other words, they both necessarily obey the rule

$$(3)\ \text{If } A_1 \leq A_2, \text{ then } \theta(A_1) \leq \theta(A_2).$$

This can be proven for a meet-morphism, for example, by using the fact that $A_1 \wedge A_2 = A_1$ whenever $A_1 \leq A_2$, for one obtains $\theta(A_1) = \theta(A_1 \wedge A_2) = \theta(A_1) \wedge \theta(A_2)$, whence $\theta(A_1) \leq \theta(A_2)$ by the definition of meet.

If $\mathcal{A}$ and $\mathcal{B}$ are lattices, then a mapping $\theta: \mathcal{A} \to \mathcal{B}$ is called a <u>lattice homomorphism</u> if it obeys both (1) and (2); i.e., if it is both a meet-morphism and a join-morphism. Finally, if $\mathcal{A}$ and $\mathcal{B}$ are Boolean algebras, then a mapping $\theta: \mathcal{A} \to \mathcal{B}$ is called a <u>Boolean homomorphism</u> if it obeys (1), (2), and

$$(4)\quad \theta(\overline{A}) = \overline{\theta(A)} \text{ for all } A \in \mathcal{A}.$$

In other words, a Boolean homomorphism is a lattice homomorphism that preserves complements. It can easily be deduced from de Morgan's law that if a mapping between two Boolean algebras obeys (4) and one of (1) and (2), then it must also obey the other. Hence in order for a mapping between two Boolean algebras to be a Boolean homomorphism, it suffices either for it to preserve complements and meets or for it to preserve complements and joins.

It is easily seen that a Boolean homomorphism $\theta: \mathcal{A} \to \mathcal{B}$ also preserves the unit and the zero; i.e., it obeys

$$(5) \qquad \theta(\perp_{\mathcal{A}}) = \perp_{\mathcal{B}},$$

and

$$(6) \qquad \theta(\top_{\mathcal{A}}) = \top_{\mathcal{B}}.$$

To prove (5), for example, note that for any $A \in \mathcal{A}$, $\theta(\perp_{\mathcal{A}}) = \theta(A \wedge \overline{A}) = \theta(A) \wedge \theta(\overline{A}) = \theta(A) \wedge \overline{\theta(A)} = \perp_{\mathcal{B}}$.

A subset $\mathcal{A}_o$ of a Boolean algebra $\mathcal{A}$ is called a subalgebra of $\mathcal{A}$ if it satisfies the following conditions:

(i) $\perp_{\mathcal{A}}$ and $\top_{\mathcal{A}}$ are in $\mathcal{A}_o$.

(ii) $\overline{A} \in \mathcal{A}_o$ for all $A \in \mathcal{A}_o$.

(iii) $A_1 \wedge A_2$ and $A_1 \vee A_2$ are in $\mathcal{A}_o$ for all pairs A_1, A_2 in $\mathcal{A}_o$.

Obviously, a subalgebra of a Boolean algebra is a Boolean algebra in its own right, its partial ordering being that inherited from the larger Boolean algebra. It is evident from equations (1), (2), (4), (5) and (6) that if $\theta: \mathcal{A} \to \mathcal{B}$ is a Boolean homomorphism, then the image $\theta(\mathcal{A})$ is a subalgebra of $\mathcal{B}$.

If a Boolean homomorphism $\theta: \mathcal{A} \to \mathcal{B}$ is one-to-one and onto, then it is called an isomorphism onto $\mathcal{B}$; it is easily verified that the inverse $\theta^{-1}: \mathcal{B} \to \mathcal{A}$ will then be an isomorphism onto $\mathcal{A}$. If a Boolean homomorphism $\theta: \mathcal{A} \to \mathcal{B}$ is merely one-to-one, it is called an isomorphism into $\mathcal{B}$; in such a case $\theta': \mathcal{A} \to \theta(\mathcal{A}): A \mapsto \theta(A)$ will be an isomorphism onto the image $\theta(\mathcal{A})$, considered as a Boolean algebra in its own right. An isomorphism into is sometimes called an embedding.

If an isomorphism onto exists between two Boolean algebras $\mathcal{A}$ and $\mathcal{B}$, then the two are said to be isomorphic. Such an isomorphism onto will necessarily preserve arbitrary meets and joins. For example, if $\{A_\gamma\}_{\gamma \in \Gamma}$ and A are elements of $\mathcal{A}$, $A = \wedge A_\gamma$ and $\theta: \mathcal{A} \to \mathcal{B}$ is an isomorphism onto, then $\wedge\, \theta(A_\gamma)$ will exist in $\mathcal{B}$ and will be equal to $\theta(A)$. This will not necessarily be true, though, if θ is merely an isomorphism into.

If $\mathcal{A}$ and $\mathcal{B}$ are both complete Boolean algebras, then a Boolean homomorphism θ: $\mathcal{A} \to \mathcal{B}$ is called complete if it preserves arbitrary meets and joins -- i.e., if

(7) $\theta(\wedge A_\gamma) = \wedge\, \theta(A_\gamma)$

and (8) $\theta(\vee A_\gamma) = \vee\, \theta(A_\gamma)$

for all collections $\{A_\gamma\}$ of elements of $\mathcal{A}$.

A subalgebra $\mathcal{A}_o$ of a complete Boolean algebra $\mathcal{A}$ is said to be a complete subalgebra if it includes meets and joins for arbitrary collections of its elements. A complete subalgebra is obviously a complete Boolean algebra. The image of a complete Boolean homomorphism is a complete subalgebra.

Similar statements can be made for σ-completeness: a subalgebra of a σ-complete Boolean algebra is called σ-complete if it is closed under countable meets and joins; a Boolean homomorphism between two σ-complete Boolean algebras is called σ-complete if it preserves countable meets and joins; and the image of a σ-complete Boolean homomorphism is a σ-complete subalgebra.

4. Filters and Ideals in Boolean Algebras

Filters and ideals are subsets of Boolean algebras that have certain closure properties. They play an important role in the general theory of Boolean algebras, and they will play an equally important role in our theory. They are so closely related that in a certain sense it would suffice to study only the one or the other, but it is more satisfying to learn about them both together.

A filter in a Boolean algebra $\mathcal{A}$ is a subset F of $\mathcal{A}$ that satisfies

(a) If $A \in F$ and $A \le B$, then $B \in F$.

(b) If $A \in F$ and $B \in F$, then $A \wedge B \in F$.

(c) $\mathcal{V} \in F$.

Notice that (c) assures that a filter cannot be empty. An ideal in a Boolean algebra $\mathcal{A}$ is a subset I of $\mathcal{A}$ that satisfies

(a) If $A \in I$ and $B \le A$, then $B \in I$.

(b) If $A \in I$ and $B \in I$, then $A \vee B \in I$.

(c) $\Lambda \in I$.

Actually, we already encountered filters and ideals in Chapter 2. Indeed, a glance at the definition of a constraint relation is that chapter will reveal that the collection of all the propositions to which a given probability mass is constrained is a filter, while the collection of all the probability masses constrained to a given proposition is an ideal.

If A is any element of a Boolean algebra $\mathcal{A}$, then the subset $\{A' \mid A' \in \mathcal{A}, A \le A'\}$ of $\mathcal{A}$ is a filter, while the subset $\{A' \mid A' \in \mathcal{A}, A' \le A\}$ is an ideal. A filter or ideal of this form is called principal. It can easily be shown that any ideal or filter in a finite Boolean algebra must be principal. Suppose, for example, that F is a filter in a finite Boolean algebra $\mathcal{A}$. Let $A_1, \ldots, A_k$ be the elements of F. Then it follows from (b) in the definition of filter that $\bigwedge_{i=1}^{k} A_i \in F$. But $\bigwedge_{i=1}^{k} A_i \le A_i$ for all i, and by (a) in the definition of filter, if $\bigwedge_{i=1}^{k} A_i \le A$ for some $A \in \mathcal{A}$, then $A \in F$. Hence $F = \{A \mid A \in \mathcal{A}, \bigwedge_{i=1}^{k} A_i \le A\}$.

The only subset of a Boolean algebra $\mathcal{A}$ that is both an ideal and a filter is $\mathcal{A}$ itself. An ideal or filter which is not equal to $\mathcal{A}$ is called a proper ideal or filter. It is evident that an ideal in $\mathcal{A}$ is proper if and only if it does not contain $\mathcal{V}$, while a filter in $\mathcal{A}$ is proper if and only if it does not contain Λ.

A filter F in a Boolean algebra $\mathcal{A}$ is called an <u>ultrafilter</u> if (i) F is proper and (ii) there is no other proper filter F' in $\mathcal{A}$ such that $F \neq F'$ and $F \subset F'$.

<u>Theorem</u>. A filter F in a Boolean algebra $\mathcal{A}$ is an ultrafilter if and only if for every element $A \in \mathcal{A}$ exactly one of the pair A, $\overline{A}$ is in F.

<u>Proof</u>: (i) Suppose that for each $A \in \mathcal{A}$, the filter F contains exactly one of the pair A, $\overline{A}$. Then $V \in F$ and $\overline{V} = \Lambda \notin F$, so F is proper. and if F' is a filter such that $F \neq F'$ and $F \subset F'$, there must be an element $A \in F'$ such that $A \notin F$. So $\overline{A}$ will be in F, whence $\overline{A} \in F'$ and $\overline{A} \wedge A = \Lambda \in F'$. Hence F' will not be proper. So F is an ultrafilter.

(ii) Now let us suppose that for some $A \in \mathcal{A}$, F does <u>not</u> contain exactly one of the pairs A, $\overline{A}$ and deduce that F is not an ultrafilter. We must consider the case where F contains both A and $\overline{A}$, and the case where it contains neither. If it contains both, then it would contain $A \wedge \overline{A} = \Lambda$ and hence would not be proper. If it contains neither, on the other hand, then the filter $F' = \{A' \mid A \wedge B \leq A' \text{ for some } B \in F\}$ satisfies $F \neq F'$ and $F \subset F'$, for F' contains both A and F. Furthermore, F' is proper. For if $\Lambda \in F$, then there exists some $B \in F$ such that $A \wedge B = \Lambda$, or $B \leq \overline{A}$, but this would contradict the assumption that $\overline{A} \notin F$. So F is contained in a larger proper filter and hence is not an ultrafilter. ▨

The notions of completeness can also be applied to filters and ideals. For example, an ideal in a σ-complete Boolean algebra is called a <u>σ-ideal</u> if it contains $\vee A_\gamma$ whenever it contains each element of a countable collection

$\{A_\gamma\}_{\gamma \in \Gamma}$. Similarly, an ideal in a complete Boolean algebra is called a complete ideal if it contains $\vee A_\gamma$ whenever it contains each element of an arbitrary collection $\{A_\gamma\}_{\gamma \in \Gamma}$. It is easily seen that an ideal in a complete Boolean algebra is complete if and only if it is principal.

5. Zorn's Lemma

In this section, I will state Zorn's lemma and us it to deduce some useful facts about Boolean algebras. Being equivalent to the principle of transfinite induction, Zorn's lemma is somewhat controversial among students of the foundations of mathematics, but it is generally accepted as a working tool. A proof of Zorn's lemma can be found on pp. 62-65 of Halmos' Naive Set Theory.

In order to state Zorn's lemma, we need the notions of a chain and of a maximal element in a partially ordered set. A maximal element in a partially ordered set is an element which is not a subelement of any other element. A unit in a partially ordered set is necessarily maximal, but a maximal element need not be a unit. A chain in a partially ordered set is a non-empty subset, any two elements A, B of which satisfy either $A \leq B$ or $B \leq A$.

Zorn's Lemma. If every chain in a partially ordered set has an upper bound, then that partially ordred set has at least one maximal element.

The two following theorems seem to require the use of Zorn's lemma in their proof.

Theorem. If F is a proper filter in a Boolean algebra $\mathcal{A}$, then F is contained in some ultrafilter in $\mathcal{A}$.

Proof: Let $\mathcal{S} = \{F' | F'$ is a proper filter in $\mathcal{A}$; $F \subset F'\}$, and let $\mathcal{S}$ be partially ordered by set inclusion. Notice that any maximal element of $\mathcal{S}$ is an ultrafilter. Hence we need only show that $\mathcal{S}$ has a maximal element. Let $\mathcal{K}$ be any chain in $\mathcal{S}$. Then it is easily seen that $\bigcup \mathcal{K}$ is a filter, and it is proper, for it does not contain Λ. Hence $\bigcup \mathcal{K}$ is in $\mathcal{S}$ and is an upper bound for the chain $\mathcal{K}$. Thus every chain in $\mathcal{S}$ has an upper bound, and by Zorn's lemma, $\mathcal{S}$ has a maximal element. ▨

Corollary. If A is a non-zero element of a Boolean algebra $\mathcal{A}$, then A is contained in some ultrafilter of $\mathcal{A}$.

Proof: A is contained in the proper filter $F = \{A' | A \leq A'\}$. ▨

Theorem. Suppose $\mathcal{A}$ is a Boolean algebra and $\mathcal{C} \subset \mathcal{A}$. Then there exists a subset $\mathcal{D} \subset \mathcal{A}$ such that (i) $\mathcal{D}$ is disjoint, (ii) for each $D \in \mathcal{D}$ there exists $C \in \mathcal{C}$ such that $D \leq C$, and (iii) $\mathcal{D}$ and $\mathcal{C}$ have the same set of upper bounds.

Proof: Set $\mathcal{S} = \{\mathcal{E} | \mathcal{E} \subset \mathcal{A}$; $\mathcal{E}$ is disjoint; and for each $E \in \mathcal{E}$ there exists $C \in \mathcal{C}$ such that $E \leq C\}$, and partially order $\mathcal{S}$ by set inclusion. If $\mathcal{S}_0$ is a chain in $\mathcal{S}$, then it is easily seen that $\bigcup \mathcal{S}_0$ is in $\mathcal{S}$; and $\bigcup \mathcal{S}_0$ will be an upper bound for $\mathcal{S}_0$ in $\mathcal{S}$. So every chain in $\mathcal{S}$ has an upper bound, and by Zorn's lemma $\mathcal{S}$ has at least one maximal element. Let $\mathcal{D}$ be such a maximal element of $\mathcal{S}$. Then it is evident that (i) $\mathcal{D}$ is disjoint, (ii) for each $D \in \mathcal{D}$ there exists an element $C \in \mathcal{C}$ such that $D \leq C$, and (iii) any upper bound of $\mathcal{C}$ is an upper bound for $\mathcal{D}$.

The proof will be complete if we can show that any upper bound A for $\mathcal{D}$ is an upper bound for $\mathcal{C}$. Consider any element $C \in \mathcal{C}$ and note that C-A will be disjoint from all the element s of $\mathcal{D}$. The set $\mathcal{D} \cup \{C-A\}$ will therefore be in $\mathcal{S}$. But $\mathcal{D}$ is already a maximal element of $\mathcal{S}$. Hence C-A must already be in $\mathcal{D}$, and this is possible only if $C-A = \Lambda$. It follows that $C \leq A$. Hence A is an upper bound for $\mathcal{C}$.

6. Fields of Subsets

I used the notion of a field of subsets extensively in the preceding chapters, and I often switched back and forth between the notions of a field of subsets and the notion of a Boolean algebra of propositions. In particular, I often used three facts: (i) any field of subsets is a Boolean algebra under the partial ordering by set inclusion; (ii) any finite Boolean algebra is isomorphic to the field of all subsets of some finite set; (iii) any Boolean algebra, whether finite or not, is isomorphic to some field of subsets of some set. Now that we have a mathematical definition for the notion of a Boolean algebra, we can verify these three facts.

A non-empty collection $\mathcal{F}$ of subsets of a non-empty set $\mathcal{S}$ is called a <u>field of subsets</u> of $\mathcal{S}$ if whenever $\mathcal{F}$ contains two sets A and B, it also contains their union, their intersection, and their set-theoretic complements. In particular, it must contain some subset A of $\mathcal{S}$, the complement $\overline{A}$ of A, their intersection $A \cap \overline{A}$, which is the empty set ϕ, and their union $A \cup \overline{A}$, which is $\mathcal{S}$ itself. A given non-empty set $\mathcal{S}$ will have, of course, many different possible fields of subsets, ranging from the two-element field $\{\phi, \mathcal{S}\}$ to the field that includes all the subsets of $\mathcal{S}$. This latter field of subsets of $\mathcal{S}$ is called the power set of $\mathcal{S}$, and I will denote it by $\mathcal{P}(\mathcal{S})$.

It is easily verified that the binary relation "$\leq$" between a field of subsets $\mathcal{F}$ and itself defined by "$A \leq B$ if an only if $A \subset B$" is a partial ordering. Furthermore, $\mathcal{F}$ is a Boolean algebra under this partial ordering, the Boolean complement of an element being its set-theoretic complement, the meet of two elements being their intersection, the join of two elements being their union, the zero being ϕ, and the unit being $\mathcal{S}$. The meet and join of any finite collection of elements will also be their intersection and union, respectively; but the same is not necessarily true of infinite collections. If the intersection, say, of a given infinite collections of elements of $\mathcal{F}$ is in $\mathcal{F}$, then it will certainly be the meet of that collection; but otherwise the collection may have some other element of $\mathcal{F}$ as its meet, or may not even have a meet in $\mathcal{F}$.

The assertion that any Boolean algebra is isomorphic to a field of subsets is called the Stone Representation Theorem.

Theorem. Let $\mathcal{A}$ be a Boolean algebra, let $\mathcal{S}$ be the set of ultrafilters in $\mathcal{A}$, and for each $A \in \mathcal{A}$, let $f(A)$ be the subset of $\mathcal{S}$ consisting of all the ultrafilters in $\mathcal{A}$ that contain A, i.e., set $f(A) = \{F | F \in \mathcal{S}, A \in F\}$. Denote by $\mathcal{B}$ the collection of all subsets S of $\mathcal{S}$ such that $S = f(A)$ for some $A \in \mathcal{A}$. Then $\mathcal{B}$ is a field of subsets of $\mathcal{S}$ and the mapping $f: \mathcal{A} \to \mathcal{B} : A \rightsquigarrow f(A)$ is an isomorphism.

Proof: It is easily verified that $f(\bigwedge_{\mathcal{A}}) = \phi$ and $f(\bigvee_{\mathcal{A}}) = \mathcal{S}$. In order to show that f preserves meets, we can use the fact that a filter in a

Boolean algebra contains both of a given pair of elements if and only if it contains their meet. Hence if $A_1, A_2 \in \mathcal{A}$, then

$$\begin{aligned} f(A_1) \cap f(A_2) &= \{F \mid F \in \mathcal{J}, A_1 \in F\} \cap \{F \mid F \in \mathcal{J}, A_2 \in F\} \\ &= \{F \mid F \in \mathcal{J}, A_1 \in F, A_2 \in F\} \\ &= \{F \mid F \in \mathcal{J}, A_1 \wedge A_2 \in F\} \\ &= f(A_1 \wedge A_2). \end{aligned}$$

In order to show that f also preserves complements, it is necessary to use the fact that the filters in $\mathcal{J}$ are ultrafilters. This means that given any $A \in \mathcal{A}$, a filter F in $\mathcal{J}$ contains exactly one of the pair A and $\overline{A}$. Hence

$$\begin{aligned} \overline{f(A)} &= \overline{\{F \mid F \in \mathcal{J}, A \in F\}} = \{F \mid F \in \mathcal{J}, A \notin F\} \\ &= \{F \mid F \in \mathcal{J}, \overline{A} \in F\} = f(\overline{A}) . \end{aligned}$$

It follows easily by de Morgan's laws that f also preserves joins. For if $A_1, A_2 \in \mathcal{A}$, then

$$\begin{aligned} f(A_1 \vee A_2) &= f(\overline{\overline{A}_1 \wedge \overline{A}_2}) = \overline{f(\overline{A}_1 \wedge \overline{A}_2)} = \overline{f(\overline{A}_1) \cap f(\overline{A}_2)} = \overline{\overline{f(A_1)} \cap \overline{f(A_2)}} \\ &= f(A_1) \cup f(A_2). \end{aligned}$$

These formulae prove in particular that $\mathcal{B}$ is a field. For if S is in $\mathcal{B}$, then there is an element $A \in \mathcal{A}$ such that $S = f(A)$ and hence $\overline{S} = f(\overline{A})$ is also in $\mathcal{B}$. And if $S_1, S_2 \in \mathcal{B}$, then there exist $A_1, A_2 \in \mathcal{A}$ such that $S_1 = f(A_1)$ and $S_2 = f(A_2)$, so that $S_1 \cup S_2 = f(A_1 \vee A_2)$ and $S_1 \cap S_2 = f(A_1 \wedge A_2)$ are also in $\mathcal{B}$.

Since B is a field and therefore a Boolean algebra, and since f preserves everything in sight, f is a Boolean homorphism. By the definition of $\mathcal{B}$, f is onto, hence it only remains to show that f is one-to-one.

In order to show that f is one-to-one, it is necessary to consider arbitrary elements $A_1, A_2 \in \mathcal{A}$ such that $A_1 \neq A_2$ and prove that $f(A_1) \neq f(A_2)$. In other words, one must show that there is some ultrafilter in $\mathcal{S}$ that contains exactly one of the pair A_1, A_2. But since $A_1 \neq A_2$, at least one of the relations $A_1 \leq A_2$ and $A_2 \leq A_1$ does not hold. We can assume that $A_1 \leq A_2$ does not hold. In that case, $A_1 \wedge \overline{A}_2 \neq \Lambda$. Hence, by the theorem proven in section 5, there must be at least one ultrafilter in $\mathcal{A}$ that contains $A_1 \wedge \overline{A}_2$. If F is such an ultrafilter, then F contains A_1, for $A_1 \wedge A_2 \leq A_1$. But F cannot contain A_2, for it contains $\overline{A}_2$. So F contains exactly one of the pair A_1, A_2. ▨

The set $\mathcal{S}$ is often called the <u>Stone space</u> of the Boolean algebra $\mathcal{A}$. Notice that the isomorphism f does map all finite meets and joins into the corresponding finite intersections and unions. It need not, however, take infinite meets and joins into the corresponding infinite set-theoretic intersections and unions.

The construction used in the Stone Representation Theorem can also be used to prove that a finite Boolean algebra is isomorphic to the field of all subsets of some set.

<u>Theorem.</u> Let $\mathcal{A}$ be a finite Boolean algebra and let $\mathcal{S}$ and $\mathcal{B}$ be as in the preceding theorem. Then $\mathcal{B}$ is the field of all subsets of $\mathcal{S}$.

<u>Proof</u>: Let F be an element of $\mathcal{S}$. Then F is an ultrafilter in $\mathcal{A}$. Since $\mathcal{A}$ is finite, F must be a principal filter; in other words, there must be a unique $A \in \mathcal{A}$ such that $F = \{A' \mid A \subset A'\}$. Since F is an ultrafilter, there cannot be any non-zero element B of $\mathcal{A}$ satisfying $B \subset A$, $B \neq A$. For if there were such an element, then the proper filter $\{A' \mid B \subset A'\}$ would properly contain F. It follows that F is the

is the only proper filter containing A; in particular, it is the only ultrafilter containing A, and hence $f(A) = \{F\}$, and $\{F\} \in \mathcal{B}$.

Since we took F to be an arbitrary element of $\mathcal{S}$, it follows that every one-element subset of $\mathcal{S}$ is in $\mathcal{B}$. Since $\mathcal{S}$ is finite, it follows that all subsets of $\mathcal{S}$ are in $\mathcal{B}$.

Notice that $\mathcal{S}$ is in a one-to-one correspondence with the set of all non-zero elements of $\mathcal{A}$ that are not majorized by any other non-zero elements. These elements were called the <u>atomic elements</u> of $\mathcal{A}$ in our earlier discussions. ▨

It should also be noted that when $\mathcal{A}$ is finite the existence of an ultrafilter containing any given proper filter can be proven by induction, so that the preceding theorem, unlike the Stone Representation theorem, does not really depend upon Zorn's lemma.

A field of subsets is called a <u>σ-field</u> if it includes the intersection and the union of any <u>countable</u> collection of its elements. Such intersections and unions will of course be the meets and joins for such collections, so a σ-field will be a σ-complete Boolean algebra. Actually, the inclusion of either all such intersections or of all such unions is sufficient to imply the inclusion of the other.

7. Closure Properties

A property of subsets of a set $\mathcal{S}$ is called a <u>closure property</u> if (i) $\mathcal{S}$ has a property, and (ii) any intersection of subsets having the given property itself has the property. Suppose that "being an X" is a closure property for subsets of a set $\mathcal{S}$.

Then if C is any given subset of $\mathcal{S}$, the collection of all X's containing C is not empty, for $\mathcal{S}$ itself is an X that contains C. The intersection of all the X's containing C is also an X containing C -- in fact it is the <u>least</u> X containing C, in the sense that it is contained in any other X containing C. Hence whenever "being an X" is a closure property for subsets of $\mathcal{S}$ and $C \subset \mathcal{S}$, one may speak of the smallest X containing C, or of the X <u>generated</u> by C.

There are several closure properties that will interest us in this essay. First, the property of "being an ideal" is a closure property for subsets of a Boolean algebra $\mathcal{A}$. Hence we may speak of the smallest ideal containing any given subset of $\mathcal{A}$, or of the ideal generated by that subset. In fact, the ideal generated by a non-empty subset $\mathcal{C} \subset \mathcal{A}$ is given by $\{A | A \in \mathcal{A}$ and $A \leq A_1 \vee \ldots \vee A_n$ for some finite collection $A_1, \ldots, A_n$ of elements of $\mathcal{C}\}$. In particular, the ideal generated by the singleton $\{A\}$ is given by $\{A' | A' \in \mathcal{A}, A' \leq A\}$.

Secondly, the property of being a filter is also a closure property for subsets of a Boolean algebra. The filter generated by a non-empty subset $\mathcal{C}$ of a Boolean algebra $\mathcal{A}$ is given by $\{A | A \in \mathcal{A}$ and $A_1 \wedge \ldots \wedge A_n \leq A$ for some finite collection $A_1, \ldots, A_n$ of elements of $\mathcal{C}\}$.

Thirdly, the property of being a subalgebra is a closure property for subsets of a Boolean algebra. The subalgebra generated by a non-empty subset $\mathcal{C}$ of a Boolean algebra $\mathcal{A}$ consists precisely of all those elements A of $\mathcal{A}$ of the form

$$A = (A_{1,1} \wedge \ldots \wedge A_{1,r_1}) \vee (A_{2,1} \wedge \ldots \wedge A_{2,r_s}) \vee \ldots \vee (A_{s,1} \wedge \ldots \wedge A_{s,r_s}),$$

where for each m, n either $A_{m,n} \in \mathcal{C}$ or $\overline{A}_{m,n} \in \mathcal{C}$. Similarly, the property of being a complete subalgebra is a closure property for subsets of a complete Boolean algebra.

Finally, the property of being a field of subsets of the set $\mathcal{S}$ is a closure property for subsets of the power set $\mathcal{P}(\mathcal{S})$. Hence we may speak of the field of subsets of $\mathcal{S}$ generated by any collection of subsets of $\mathcal{S}$. The property of being a σ-field of subsets of $\mathcal{S}$ will also be a closure property for subsets of the power set $\mathcal{P}(\mathcal{S})$.

8. Quotients of Boolean Algebras

This section is devoted to the notion of dividing a Boolean algebra by an ideal.

If A and B are two elements in a Boolean algebra $\mathcal{A}$, then the element $(A - B) \vee (B - A) = (A \vee B) - (A \wedge B)$ is called the <u>symmetric difference</u> of A and B and is denoted $A \Delta B$. Notice that if A, B, and C are elements of $\mathcal{A}$, then

$$\begin{aligned} A \,\Delta\, C &= (A \,\Delta\, C) \wedge (B \vee \overline{B}) \\ &= (A \wedge B \wedge \overline{C}) \vee (\overline{A} \wedge B \wedge C) \vee (A \wedge \overline{B} \wedge \overline{C}) \vee (\overline{A} \wedge \overline{B} \wedge C) \\ &\leq (B \wedge \overline{C}) \vee (\overline{A} \wedge B) \vee (A \wedge \overline{B}) \vee (\overline{B} \wedge C) \\ &= (A \,\Delta\, B) \vee (B \,\Delta\, C). \end{aligned}$$

Suppose we fix a proper ideal I in $\mathcal{A}$ and write "$A \approx B$" whenever A and B are in $\mathcal{A}$ and $A \,\Delta\, B \in I$. Then the relation "$\approx$" is a equivalence relation for elements of $\mathcal{A}$. In other words, it is reflexive, symmetric and transitive:

(i) If $A \in \mathcal{A}$, then $A \,\Delta\, A = \Lambda \in I$ and $A \approx A$.

(ii) If $A \approx B$, then $B \approx A$.

(iii) If $A \approx B$ and $B \approx C$, then $A \,\Delta\, C \leq (A \,\Delta\, B) \vee (B \,\Delta\, C)$;

so $A \,\Delta\, C \in I$ and $A \approx C$.

The set of equivalence classes induced by this equivalence relation is called the quotient of $\mathcal{Q}$ by I and denoted $\mathcal{Q}/I$. In other words,

$$\mathcal{Q}/I = \left\{ \{B \mid B \in \mathcal{Q} \text{ and } B \approx A\} \mid A \in \mathcal{Q} \right\}.$$

It is convenient to denote by [A] the equivalence class

$$\{B \mid B \in \mathcal{Q} \text{ and } B \approx A\}.$$

A binary relation "$\leq$" between $\mathcal{Q}/I$ and itself can be defined by setting $S_1 \leq S_2$ whenever S_1, $S_2 \in \mathcal{Q}/I$ and there are elements $A_1 \in S_1$ and $A_2 \in S_2$ such that $A_1 \leq A_2$. It is straightforward but tedious to verify that this binary relation is a partial ordering and that it makes $\mathcal{Q}/I$ into a Boolean algebra.

Furthermore, the mapping

$$f: \mathcal{Q} \to \mathcal{Q}/I : A \leadsto [A]$$

is a Boolean homomorphism. It is onto, of course, and it is called the canonical homomorphism of $\mathcal{Q}$ onto $\mathcal{Q}/I$. Notice in particular that the zero of $\mathcal{Q}/I$ is $I = [\Lambda_{\mathcal{Q}}]$, while the unit of $\mathcal{Q}/I$ is $[V_{\mathcal{Q}}]$.

When $\mathcal{Q}$ is a Boolean algebra of propositions, the quotient $\mathcal{Q}/I$ has an epistemic interpretation. Suppose, indeed, that one first contemplates $\mathcal{Q}$ without knowing whether any given propositions in it are true or false, except for V and Λ, which one knows to be true and false, respectively, Suppose further that one then learns that all the propositions in a given proper ideal I are also false. When this happens, one can regard all the propositions in I as "logically equivalent" to the impossible proposition . And we can say even more. If A, B $\in \mathcal{Q}$, then one of the pair A, B can be false and the other true only if $A \Delta B$ is true. Hence if $A \Delta B$ is in I, the knowledge that all the propositions in I are false tells one that A is true if and only if B is true -- i.e., A and B become logically equivalent. In general, then, the knowledge that all the propositions in I are false leads one to regard all the propositions in any given equivalence class as logically equivalent. Identifying propositions that are now seen as equivalent then amounts to replacing the Boolean algebra $\mathcal{Q}$ by the Boolean algebra $\mathcal{Q}/I$.

Of course, one might learn that the set J of propositions is false, where $J \subset \mathcal{Q}$ but J is not an ideal. In this case, the falsity of the propositions in J would imply the falsity of all the propositions that imply some proposition in J or the disjunction of some finite collection of propositions in J. But this latter collection of propositions, $I = \{A \mid A \leq A_1 \vee \ldots \vee A_n$ for some finite collection $A_1, \ldots, A_n$ of elements of $J\}$, is the ideal generated by J. Hence the total collection of propositions learned to be false will be an ideal, and the preceding analysis will apply. An important special case occurs when J is a singleton $\{A\}$; in this case, I is the principal ideal generated by A.

We will often be interested, of course, in the case where $\mathcal{Q} = \mathcal{P}(\mathcal{S})$ for some set $\mathcal{S}$, and the propositions all concern the true value of some parameter that takes values in $\mathcal{S}$. In this case, the ideal I usually arises by the discovery that the true value is in some subset $\mathcal{S}_0 \subset \mathcal{S}$. The ideal I of propositions learned to be false by virtue of such a discovery is precisely the principal ideal generated by $\overline{\mathcal{S}_0}$, and $\mathcal{Q}/I$ will be isomorphic to $\mathcal{P}(\mathcal{S}_0)$.

In Chapters 4 and 7, we will study another application of the notion of a quotient of a Boolean algebra by an ideal, this time to the case of a Boolean algebra of probability masses.

I will conclude this section with two theorems, one with a proof and the other without.

<u>Theorem</u>. Suppose $\mathcal{A}$ is a σ-complete Boolean algebra and I is a proper σ-ideal of $\mathcal{A}$. Then $\mathcal{A}/I$ is σ-complete, and the canonical Boolean homomorphism

$$f: \mathcal{A} \to \mathcal{A}/I : A \rightsquigarrow [A]$$

is σ-complete.

<u>Proof</u>: The σ-completeness of both $\mathcal{A}/I$ and f follows from the following fact: If $A_1, A_2, \ldots$ is a sequence of elements of $\mathcal{A}$, then $[A_1], [A_2], \ldots$ has a join in $\mathcal{A}/I$, and in fact $\vee[A_i] = [\vee A_i]$. To prove this fact, note first that by the monotonicity of the Boolean homomorphism f, $[A_i] \leq [\vee A_i]$ for all i. Hence we need only prove that if $[A_i] \leq [B]$ for all i, then $[\vee A_i] \leq [B]$. But since the Boolean homomorphism f preserves differences, $[A_i] \leq [B]$ means that $\Lambda = [A_i] - [B] = [A_i - B]$, whence $A_i - B \in I$ for each i. But I is a σ-ideal, so $\vee(A_i - B) = (\vee A_i - B) \in I$. Hence $\Lambda = f(\vee A_i - B)$ $= [\vee A_i - B] = [\vee A_i] - [B]$, or $[\vee A_i] \leq [B]$. ▨

<u>The Loomis-Sikorski Representation Theorem</u>. For every σ-complete Boolean algebra $\mathcal{A}$ there exists a σ-field of sets $\mathcal{F}$ and a σ-ideal I of $\mathcal{F}$ such that $\mathcal{A}$ is isomorphic to $\mathcal{F}/I$.

Proofs of this theorem can be found in Sikorski (p. 117), Birkhoff (p. 255), or in Halmos' <u>Lectures</u> (p. 102).

9. Independent Sums of Boolean Algebras

Suppose $A_1, \ldots, A_n$ are subalgebras of a Boolean algebra $\mathcal{Q}$. Then these n subalgebras are said to be <u>independent</u> if

$$A_1 \wedge \ldots \wedge A_n \neq \Lambda$$

whenever $A_i \varepsilon \mathcal{Q}_i$ and $A_i \neq \Lambda$ for each i, i = 1, ..., n. When $\mathcal{Q}$ is a Boolean algebra of propositions, this notion corresponds to the intuitive idea of logical independence. Indeed, two subalgebras $\mathcal{Q}_1$ and $\mathcal{Q}_2$ will be independent if and only if a non-sure proposition in one of them is never implied by a non-impossible proposition in the other. And more generally, n subalgebras $\mathcal{Q}_1, \ldots, \mathcal{Q}_n$ are independent if and only if a non-sure proposition in one of them is never implied by the conjunction of non-impossible propositions from the others.

Some of the implications of independence are developed by the following propositions.

<u>Theorem</u>. Suppose $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are independent subalgebras of a Boolean algebra $\mathcal{Q}$. Suppose $A_1, A_1' \varepsilon \mathcal{Q}_1, A_2, A_2' \varepsilon \mathcal{Q}_2$, and $\Lambda \neq A_1' \wedge A_2' \leq A_1 \wedge A_2$. Then $A_1' \leq A_1$ and $A_2' \leq A_2$.

<u>Proof:</u>

$$\begin{aligned} A_1' \wedge A_2' &= [(A_1' \wedge A_1) \vee (A_1' - A_1)] \wedge A_2' \\ &= [(A_1' \wedge A_1) \wedge A_2'] \vee [(A_1' - A_1) \wedge A_2'] \end{aligned}$$

is a disjoint partition. But $A_1' \wedge A_2' = (A_1' \wedge A_2') \wedge A_1 = (A_1' \wedge A_1) \wedge A_2'$. So $(A_1' - A_1) \wedge A_2' = \Lambda$. Hence, by independence, $A_1' - A_1 = \Lambda$. This means $A_1' \leq A_1$. Similarly, $A_2' \leq A_2$. ▨

Corollary. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$. Suppose $A \in \mathcal{A}$, $A \neq \Lambda$, and $A = A_1' \wedge A_2' = A_1 \wedge A_2$, where $A_1, A_1' \in \mathcal{A}_1$, and $A_2, A_2' \in \mathcal{A}_2$. Then $A_1 = A_1'$ and $A_2 = A_2'$.

Corollary. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$, that $A_1, A_1' \in \mathcal{A}$, that $A_2, A_2' \in \mathcal{A}_2$, and that $\Lambda \neq A_1' \wedge A_2' < A_1 \wedge A_2$. Then $A_1' \leq A_1$, $A_2' \leq A_2$ and either $A_1 < A_1'$ or $A_2 < A_2'$.

Theorem. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$ and generate $\mathcal{A}$. Suppose $A_o \in \mathcal{A}$, $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$, and $\Lambda \neq A_o < A \wedge B$. Then there exists an integer $s \geq 1$ and elements $A_1, \ldots, A_s \in \mathcal{A}_1$ and elements $B_1, \ldots, B_s \in \mathcal{A}_2$ such that $A_i \leq A$ and $B_i \leq B$ for $i = 1, \ldots, s$, and

$$A_o = (A_1 \wedge B_1) \vee \ldots \vee (A_s \wedge B_s) \qquad (1)$$

Proof: Since $\mathcal{A}$ is generated by $\mathcal{A}_1 \cup \mathcal{A}_2$, the element A_o must, by section 7, be of the form

$$A_o = (A_{1,1} \wedge \ldots \wedge A_{1,r_1}) \vee (A_{2,1} \wedge \ldots \wedge A_{2,r_2}) \vee \ldots \vee$$

$$(A_{s,1} \wedge \ldots \wedge A_{s,r_s}),$$

where for each m, n either $A_{m,n} \in \mathcal{A}_1 \cup \mathcal{A}_2$ or $\overline{A}_{m,n} \in \mathcal{A}_1 \cup \mathcal{A}_2$. And since $\mathcal{A}_1$ and $\mathcal{A}_2$ are subalgebras, this means that A must be of the form (1). Since we may assume that the $A_i \wedge B_i$ are non-zero, the fact that the $A_i \leq A$ and the $B_i \leq B$ follows from the preceding theorem. ▨

Theorem. Suppose $A_1 \in \mathcal{Q}_1$, $A_1 \neq \Lambda$, and $A_2, A_2' \in \mathcal{Q}_2$, where $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are independent subalgebras of $\mathcal{Q}$. And suppose that $A_1 \wedge A_2 \leq A_2'$. Then $A_2 \leq A_2'$.

Proof: If $A_1 \wedge A_2 \leq A_2'$, then $A_1 \wedge A_2 \wedge \overline{A_2'} = \Lambda$, whence $A_2 \wedge \overline{A_2'} = \Lambda$, or $A_2 \leq A_2'$. ▨

Theorem. Suppose $A_1, A_1' \in \mathcal{Q}_1$ and $A_2, A_2' \in \mathcal{Q}_2$, where $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are independent subalgebras of $\mathcal{Q}$. And suppose that $A_1 \wedge A_2 \leq A_1' \vee A_2'$. Then $A_1 \leq A_1'$ or $A_2 \leq A_2'$.

Proof: $(A_1 \wedge A_2) \wedge \overline{A_1'} \leq (A_1' \vee A_2') \wedge \overline{A_1'} \leq A_2'$. Hence $(A_1 \wedge \overline{A_1'}) \wedge A_2 \leq A_2'$. So by the preceding theorem, either $A_1 \wedge A_1' = \Lambda$ and $A_1 \leq A_1'$, or $A_2 \leq A_2'$. ▨

Now suppose we have a collection $\mathcal{Q}_1, \ldots, \mathcal{Q}_n$ of Boolean algebras and that we conceive of them in the first instance as having nothing to do with each other. Then we might still wish to think of them as independent subalgebras of some larger Boolean algebra. If, for example, they are Boolean algebras of propositions, each dealing with a different subject, then we might wish to embed them in an overall Boolean algebra of propositions which would also contain propositions of the form (1) -- propositions dealing with more than one subject at a time.

Abstractly, what would it mean to embed the Boolean algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ as independent subalgebras of a larger algebra $\mathcal{A}$? Well, one would require a collection of isomorphisms $f_1, \ldots, f_n$ such that for each i, $i = 1, \ldots, n$, f_i is an isomorphism of $\mathcal{A}_i$ into $\mathcal{A}$; and one would require that the images $f_1(\mathcal{A}_1), \ldots, f_n(\mathcal{A}_n)$ should be independent subalgebras of $\mathcal{A}$.

Now we might carry out such an embedding and then find out that the algebra $\mathcal{A}$ is larger than it needs to be. In other words, the subalgebra of $\mathcal{A}$ generated by $f_1(\mathcal{A}_1) \cup \ldots \cup f_n(\mathcal{A}_n)$ might be a proper subalgebra of $\mathcal{A}$. If this occurs, though, we can replace $\mathcal{A}$ by that proper subalgebra and still have an embedding -- one which would now be "minimal". This leads us to the following definition:

Definition. Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_n, \mathcal{A}$ are Boolean algebras and that for each i, $i = 1, \ldots, n$, $f_i: \mathcal{A}_i \to \mathcal{A}$ is an isomorphism into. Suppose further that $f_1(\mathcal{A}_1), \ldots, f_n(\mathcal{A}_n)$ are independent subalgebras of $\mathcal{A}$ and that $f_1(\mathcal{A}_1) \cup \ldots \cup f_n(\mathcal{A}_n)$ generates $\mathcal{A}$. Then $(f_1, \ldots, f_n; \mathcal{A})$ is called an <u>independent sum</u> of $\mathcal{A}_1, \ldots, \mathcal{A}_n$.

As is turns out, an independent sum of $\mathcal{A}_1, \ldots, \mathcal{A}_n$ always exists. (See Sikorski, pp. 40-41.) Furthermore, all such independent sums are isomorphic, in the sense that for any two of them, say $(f_1, \ldots, f_n, \mathcal{A})$ and $(f_1', \ldots, f_n', \mathcal{A}')$ there is an isomorphism h of $\mathcal{A}$ onto $\mathcal{A}'$ such that $f_i' = h \circ f_i$ for each i, $i = 1, \ldots, n$. Hence the independent sum of a collection of Boolean algebras is essentially unique.

When the Boolean algebras are thought of as Boolean algebras of propositions this uniqueness is reassuring, for each element of the independent sum is given an intuitive interpretation by formula (1).

Often, of course, each of the Boolean algebras $\mathcal{A}_i$ is conceived of as a field of subsets of some set $\mathcal{S}_i$. In this case, the sum can be thought of as a field of subsets of the Cartesian product $\mathcal{S} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_n$. Indeed, the isomorphisms f_i are defined by $f_i(A) = \mathcal{S}_1 \times \ldots \times \mathcal{S}_{i-1} \times A \times \mathcal{S}_{i+1} \times \ldots \mathcal{S}_n$, and the sum $\mathcal{A}$ is then the field of subsets of $\mathcal{S}$ generated by the collection $f_1(\mathcal{A}_1) \cup \ldots \cup f_n(\mathcal{A}_n)$.

In general, I will denote by $\mathcal{A}_1 \oplus \ldots \oplus \mathcal{A}_n$ the independent sum of the collection $\mathcal{A}_1, \ldots, \mathcal{A}_n$. Properly speaking, the Boolean algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ will only be isomorphic to independent subalgebras of $\mathcal{A}_1 \oplus \ldots \oplus \mathcal{A}_n$. But I will often speak of them as if they actually were subalgebras of $\mathcal{A}_1 \oplus \ldots \oplus \mathcal{A}_n$. This practice is often quite convenient and does not seem to cause confusion.

CHAPTER 4. THE MATHEMATICAL REPRESENTATION OF OUR PROBABILITY

Now that we have a better technical grasp of the theory of Boolean algebras, we can improve the mathematical representation of our intuitive "probability masses." In this chapter, that representation is improved and developed.

1. Probability Algebras

In section 1 of Chapter 2, I gave the following definition of a measure on a Boolean algebra:

Definition. If $\mathcal{M}$ is a Boolean algebra, then a function $\mu : \mathcal{M} \to [0, 1]$ is a measure if

(1) $\mu(\Lambda_{\mathcal{M}}) = 0$,

(2) $\mu(V_{\mathcal{M}}) = 1$,

and (3) $\mu(M_1) + \mu(M_2) = \mu(M_1 \vee M_2)$ whenever $M_1, M_2 \in \mathcal{M}$ and $M_1 \wedge M_2 = \Lambda_{\mathcal{M}}$.

I then declared that any Boolean algebra $\mathcal{M}$ with an accompanying measure μ could be called a measure algebra -- the intuitive idea being that the elements of $\mathcal{M}$ could be regarded as probability masses. But as I later observed, there are properties that our "probability masses" ought ideally to have that are not imposed by this definition. At the end of Chapter 2, I listed three such properties: positivity, completeness and complete additivity. Now that we have a stronger technical grasp of the theory of Boolean algebras, we can describe these properties more precisely.

Definition. A measure algebra $(\mathcal{M}, \mu)$ is called a probability algebra if

(1) $(\mathcal{M}, \mu)$ is positive: If $M \in \mathcal{M}$ and $M \neq \Lambda_{\mathcal{M}}$, then $\mu(M) > 0$;

(2) $\mathcal{M}$ is complete;

(3) $(\mathcal{M}, \mu)$ is completely additive: If $\mathcal{C} \subset \mathcal{M}$ and the elements of $\mathcal{C}$ are pairwise disjoint, then $\sum_{M \in \mathcal{C}} \mu(M) = \mu(\vee \mathcal{C})$.

The conditions listed in this definition add up to a rather strong package, and the reader might well question whether there even exist any probability algebras. As it turns out, though, there are quite a few of them. In fact, for every measure algebra $(\mathcal{M}, \mu)$, there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism $h: \mathcal{M} \to \mathcal{N}$ such that $\mu = \nu \circ h$. This fact will be proven in the next section.

The condition of complete additivity may require some explanation. The symbol $\sum_{M \in \mathcal{C}} \mu(M)$ ostensibly requires the addition of a number of non-negative quantities that may be infinite and perhaps even uncountably infinite. But the sum of an uncountable number of positive quantities does not really exist, or at any rate must be considered infinite, while $\sum_{M \in \mathcal{C}} \mu(M) = \mu(\vee \mathcal{C})$ is supposed to be finite. Hence the condition of complete additivity requires in particular that at most a countable number of the elements of $\mathcal{C}$ can have non-zero measure. If $(\mathcal{M}, \mu)$ is also positive, then this means that only a countable number of the elements of $\mathcal{M}$ can be non-zero. Hence we may conclude that any collection of disjoint non-zero elements in a probability algebra must be countable.

These considerations make the following theorem less surprising than it seems at first:

Theorem. Suppose $(\mathcal{M}, \mu)$ is a measure algebra and satisfies the following conditions:

(i) $\mathcal{M}$ is σ-complete.

(ii) $\mathcal{M}$ is positive.

(iii) $(\mathcal{M}, \mu)$ is countably additive: If $\mathcal{C} \subset \mathcal{M}$, $\mathcal{C}$ is countable and the elements of $\mathcal{C}$ are pairwise disjoint, then $\sum_{M \in \mathcal{C}} \mu(M) = \mu(\vee \mathcal{C})$.

Then $(\mathcal{M}, \mu)$ is a probability algebra.

Proof: It follows from the (finite) additivity of μ that $\mathcal{M}$ cannot contain, for any positive integer n, as many as n elements of measure greater than $1/n$. Hence any disjoint set of elements of $\mathcal{M}$ must be countable.

Let $\mathcal{C}$ be any non-empty subset of $\mathcal{M}$. Then it follows from the second theorem of section 5 of Chapter 3 that there exists a disjoint subset $\mathcal{D}$ of $\mathcal{M}$ with exactly the same upper bounds as $\mathcal{C}$. Since $\mathcal{D}$ is disjoint, it must be countable; and since $\mathcal{M}$ is σ-complete, $\mathcal{D}$ must have a least upper bound or join. The same element will also be the join of $\mathcal{C}$. Hence any non-empty subset of $\mathcal{M}$ has a join. The existence of meets follows; $\mathcal{M}$ is complete. And complete additivity follows from countable additivity. ▨

The fact that any collection of disjoint non-zero elements in a probability algebra must be countable also gives the following interesting result:

Theorem: Suppose $\mathcal{M}$ is a probability algebra, $M \in \mathcal{M}$, $\{M_\gamma\}_{\gamma \in \Gamma}$ is a collection of elements of $\mathcal{M}$ and $M = \vee M_\gamma$. Then there exists a disjoint sequence $M_1, M_2, \ldots$ of elements of $\mathcal{M}$ such that (i) $M = \vee M_i$ and (ii) for each i there exists $\gamma \in \Gamma$ such that $M_i \leq M_\gamma$.

Proof: By the second theorem of section 5 of Chapter 3, there exists a disjoint subset $\mathcal{D}$ of $\mathcal{M}$ with the same set of upper bounds as $\{M_\gamma\}_{\gamma \in \Gamma}$, and such that for each $D \in \mathcal{D}$ there exists $\gamma \in \Gamma$ with $D \leq M_\gamma$. But since $\mathcal{D}$ is disjoint, it can have at most a countably infinite number of non-zero elements. Denoting these by M_1, M_2, $\ldots$ yields the theorem. ▨

A probability algebra also has very strong properties of the type that are often called continuity properties. For a start, the measures of a monotone sequence of elements of the probability algebra will converge to the measure of the limit of the monotone sequence, as shown in the following theorem.

Theorem. Suppose $(\mathcal{M}, \mu)$ is a probability algebra. Then for any monotonically increasing sequence $M_1 \leq M_2 \leq \ldots$ in $\mathcal{M}$,

$$\mu(\vee M_i) = \sup_i \mu(M_i).$$

And for any monotonically decreasing sequence $M_1 \geq M_2 \geq \ldots$ in $\mathcal{M}$,

$$\mu(\wedge M_i) = \inf_i \mu(M_i).$$

Proof: First suppose $M_1 \leq M_2 \leq \ldots$ is an increasing sequence, and set $M_o = \Lambda$. Then it is easily verified from the definition of join that

$$\bigvee_{i=1}^{\infty} M_i = \bigvee_{i=1}^{\infty} (M_i - M_{i-1}).$$

But the elements in the join on the right-hand side are disjoint; hence

$$\mu(\bigvee_{i=1}^{\infty} M_1) = \mu(\bigvee_{i=1}^{\infty} [M_i - M_{i-1}]) = \sum_{i=1}^{\infty} \mu(M_i - M_{i-1})$$

$$= \sup_n \sum_{i=1}^{n} \mu(M_i - M_{i-1}) = \sup_n \mu(\bigvee_{i=1}^{n} (M_i - M_{i-1}))$$

$$= \sup_n \mu(M_n).$$

In the case where $M_1 \geq M_2 \geq \ldots$ is a decreasing sequence, $\overline{M_1} \leq \overline{M_2} \leq \ldots$ an increasing sequence, and $\wedge M_i = \overline{\vee \overline{M_i}}$. Hence,

$$\mu(\wedge M_i) = \mu(\overline{\vee \overline{M_i}}) = 1 - \mu(\vee \overline{M_i}) = 1 - \sup_i \mu(\overline{M_i})$$

$$= 1 - \sup_i (1 - \mu(M_i)) = \inf_i \mu(M_i).$$

The proof just given uses the property of additivity only for countable subsets of $\mathcal{M}$. Using the full force of the property of complete additivity, we can prove a rather stronger statement, the formulation of which requires the notion of a net.

A non-empty subset $\mathcal{B}$ of a Boolean algebra $\mathcal{Q}$ is called a downward net in $\mathcal{Q}$ if for every pair of elements A, B $\in \mathcal{B}$ there exists an element C $\in \mathcal{B}$ such that $C \leq A \wedge B$. A non-empty subset $\mathcal{B}$ of a Boolean algebra is called an upward net in $\mathcal{Q}$ if for every pair of elements A, B $\in \mathcal{B}$ there exists an element $C \in \mathcal{B}$ such that $A \vee B \leq C$. Notice that a filter is a downward net and that an ideal is an upward net.

<u>Theorem</u>. Suppose $(\mathcal{M}, \mu)$ is a probability algebra. Then for any downward net $\mathcal{B} \subset \mathcal{M}$,

$$\mu(\wedge \mathcal{B}) = \inf_{B \in \mathcal{B}} \mu(B).$$

And for any upward net $\mathcal{B} \subset \mathcal{M}$,

$$\mu(\vee \mathcal{B}) = \sup_{B \in \mathcal{B}} \mu(B).$$

<u>Proof</u>: Consider first the case of a downward net $\mathcal{B}$. Since $\wedge \mathcal{B} \leq B$ for all $B \in \mathcal{B}$, $\mu(\wedge \mathcal{B}) \leq \inf_{B \in \mathcal{B}} \mu(B)$. Choose a decreasing sequence $B_1 \geq B_2 \geq B_3, \ldots$ in $\mathcal{B}$ such that $\inf_i \mu(B_i) = \inf_{B \in \mathcal{B}} \mu(B)$. Then by the preceding theorem, $\mu(\wedge_i B_i) = \inf_i \mu(B_i) = \inf_{B \in \mathcal{B}} \mu(B)$. Now suppose $\mu(\wedge \mathcal{B}) < \inf_{B \in \mathcal{B}} \mu(B)$. Then $\wedge \mathcal{B}$ is a proper subelement of $\wedge B_i$. This implies the existence of some element $M_1 \in \mathcal{B}$ such that $\wedge B_i$ is not a subelement of M_1, or $\wedge B_i - M_1 \neq \Lambda$. Denote $\mu(\wedge B_i - M_1) = \epsilon > 0$. We can choose an integer K so that $\mu(B_K - \wedge B_i) = \mu(B_K) - \mu(\wedge B_i) < \epsilon/2$, and if we then choose $M_2 \in \mathcal{B}$ so that $M_2 \leq B_K \wedge M_1$, we will have

$$\mu(\wedge B_i - M_2) \geq \epsilon$$

and

$$\mu(M_2 - \wedge B_i) < \epsilon/2.$$

This implies that $\mu(\wedge B_i) > \mu(M_2)$, contradicting the assumption that $\mu(\wedge B_i) = \inf_{B \in \mathcal{B}} \mu(B)$. Hence $\mu(\wedge B) = \inf_{B \in \mathcal{B}} \mu(B)$. ▨

If $(\mathcal{M}, \mu)$ is a probability algebra and $\mathcal{N}$ is a complete subalgebra of $\mathcal{M}$, then $(\mathcal{N}, \mu | \mathcal{N})$ will be a probability algebra. We can describe this situation by saying that $(\mathcal{N}, \mu | \mathcal{N})$ is embedded in $(\mathcal{M}, \mu)$. More generally, if $(\mathcal{M}, \mu)$ and $(\mathcal{N}, \nu)$ are probability algebras, then an isomorphism θ of $\mathcal{M}$ into $\mathcal{N}$ is called an embedding or isomorphism of $(\mathcal{M}, \mu)$ into $(\mathcal{N}, \nu)$ if $\mu = \nu \circ \theta$. And of course if θ is also onto, then it is called an isomorphism between the two probability algebras, and they are said to be isomorphic.

2. Constructing Probability Algebras

In this section, I will show that for any measure algebra $(\mathcal{M}, \mu)$ there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism $h: \mathcal{M} \rightarrow \mathcal{N}$ such that $\mu = \nu \circ h$. One important tool in this demonstration is Carathéodory's Extension Theorem, a standard theorem in measure theory that I will state and use without proof.

Carathéodory's Extension Theorem. Suppose $\mathcal{F}$ is a field of subsets of a set $\mathcal{S}$ and $\delta: \mathcal{F} \rightarrow [0,1]$ satisfies

(1) $\delta(\phi) = 0$

(2) $\delta(\mathcal{S}) = 1$

(3) $\delta(S_1) + \delta(S_2) = \delta(S_1 \cup S_2)$ whenever $S_1, S_2 \in \mathcal{F}$ and $S_1 \cap S_2 = \phi$.

(4) If $S_1 \supset S_2 \supset \ldots$ and $\bigcap_{i=1}^{\infty} S_i = \phi$, then $\lim_{i \to \infty} \delta(S_i) = 0$.

Let $\mathcal{F}^*$ be the σ- field of subsets of $\mathcal{S}$ generated by $\mathcal{F}$. Then there exists an extension δ^* of δ to $\mathcal{F}^*$ such that

(5) $\sum_{i=1}^{\infty} \delta^*(S_i) = \delta^*(\bigcup_{i=1}^{\infty} S_i)$ for all disjoint sequences $S_1, S_2, \ldots$ of elements of $\mathcal{F}^*$.

<u>Lemma 1</u>. Suppose $(\mathcal{M}, \mu)$ is a measure algebra. Then there exists a measure algebra $(\mathcal{F}^*, \delta^*)$ that is σ - complete and countably additive and a Boolean homomorphism $f: \mathcal{M} \rightarrow \mathcal{F}^*$ such that $\mu = \delta^* \circ f$.

<u>Proof</u>: Let $f_o: \mathcal{M} \rightarrow \mathcal{F}$ be the isomorphism established by the Stone Representation Theorem; $\mathcal{F}$ being a field of subsets of the set $\mathcal{S}$ of all ultrafilters in $\mathcal{M}$, with $f_o(M) = \{F \mid F \in \mathcal{S} \text{ and } M \in F\}$. Set $\delta = \mu \circ f_o^{-1}$.

Then $\mathcal{F}$ and δ obviously satisfy (1), (2) and (3) in the hypothesis of Caratheodory's Extension Theorem. In fact, it also satisfies (4). To see this, let $S_1 \supset S_2 \supset \ldots$ be a decreasing sequence in with $\bigcap S_i = \phi$, and set $M_i = f^{-1}(S_i)$. Then $M_1 \geq M_2 \geq \ldots$, and

$$\phi = \bigcap S_i = \bigcap \{F \mid F \in \mathcal{S} \text{ and } M_i \in F\}$$
$$= \{F \mid F \in \mathcal{S} \text{ and } M_i \in F \text{ for all } i\}.$$

Now set $F_o = \{M \mid M_i \leq M \text{ for some } i\}$. It is easily seen that F_o is a filter. Furthermore, F_o is improper. For if it were proper, it would be contained in an ultrafilter F_1; F_1 would then contain all the M_i and hence would be in S_i, contradicting the assumption that $\bigcap S_i = \phi$. So F_o is improper and thus contains $\Lambda_{\mathcal{M}}$. But this implies that $M_i = \Lambda_{\mathcal{M}}$ for some i and hence for all $j \geq i$. Thus $\lim_{i \to \infty} \delta(S_i) = \lim_{i \to \infty} \mu(M_i) = 0$.

So by Carathéodory's Extension Theorem, δ can be extended to a countably additive measure δ^* on the σ - field $\mathcal{F}^*$ generated by $\mathcal{F}$. Evidently, $(\mathcal{F}^*, \delta^*)$ is a σ - complete and countably additive probability algebra. If we denote by i the identity mapping from $\mathcal{F}$ into $\mathcal{F}^*$ then $f = i \circ f_o$ is a Boolean homomorphism of $\mathcal{M}$ into $\mathcal{F}^*$. Furthermore, $\mu = \delta \circ f_o = \delta^* \circ i \circ f_o = \delta^* \circ f$. ∎

Lemma 2. Suppose $(\mathcal{F}^*, \delta^*)$ is a σ-complete and countably additive measure algebra. Then there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism $g: \mathcal{F}^* \to \mathcal{N}$ such that $\delta^* = \nu \circ g$.

Proof: Consider the set $I = \left\{M \middle| M \in \mathcal{F}^*, \mu(M) = 0\right\}$. It is easily shown that I is a proper ideal in $\mathcal{F}^*$. Hence one may construct the quotient $\mathcal{N} = \mathcal{F}^*/I$ and the Boolean homomorphism $g: \mathcal{F}^* \to \mathcal{N}$: $M \rightsquigarrow \left\{N \middle| N \in \mathcal{F}^*, N \Delta M \in I\right\}$ as in section 8 of Chapter 3. Recall that each element of $\mathcal{N}$ is an equivalence class of elements of $\mathcal{F}^*$. If M and N are both in the equivalence class $E \in \mathcal{N}$, then $N \Delta M \in I$, whence $\delta^*(N \Delta M) = 0$ and $\delta^*(N) = \delta^*(M)$. Hence one may define a function $\nu: \mathcal{N} \to [0, 1]$ by setting $\nu(E) = \delta^*(M)$ for any $M \in E$. Evidently, $\nu = \delta^* \circ g$.

Since $\Lambda_{\mathcal{F}^*}$ is in the equivalence class $\Lambda_{\mathcal{N}}$ and $V_{\mathcal{F}^*}$ is in the equivalence class $V_{\mathcal{N}}$, $\nu(\Lambda_{\mathcal{N}}) = 0$ and $\nu(V_{\mathcal{N}}) = 1$. And if $E_1, E_2 \in \mathcal{N}$ with $E_1 \wedge E_2 = \Lambda_{\mathcal{N}}$, then choosing $M \in E_1$ and $N \in E_2$ gives $g(M \wedge N) = g(M) \wedge g(N) = E_1 \wedge E_2 = \Lambda_{\mathcal{N}} = I$, whence $\delta^*(M \wedge N) = 0$. Hence $\nu(E_1) + \nu(E_2) = \delta^*(M) + \delta^*(N) = \delta^*(M \vee N) = \nu(g(M \vee N)) = \nu(g(M) \vee g(N)) = \nu(E_1 \vee E_2)$. So $(\mathcal{N}, \nu)$ is a probability algebra.

Furthermore, $(\mathcal{N}, \nu)$ is positive. For if $\nu(E) = 0$, then choosing $M \in E$ gives $\delta^*(M) = 0$, whence $M \in I$ and $E = I = \Lambda_{\mathcal{N}}$.

Now I is a σ-ideal. In order to prove this, take any countable collection $A_1, A_2, \ldots$ of elements of I and set $B_i = A_i - \bigvee_{j<i} A_j$, and note that $\vee A_i = \vee B_i$ and $\delta^*(\vee B_i) = \Sigma\, \delta^*(B_i) = 0$. Since I is a σ-ideal, the quotient $\mathcal{N}$ is σ-complete and g preserves countable joins. From the fact that g preserves countable joins, one may deduce that $(\mathcal{N}, \nu)$ is countably additive. It then follows from the first theorem in section 1 that $(\mathcal{N}, \nu)$ is a probability algebra. ▨

Theorem. Suppose $(\mathcal{M}, \mu)$ is a measure algebra. Then there exists a probability algebra $(\mathcal{N}, \nu)$ and a Boolean homomorphism h: $\mathcal{M} \to \mathcal{N}$ such that $\mu = \nu \circ h$.

Proof: The theorem follows directly from the constructions in the two lemmas. For setting $h = g \circ f$, we have $\mu = \delta^* \circ f = \nu \circ g \circ f = \nu \circ h$. ▨

In the proof of the second lemma above, we took a σ-field of subsets that had a countably additive measure and divided it by the ideal consisting of those of its elements with zero measure. As we saw, such a process necessarily results in a probability algebra. With the help of the Loomis-Sikorski Representation Theorem, it is easily shown that any probability algebra can be represented as such a quotient.

Theorem. Suppose $(\mathcal{M}, \mu)$ is a probability algebra. Then there exists a set $\mathcal{S}$, a σ-field $\mathcal{F}$ of subsets of $\mathcal{S}$, a countably additive measure ν on $\mathcal{F}$, and an isomorphism i of $\mathcal{M}$ onto the quotient of $\mathcal{F}$ by the σ-ideal of sets of measure zero such that $\nu(F) = \mu(M)$ whenever $F \in i(M)$.

Proof: The Loomis-Sikorski Representation Theorem supplies us with a σ-field $\mathcal{F}$ of subsets of a set $\mathcal{S}$, σ-ideal I of $\mathcal{F}$ and an isomorphism i of $\mathcal{M}$ onto $\mathcal{F}/I$. Hence we need only verify that the function $\nu: \mathcal{F} \to [0, 1]$ defined by $\nu(F) = \mu(M)$ whenever $F \in i(M)$ is countably additive measure and that I consists precisely of the sets F for which $\nu(F) = 0$.

The second part is easy: the sets F for which $\nu(F) = 0$ are precisely those in $i(\Lambda_{\mathcal{M}}) = \Lambda_{\mathcal{F}/I} = I$. On the other hand, $\mathcal{S} \in i(V_{\mathcal{M}})$, so $\nu(\mathcal{S}) = 1$. Hence we need only show countable additivity

for ν. But the canonical homomorphism $f: \mathcal{F} \rightarrow \mathcal{F}/I$ is σ-complete. So if we take any disjoint sequence $S_1, S_2, \ldots$ of elements of $\mathcal{F}$, we have $\nu(\cup S_i) = \mu\left(i^{-1}\left(f(\cup S_i)\right)\right) = \mu\left(i^{-1}\left(\vee f(S_i)\right)\right) = \mu\left(\vee i^{-1}\left(f(S_i)\right)\right)$
$= \sum \mu\left(i^{-1}\left(f(S_i)\right)\right) = \sum \nu(S_i)$. ▨

3. Standard Representations for Belief Functions

It follows from the preceding theorem that any belief function can be represented by an allocation into a probability algebra. Suppose, indeed, that we have a belief function $\mathrm{Bel}: \mathcal{Q} \rightarrow [0,1]$, a measure algebra $(\mathcal{M}, \mu)$ and an allocation $\rho_o: \mathcal{Q} \rightarrow \mathcal{M}$ such that $\mathrm{Bel} = \mu \circ \rho_o$. Then using the probability algebra $(\mathcal{N}, \nu)$ and the Boolean homomorphism $h: \mathcal{M} \rightarrow \mathcal{N}$ supplied by our theorem, we may set $\rho = h \circ \rho_o$. The mapping $\rho: \mathcal{Q} \rightarrow \mathcal{N}$ will then be an allocation into the probability algebra $(\mathcal{N}, \nu)$ and it will represent Bel; for $\mathrm{Bel} = \mu \circ \rho_o = \nu \circ h \circ \rho_o = \nu \circ \rho$.

In the sequel, I will generally mean an allocation into a probability algebra whenever I use the term "allocation of probability." When confusion is possible, I will use the word <u>standard</u> to specifically refer to allocations into probability algebras. I will say that an allocation into a probability algebra is a <u>standard allocation</u>, and I will say that it is a <u>standard representation</u> of the belief function it represents.

As we will see, the existence of standard representations will often facilitate our thinking about allocations of probability and belief functions.

It should be noted that when $\rho: \mathcal{Q} \rightarrow \mathcal{M}$ is a standard representation for the belief function $\mathrm{Bel}: \mathcal{Q} \rightarrow [0,1]$ and $\mathcal{Q}_o$ is a subalgebra of $\mathcal{Q}$, $\rho|\mathcal{Q}_o$ will be a standard representation for the belief function $\mathrm{Bel}|\mathcal{Q}_o$ on $\mathcal{Q}_o$.

4. Quotients of Probability Algebras

One of our fundamental conventions is that the measure of our total probability mass should equal one. It sometimes happens, though, that we want to discard a given probability mass and to regard the probability mass that is left over as our total probability; in this circumstance the measure of our total probability will decrease unless we "renormalize" it. In this section I will briefly describe the process of discarding a probability mass and renormalizing the measure of the remainder.

Essentially, to discard a probability mass means (1) to put the null probability mass in its place and (2) to deduct its contribution from every probability mass to which it contributed. In symbols, the discarding of the probability mass M from a probability algebra $(\mathcal{M}, \mu)$ involves replacing every probability mass $M' \epsilon \mathcal{M}$ by $M' \wedge \overline{M} = M' - M$. Or, to put it a different way, it means identifying all pairs M', M'' of probability masses in $\mathcal{M}$ for which $M' - M = M'' - M$.

But this is precisely what is done when $\mathcal{M}$ is divided by the principal ideal I generated by M. For under that division M' goes into the equivalence class $\{M'' | M' \triangle M'' \epsilon I\} = \{M'' | M'' \triangle M' \leq M\} = \{M'' | M'' - M = M' - M\}$. Hence discarding a probability mass means dividing by a principal ideal.

Denote by f the canonical homomorphism of $\mathcal{M}$ onto $\mathcal{M}/I$. Then what measure should be assigned to a given element $f(M') \epsilon \mathcal{M}/I$? Well, $M' = (M' - M) \vee (M' \wedge M)$ and $M' \wedge M$ is being discarded; so $M' - M$ is what is left of M', and it would be natural to adopt $\mu(M' - M)$ as the measure of f(M'). But this procedure will result in a measure of $\mu(\bigvee_{\mathcal{M}} - M) = \mu(\overline{M}) = 1 - \mu(M)$ for the unit $\bigvee_{\mathcal{M}}/I = f(\bigvee_{\mathcal{M}})$. If $\mu(\overline{M}) > 0$ -- i.e., if $M \neq \bot_{\mathcal{M}}$, then this conflicts with the requirement that the measure of $\bigvee_{\mathcal{M}}/I$ should be one. We can correct this difficulty by multiplying all the quantities $\mu(M' - M)$ by a constant in order to increase the measure of $\bigvee_{\mathcal{M}}/I$ to one. The appropriate constant is, of course, $1/(1-\mu(M))$. In other words, we define a measure ν on $\mathcal{M}/I$ by

$$\nu(f(M')) = \frac{1}{1 - \mu(M)} \mu(M' - M).$$

It is easily verified that this is indeed a measure on $\mathcal{M}/I$. In fact, $(\mathcal{M}/I, \nu)$ is a probability algebra, provided only that $M \neq \bigvee_{\mathcal{M}}$. In the sequel, I will refer to $(\mathcal{M}/I, \nu)$ as the probability algebra obtained from $(\mathcal{M}, \mu)$ by discarding M.

5. Orthogonal Sum of Probability Algebras

As I mentioned above, if $(\mathcal{M}, \mu)$ is a probability algebra and $\mathcal{N}$ is a complete subalgebra of $\mathcal{M}$, then $(\mathcal{N}, \mu|\mathcal{N})$ is a probability algebra and is said to be embedded in $(\mathcal{M}, \mu)$. Now suppose that $\mathcal{M}_1, \ldots, \mathcal{M}_n$ are independent complete subalgebras of $\mathcal{M}$. Then they are said to be <u>orthogonal</u> if

$$\mu(M_1 \wedge \ldots \wedge M_n) = \mu(M_1) \ldots \mu(M_n)$$

whenever $M_i \in \mathcal{M}_i$ for each i, i = 1, ..., n.

In the sequel we will sometimes deal with a collection of probability algebras that are conceived of as having nothing to do with one another and yet which we wish to embed as orthogonal subalgebras of a single overall probability algebra. In this section, we will see how this can be done.

Definition. Suppose $(\mathcal{M}_1, \mu_1)$, ..., $(\mathcal{M}_n, \mu_n)$, $(\mathcal{M}, \mu)$ are probability algebras and that for each i, i = 1, ..., n, $f_i : \mathcal{M}_i \to \mathcal{M}$ is a complete isomorphism into with $\mu_i = \mu \circ f_i$. Suppose further that $f_1(\mathcal{M}_1)$, ..., $f_n(\mathcal{M}_n)$ are independent and orthogonal subalgebras of $\mathcal{M}$. Then $(f_1, \ldots, f_n; (\mathcal{M}, \mu))$ is called an orthogonal sum of $(\mathcal{M}_1, \mu_1)$, ..., $(\mathcal{M}_n, \mu_n)$.

The rest of this section is devoted to showing that an orthogonal sum exists for any finite collection of probability algebras. This will be done by appealing to the construction of "product measures" in measure theory. In particular, I will appeal to the following theorem, which is a long-winded version of the assertion that product measures exist:

Theorem. Let $(\mathcal{S}_1, \mathcal{F}_1, \nu_1)$, ..., $(\mathcal{S}_n, \mathcal{F}_n, \nu_n)$ be "measure spaces." In other words, for each i, i = 1, ..., n, $\mathcal{F}_i$ is a σ-field of subsets of the set $\mathcal{S}_i$ and ν_i is a countably additive measure on $\mathcal{F}_i$. Denote by $\mathcal{S}$ the Cartesian product $\mathcal{S}_1 \times \ldots \times \mathcal{S}_n$. And for each i, i = 1, ..., n, define a mapping $k_i : \mathcal{F}_i \to P(\mathcal{S})$ by setting $k_i(A) = \mathcal{S}_1 \times \ldots \times \mathcal{S}_{i-1} \times A \times \mathcal{S}_{i+1} \times \ldots \times \mathcal{S}_n$. Let $\mathcal{F}$ be the σ-field of subsets of $\mathcal{S}$ generated by $k_1(\mathcal{F}_1) \cup \ldots \cup k_n(\mathcal{F}_n)$. Then

(i) for each i, i = 1, ..., n, k_i is a σ-complete Boolean isomorphism of $\mathcal{F}_i$ into $\mathcal{F}$, and

(ii) there exists a unique countably additive measure ν on $\mathcal{F}$ such that $\nu_i = \nu \circ k_i$ for all i and

$$\nu(A_1 \cap \ldots \cap A_n) = \nu(A_1) \ldots \nu(A_n)$$

whenever $n \geq 1$ and $A_i \in k_i(\mathcal{F}_i)$ for each i, $i = 1, \ldots, n$.

This theorem is proven, for example, in section 37 of Halmos' Measure Theory.

Suppose now that we begin with a collection $(\mathcal{M}_1, \mu_1), \ldots, (\mathcal{M}_n, \mu_n)$ of probability algebras and that we wish to construct an orthogonal sum. Then by the last theorem in section 2, we can suppose that for each i, $i = 1, \ldots, n$, there exists a set $\mathcal{S}_i$, a σ- field $\mathcal{F}_i$ of subsets of $\mathcal{S}_i$, a countably additive measure ν_i on $\mathcal{F}_i$, and an isomorphism j_i of $\mathcal{M}_i$ onto the quotient $\mathcal{F}_i/I_i$, where I_i is the σ - ideal of sets of measure zero and $\nu_i(F) = \mu_i(M)$ whenever $F \in j_i(M)$. Suppose, then, that we let $(\mathcal{S}, \mathcal{F}, \nu)$ and $k_1, \ldots, k_n$ be as in the preceding theorem. Then denoting by I the σ-ideal of $\mathcal{F}$ consisting of all sets of measure zero, we may set $\mathcal{M} = \mathcal{F}/I$ and let μ be the measure on $\mathcal{M}$ inherited from the measure ν on $\mathcal{F}$. Then $(\mathcal{M}, \mu)$ will be a probability algebra and a candidate as an orthogonal sum of $(\mathcal{M}_1, \mu_1), \ldots, (\mathcal{M}_n, \mu_n)$. But we still require the embeddings $f_1, \ldots, f_n$.

First we must use the isomorphisms $k_i: \mathcal{F}_i \to \mathcal{F}$ to construct isomorphisms $k_i': \mathcal{F}_i/I_i \to \mathcal{F}/I$. It is easily seen that whenever A, B $\in \mathcal{F}_i$ differ only by a set of measure zero, their images $k_i(A)$ and $k_i(B)$ differ only by a set of measure zero. Hence k_i' may be defined by setting $k_i'([E]) = [k_i(E)]$. It is easily verified that the k_i' defined in this way are indeed isomorphisms into. Finally, setting $f_i = k_i' \circ j_i$ for each i, $i = 1, \ldots, n$, we obtain the desired embeddings.

6. Bibliographic Notes

With the exception of the ideas in section 3, most of the material in this chapter is fairly well known to students of Boolean algebra. But it is not as widely accessible as the material of the preceding chapter. Several of the proofs in sections 1 and 2 can be gleaned from pp. 55-68 of Halmos' Lectures on Boolean Algebras, but for others I have been unable to find any references.

For a proof of Caratheodory's Extension Theorem, the reader may consult Robert Bartle's Theory of Integration, pp. 98-104.

Another method of proving the main theorem of section 2 would be to take the quotient first and then embed the resulting positive measure algebra in a probability algebra by completing the metric space given by the distance $d(A, B) = \mu(A \Delta B)$. This approach is spelled out in Demetrios A. Kappos' Probability Algebras and Stochastic Spaces, p. 12 and pp. 16-28.

CHAPTER 5. CONDENSABLE ALLOCATIONS

An allocation of probability on a power set $\mathcal{P}(\mathcal{S})$ is <u>condensable</u> if its upper probability function P^* obeys

$$P^*(A) = \sup \{P^*(B) \mid B \subset A; \text{ B is finite}\} .$$

Condensability is a very important property. It is a property that can generally be expected for belief functions based on empirical evidence; and belief functions that are condensable are intuitively much more transparent than belief functions in general.

This chapter is devoted to the mathematical and intuitive aspects of condensability, and aims at an understanding of the commonality numbers, which provide the best way of describing condensable allocations.

1. Condensability

The theory of degrees of belief set out in the preceding chapters is really built on a single simple intuition: if a given portion of our belief is committed to both of two propositions A and B, then it should be committed to the conjunction $A \wedge B$. It has been my claim that this intuition practically imposes itself -- that a probability mass's being committed to both of two propositions can only mean its being committed to their conjunction.

But one who finds this perception convincing is not likely to stop with pairs of propositions; instead, he is likely to apply the idea to larger, even to infinite collections of propositions. In other words, he will insist that if a given probability mass M is committed to each of a collection $\mathcal{B}$

of propositions, then it must be committed to the logical conjunction of all the elements of $\mathcal{B}$.

If we begin with an arbitrary Boolean algebra of propositions, $\mathcal{A}$, there is no guarantee that $\mathcal{A}$ will contain an element corresponding to the logical conjunction of a given infinite collection of propositions $\mathcal{B} \subset \mathcal{A}$. But suppose that $\mathcal{A}$ can be thought of as the power set of a set $\mathcal{S}$ of possible states of nature, so that a given proposition A in $\mathcal{A}$ asserts that the true state of nature is one of those in a certain subset A of $\mathcal{S}$. Then for any collection $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$, the set-theoretic intersection $\cap\mathcal{B}$ must be interpreted as the logical conjunction of the propositions in $\mathcal{B}$; it says that the true state of nature is in all the sets $B \in \mathcal{B}$, i.e., in their intersection $\cap\mathcal{B}$. In this case, our intuition tells us that a probability mass that is constrained to all the elements of $\mathcal{B}$ should also be constrained to $\cap\mathcal{B}$.

This intuition goes beyond the intuition we have used thus far, and not all allocations of probability on a power set will satisfy it; our rules for allocations imply it for finite collections $\mathcal{B}$, but not for infinite ones. So those allocations that do meet this intuition deserve a special name: A standard allocation of probability $\rho : \mathcal{P}(\mathcal{S}) \to \mathcal{M}$ over a set $\mathcal{S}$ will be called <u>condensable</u> if for each $M \in \mathcal{M}$ and $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$, M is constrained to $\cup\mathcal{B}$ if and only if it is constrained to each element $B \in \mathcal{B}$.

The requirement that ρ must be standard should not be overlooked; it means that the properties of condensable allocations depend on our intuition about what our probability itself looks like, as well as upon our intuitive understanding of the logical structure of $\mathcal{P}(\mathcal{S})$. In fact, though, condensability is a property of the belief function or the upper probability function and does not depend on which standard representation is used.

Theorem. Suppose $\rho: \mathcal{P}(\mathcal{S}) \to \mathcal{M}$ is an allocation into the probability algebra $(\mathcal{M}, \mu)$. Denote by ζ the allowment $\zeta: \mathcal{P}(\mathcal{S}) \to \mathcal{M}: A \leadsto \overline{\rho(\tilde{A})}$, by Bel the belief function $\mu \circ \rho$, by P* the upper probability $\mu \circ \zeta$, and by ct the constraint relation defined by "A ct M if and only if $M \leq \rho(A)$. Then the following seven conditions are all equivalent.

(i) ρ is condensable -- i.e., if $\mathcal{B}$ is a non-empty subset of $\mathcal{P}(\mathcal{S})$, $M \in \mathcal{M}$, and M ct B for all $B \in \mathcal{B}$, then M ct $\cap \mathcal{B}$.

(ii) (ii) $\rho(\cap \mathcal{B}) = \bigwedge_{B \in \mathcal{B}} \rho(B)$ for all non-empty $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$.

(iii) $\zeta(\cup \mathcal{B}) = \bigvee_{B \in \mathcal{B}} \zeta(B)$ for all non-empty $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$.

(iv) For every non-empty $A \subset \mathcal{S}$, there exists a sequence $s_1, s_2, \ldots$ of elements of A and a countable disjoint partition $M_1, M_2, \ldots$ of $\zeta(A)$ such that $M_i \leq \zeta(\{s_i\})$ for each positive integer i.

(v) For each $A \subset \mathcal{S}$, $P^*(A) = \sup_{\substack{A' \subset S \\ A' \text{ finite}}} P^*(A')$.

(vi) If $\mathcal{B}$ is an upward net in $\mathcal{P}(\mathcal{S})$, then $P^*(\cup \mathcal{B}) = \sup_{B \in \mathcal{B}} P^*(B)$

(vii) If $\mathcal{B}$ is a downward net in $\mathcal{P}(\mathcal{S})$, then $\mathrm{Bel}(\cap \mathcal{B}) = \inf_{B \in \mathcal{B}} \mathrm{Bel}(B)$.

Proof: (i) $\Rightarrow$ (ii). Since allocations are isotone, $\rho(\cap \mathcal{B}) \leq \rho(B)$ for all $B \in \mathcal{B}$, and hence $\rho(\cap \mathcal{B}) \leq \bigwedge_{B \in \mathcal{B}} \rho(B)$. On the other hand, $\bigwedge_{B \in \mathcal{B}} \rho(B) \leq \rho(B)$ for all $B \in \mathcal{B}$ -- i.e., $\bigwedge_{B \in \mathcal{B}} \rho(B)$ ct B for all $B \in \mathcal{B}$. So by condensability, $\bigwedge_{B \in \mathcal{B}} \rho(B)$ ct $\cap \mathcal{B}$ -- i.e., $\bigwedge_{B \in \mathcal{B}} \rho(B) \leq \rho(\cap \mathcal{B})$.

(ii) $\Rightarrow$ (iii). $\zeta(\cup \mathcal{B}) = \overline{\rho(\widetilde{\cup \mathcal{B}})} = \overline{\rho(\widetilde{\bigcup_{B \in \mathcal{B}} B})}$

$= \overline{\rho(\bigcap_{B \in \mathcal{B}} \tilde{B})} = \overline{\bigwedge_{B \in \mathcal{B}} \rho(\tilde{B})} = \bigvee_{B \in \mathcal{B}} \overline{\rho(\tilde{B})} = \bigvee_{B \in \mathcal{B}} \zeta(B)$.

(iii) $\Rightarrow$ (iv). For every non-empty $B \subset \mathcal{S}$, $\zeta(B) = \bigvee_{s \in B} \zeta(\{s\})$. Hence (iv) follows by the second theorem of Chapter 4, section 1.

(iv) $\Rightarrow$ (v). We can suppose A is non-empty, and in that case we can choose a sequence $s_1, s_2, \ldots$ of points of A such that $\vee \zeta(\{s_i\}) = \zeta(A)$. But $\bigvee_{i=1}^{\infty} \zeta(\{s_i\}) = \bigvee_{n=1}^{\infty} \zeta(\{s_1, \ldots, s_n\})$, and $\zeta(\{s_1\}) \leq \zeta(\{s_1, s_2\}) \leq \zeta(\{s_1, s_2, s_3\}) \leq \ldots$ is an increasing sequence in $\mathcal{M}$. Hence, by the third theorem of Chapter 4, section 1, $P^*(A) = \mu(\zeta(A)) = \mu(\vee \zeta(\{s_1, \ldots, s_n\})) = \sup_n \mu(\zeta(\{s_1, \ldots, s_n\})) = \sup_n P^*(\{s_1, \ldots, s_n\}) \leq \sup_{\substack{A' \subset A \\ A' \text{ finite}}} P^*(A')$. The inequality $\sup_{\substack{A' \subset A \\ A' \text{finite}}} P^*(A') \leq P^*(A)$ follows, of course, from the monotonicity of P^*.

(v) $\Rightarrow$ (vi). Suppose $\mathcal{B}$ is an upward net and A is a finite subset of $\cup\mathcal{B}$. Then it is easily verified by induction on the number of elements of A that there exists an element $B \in \mathcal{B}$ such that $A \subset B$. Hence if $\mathcal{B}$ is an upward net, $\sup_{B \in \mathcal{B}} P^*(B) \geq \sup_{\substack{A \subset \cup\mathcal{B} \\ A \text{ finite}}} P^*(A) = P^*(\cup \mathcal{B})$. The inequality $\sup_{B \in \mathcal{B}} P^*(B) \leq P^*(B)$ follows, of course, from the monotonicity of P^*.

(vi) $\Rightarrow$ (vii). Suppose $\mathcal{B}$ is a downward net in $P(\mathcal{S})$. Then $\mathcal{C} = \{\tilde{B} | B \in \mathcal{B}\}$ is an upward net in $P(\mathcal{S})$. Hence

$$\mathrm{Bel}(\cap\mathcal{B}) = 1 - P^*(\widetilde{\cap\mathcal{B}}) = 1 - P^*(\cup\mathcal{C}) = 1 - \sup_{C \in \mathcal{C}} P^*(C)$$

$$= 1 - \sup_{B \in \mathcal{B}} P^*(\tilde{B}) = \inf_{B \in \mathcal{B}} (1 - P^*(\tilde{B})) = \inf_{B \in \mathcal{B}} \mathrm{Bel}(B).$$

(vii) $\Rightarrow$ (ii). Suppose $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$ is non-empty. Then $\mathcal{C} = \{\cap\mathcal{A} | \mathcal{A} \subset \mathcal{B}, \mathcal{A} \text{ finite}\}$ is a downward net in $\mathcal{P}(\mathcal{S})$. But $\cap\mathcal{B} = \cap\mathcal{C}$ and $\bigwedge_{C \in \mathcal{C}} \rho(C) = \bigwedge_{B \in \mathcal{B}} \rho(B)$. Hence

$$\mu(\rho(\cap\mathcal{B})) = \mu(\rho \cap \mathcal{C})) = \mathrm{Bel}(\cap\mathcal{C}) = \inf_{C \in \mathcal{C}} \mathrm{Bel}(C)$$

$$= \inf_{C \in \mathcal{C}} \mu(\rho(C)) = \mu(\bigwedge_{C \in \mathcal{C}} \rho(C)) = \mu(\bigwedge_{B \in \mathcal{B}} \rho(B)).$$

Since $\rho(\cap \mathcal{B}) \subset \bigwedge_{B \in \mathcal{B}} \rho(B)$ and μ is positive, it follows that $\rho(\cap \mathcal{B}) = \bigwedge_{B \in \mathcal{B}} \rho(B)$.

(ii) $\Rightarrow$ (i) If M ct B for all $B \in \mathcal{B}$, then $M \subset \rho(B)$ for all $B \in \mathcal{B}$, or $M \subset \bigwedge_{B \in \mathcal{B}} \rho(B)$. Hence $M \subset \rho(\cap \mathcal{B})$, or M ct $\cap \mathcal{B}$. ▨

Since conditions (v), (vi) and (vii) make no reference to any particular standard representation for the belief function or upper probability function, this theorem justifies the assertion that condensability is a property of the belief function or upper probability function and does not depend on which standard representation is used. More generally, the theorem shows that the adjective condensable can properly be applied to the constraint relation, the allowment, the upper probability function or the belief function, as well as to the allocation ρ. I will follow such a usage in the sequel.

Condition (iv) is of particular interest for the intuitive understanding of condensability. It states that the probability mass $\zeta(A)$ -- the total probability mass that can get into B -- can be divided into a countable number of discrete pieces, each of which can get into some single point of B. We will shortly see why this property deserves to be called "condensability."

It is condition (v) that we will deal with most often in the sequel. Its utility is obvious -- it means that the entire upper probability function is determined by its values on finite subsets and thus allows us to examine the structure of condensable upper probability functions much more closely. We will begin this closer examination in section 3.

In my definition of condensability, I have required that the allocation or belief function be on a power set $\mathcal{P}(\mathcal{S})$. This may seem unnecessarily restrictive, for the definition could easily be extended to any complete Boolean algebra in which arbitrary meets and joins can be understood as conjunctions and disjunctions. It is not clear, however, that there are any such Boolean algebras which are not isomorphic to power sets; and hence it is not clear whether the seemingly more general formulation is of any real interest. In any case, the upper probability functions that we will be concerned with will be on power sets.

There are many ways in which condensable belief functions are more attractive than belief functions in general. Consider, for example, the problem of sets of "upper probability zero." If the upper probability function $P^*:\mathcal{P}(\mathcal{J}) \longrightarrow [0,1]$ is condensable, then the set

$$S = \bigcup \{ S' \mid P^*(S') = 0 \}$$

will obey $P^*(S) = 0$. (This follows from condition (vi) in the preceding theorem.) The significance of this fact is that it makes it possible to interpret "$P^*(S) = 0$" as really meaning that the upper probability function P^* holds S to be impossible. In the case of non-condensable belief functions -- for example, in the case of "continuous" probability functions -- such an interpretation is, somewhat paradoxically, impossible.

2. Mobile Probability Masses

A condensable allocation on a power set $\mathcal{P}(\mathcal{J})$ can be interpreted in a very vivid way if we think of the set $\mathcal{J}$ geometrically and think of our probability as being distributed over it. More precisely, let us think of our probability not as being distributed in a fixed way, but rather as having a certain degree of mobility. In other words, the various probability masses in $\mathcal{M}$ are to be allowed to move around, to some extent, within $\mathcal{S}$.

The extent of the mobility is specified by the constraint relation ct between $\mathcal{M}$ and $\mathcal{P}(\mathcal{S})$; if a probability mass $M \in \mathcal{M}$ is constrained to a set $A \subset \mathcal{S}$, this means precisely that neither M nor any subelement of M can get out of A. A glance at the rules for constraint relations in section 2 of Chapter 4 will reveal that those rules are all immediately obvious from this geometric picture. And the condition of condensability is equally obvious; for if all of a probability mass is constrained to stay inside A for each A in some subset $\mathcal{B}$ of $\mathcal{P}(\mathcal{S})$, then it must be constrained to stay inside $\cap \mathcal{B}$.

An even more vivid understanding of condensability can be obtained from condition (iv) of the theorem in the preceding section. Intuitively, this condition means that though the constraints on the probability mass ζ(A) might allow it to become spread out over A in a completely diffuse fashion (as in the case of a "continuous distribution" of probability), it must always be possible to condense it into a collection of discrete pieces, just as a diffuse mass of water vapor can be condensed into a collection of drops. The word "condensability" is meant to bring to mind the possibility of such a condensation.

It is easy to think about a subset A's degree of belief Bel(A) and upper probability P*(A) in terms of this picture. Bel(A) is simply the amount of probability that cannot get out of A, while P*(A) is the amount of probability that can get into A.

If we concentrate on a probability mass $M \in \mathcal{M}$, it is natural to ask just how constrained M is. Evidently there will be a whole, possibly quite large, set $\mathscr{B} \subset \mathcal{P}(\mathscr{S})$ of subsets of $\mathscr{S}$ to which M is constrained. By condensability, M will also be constrained to $\cap \mathscr{B}$, and this will be the smallest region to which all of it is constrained -- its "tightest" constraint. But as we saw in section 9 of Chapter 2, the existence of such a "tightest" constraint for each probability mass can be described by saying that there exists a "constraint mapping" $\lambda: \mathcal{M} \rightarrow \mathcal{P}(\mathscr{S})$ that maps each probability mass to its tightest constraint. So condensability has to do with the existence of a constraint mapping.

This may be puzzling, for in Chapter 2 we saw that any belief function can be represented by an allocation of probability for which a constraint mapping exists. But the allocation constructed there was not necessarily standard -- it was into a "measure algebra" but not necessarily into a "probability algebra." And when the allocation is extended to one into a probability algebra, the constraint mapping may be lost. In fact, it will be unless the belief function is condensable.

Theorem. Suppose $\rho: \mathcal{P}(\mathscr{S}) \rightarrow \mathcal{M}$ is a standard allocation of probability. Then ρ is condensable if and only if a constraint mapping $\lambda: \mathcal{M} \rightarrow \mathcal{P}(\mathscr{S})$ exists for ρ.

Proof: If ρ is condensable, then the mapping $\lambda: \mathcal{M} \rightarrow \mathcal{P}(\mathscr{S}): M \rightsquigarrow \cap\{A \mid A \subset \mathscr{S}, \text{ M ct A}\}$ is a constraint mapping for ρ. If a constraint mapping $\lambda: \mathcal{M} \rightarrow \mathcal{P}(\mathscr{S})$ exists, then M ct A if and only $\lambda(M) \subset A$; so if M ct B for all $B \in \mathscr{B}$ it follows that $\lambda(M) \subset \mathscr{B}$ for all $B \in \mathscr{B}$ and $\lambda(M) \subset \cap \mathscr{B}$, whence M ct $\cap \mathscr{B}$. ▨

3. Upper Probabilities for Finite Subsets

A condensable upper probability function $P^*: \mathbb{P}(\mathcal{S}) \rightarrow [0,1]$ is determined by its values on finite subsets of $\mathcal{S}$. Denoting by $\mathcal{F}(\mathcal{S})$ the set of all finite subsets of $\mathcal{S}$, we can express this by saying that P^* is completely determined by $P_o^*: \mathcal{F}(\mathcal{S}) \rightarrow [0,1]$, where $P_o^* = P^* \mid \mathcal{F}(\mathcal{S})$.

This fact leads us naturally to inquire about the properties of P_o^*. On the one hand, we might ask what properties P_o^* will have on account of P^*'s being a condensable upper probability function; and on the other hand, we might look for conditions on a function $f: \mathcal{F}(\mathcal{S}) \rightarrow [0,1]$ that are sufficient to assure that the function $P^*: \mathbb{P}(\mathcal{S}) \rightarrow [0,1]: A \rightsquigarrow \sup_{\substack{A' \subset A \\ A' \text{finite}}} f(A)$ should be a condensable upper probability function. The following lemma will help us state such conditions:

<u>Lemma.</u> Let f be a real function on the set $\mathcal{F}(\mathcal{S})$ of all finite subsets of a non-empty set $\mathcal{S}$, and denote

$$\nabla_n^f (B; A_1, \ldots, A_n) = \sum_{J \subset \{1, \ldots, n\}} (-1)^{\text{card } J} f(B \cup (\bigcup_{i \in J} A_i))$$

whenever $n \geq 1$ and $B, A_1, \ldots, A_n \in \mathcal{F}(\mathcal{S})$. Now fix $A_1, \ldots, A_n$, and for each i, $i = 1, \ldots, n$, set

$$A_i = \{a_{i1}, \ldots, a_{ik_i}\}$$

and for each j, $j = 0, 1, \ldots, k_i$, set

$$A_i^j = \{a_{i1}, \ldots, a_{ij}\}.$$

($A_i^o = \phi$ for all i, $i = 1, \ldots, n$.) Then

$$\nabla_n(B; A_1, \ldots, A_n) = \sum\left\{\nabla_n\left(B \cup A_1^{j_1-1} \cup \cdots \cup A_n^{j_n-1};\right.\right.$$

$$\left.\left.\{a_{1j_1}\}, \ldots, \{a_{nj_n}\}\right) \middle| 1 \le j_i \le k_i\right\}.$$

Proof: If $k_i = 0$ for some i, then $A_i = \phi$, and it is evident from the fact that ∇_n is a successive difference (cf. Chapter 1, section 3) that $\nabla_n(B, A_1, \ldots, A_n) = 0$; on the other hand, the right-hand side above would also be zero, for there would be no terms in the summation. Hence we may assume that $k_i > 0$ for $i = 1, \ldots, n$.

In that case,

$$\text{r.h.s.} = \sum_{\substack{(j_1,\ldots,j_n), \\ 0 \le j_i \le k_i}} \sum_{J \subset \{1,\ldots,n\}} (-1)^{\text{card } J} f\left(B \cup \left(\bigcup_{i \in J} A_i^{j_i}\right) \cup \left(\bigcup_{i \notin J} A_i^{j_i - 1}\right)\right)$$

$$= \sum_{\substack{(j_1,\ldots,j_n), \\ 0 \le j_i \le k_i}} f\left(B \cup \left(\bigcup_{i=1}^{n} A_i^{j_i}\right)\right) \sum \left\{(-1)^{\text{card } J} \middle| \{i \mid j_i = k_i\} \subset J \subset \{i \mid j_i \ne 0\}\right\}$$

$$= \sum_{\substack{(j_1,\ldots,j_n), \\ j_i = 0 \text{ or } k_i \text{ for each } i}} f\left(B \cup \left(\bigcup_{i=1}^{n} A_i^{j_i}\right)\right) (-1)^{\#\text{ of } i \text{ for which } j_i = k_i}$$

$$= \sum_{J \subset \{1,\ldots,n\}} f\left(B \cup \left(\bigcup_{i \in J} A_i\right)\right) (-1)^{\text{card } J}$$

$$= \nabla_n(B; A_1, \ldots, A_n).$$

▨

Theorem. Suppose f is a real function on $\mathcal{F}(\mathcal{S})$, the set of all finite subsets of a non-empty set $\mathcal{S}$. Then the real function P^* on $\mathcal{P}(\mathcal{S})$ defined by $P^*(A) = \sup_{\substack{A' \subseteq A \\ A' \text{ finite}}} f(A')$ is a condensable upper probability function if and only if

(i) $f(\emptyset) = 0$

(ii) $\sup_{A \in \mathcal{F}(\mathcal{S})} f(A) = 1$

(iii) If $A, B \in \mathcal{F}(\mathcal{S})$ and $A \neq \emptyset$, then $\sum_{T \subset A} (-1)^{\text{card } T} f(B \cup T) \leq 0$.

Proof: First we must show that if $P^*: \mathcal{P}(\mathcal{S}) \to [0,1]$ is a condensable upper probability function, then $f = P^* | \mathcal{F}(\mathcal{S})$ satisfies the three conditions. But (i) and (ii) are obvious. Now we may write $A = \{s_1, \ldots, s_n\}$ for some $n \geq 1$, and $\sum_{T \subset B} (-1)^{\text{card } T} f(B \cup T)$ then becomes $\sum_{J \subset \{1, \ldots, n\}} (-1)^{\text{card } J} P^*(B \cup (\bigcup_{i \in J} \{s_i\})) = \nabla^{P^*}_n (B; \{s_1\}, \ldots, \{s_n\})$, and this is non-positive according to section 3 of Chapter 1.

Next, we must show that P^* is a condensable upper probability function if f satisfies the three conditions and P^* is defined by $P^*(A) = \sup_{\substack{A' \subset A \\ A' \text{finite}}} f(A')$. But the relations $P^*(\emptyset) = 0$ and $P^*(\mathcal{S}) = 1$ are evident from (i) and (ii). Hence, by the last theorem of section 3 of Chapter 1, we need only show that $\nabla_n(B; A_1, \ldots, A_n) \leq 0$ for all $B, A_1, \ldots, A_n \in \mathcal{P}(\mathcal{S})$. But we have just seen that $\sum_{T \subset A} (-1)^T (B \cup T) = \nabla^f_n(B; \{s_1\}, \ldots, \{s_n\})$, where $A = \{s_1, \ldots, s_n\}$, so (iii) above asserts that $\nabla_n(B; A_1, \ldots, A_n) \leq 0$ in the case where B is finite and the A_i are singletons. The case where B and the A_i are all finite follows by the lemma.

By the definition of P^*, the values $P^*(A)$ can always be approximated by values $P^*(A')$, where $A' \subset A$ and A' is finite, so we can easily establish that $\nabla_n(B; A_1, \ldots, A_n) \leq 0$ in general by approximating each upper probability with the upper probability of a finite subset. Suppose, indeed, that

$$O < \epsilon = \nabla_n(B; A_1, \ldots, A_n) = P^*(B) - \Sigma P^*(B \cup A_i) + \cdots + (-1)^n P^*(B \cup A_i, \cup \ldots \cup A_n).$$

Then since there are 2^n terms on the right-hand side of this inequality, we can approximate B, $A_1, \ldots, A_n$ by finite subsets B', $A_1', \ldots, A_n'$ such that $P^*(B \cup A_{i_1} \cup \ldots \cup A_{i_k})$ differs from $P^*(B' \cup A_{i_1}' \cup \ldots \cup A_{i_k}')$ by less than $1/2 \cdot \epsilon / 2^n$ for each$(i_1, \ldots, i_k)$ such that $1 \leq i_1 \leq \ldots \leq i_k \leq n$. Hence, the quantity $\nabla_n(B', A_1', \ldots, A_n')$ would also be positive, contradicting our conclusion in the preceding paragraph. ▨

Theorem. Suppose $\mathcal{S}$ is a non-empty set, $(\mathcal{M}, \mu)$ is a measure algebra and

$$\zeta_o : \mathcal{S} \to \mathcal{M}$$

is such that for any $\epsilon \geq 0$ there exists a finite subset $\{s_1, \ldots, s_n\}$ of $\mathcal{S}$ such that

$$\mu(\zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n)) \geq 1 - \epsilon.$$

Then the function $P^*: \mathcal{P}(\mathcal{S}) \to [0,1]$ defined by $P^*(\phi) = 0$ and

$$P^*(A) = \sup \{\mu(\zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n)) \mid n \geq 1; \{s_1, \ldots, s_n\} \subset A\}$$

for $A \neq \phi$ is a condensable upper probability function.

Proof: Evidently, $P^*(A) = \sup_{\substack{A' \subset A \\ A' \text{finite}}} f(A)$, where $f(\phi) = 0$ and $f(\{s_1, \ldots, s_n\}) = \mu(\zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n))$. And $\sup_{A \in \mathcal{F}(\mathcal{S})} f(A) = 1$. So by the preceding theorem we need only show that if A, $B \in \mathcal{F}(\mathcal{S})$

and $A \neq 0$, then

$$\sum_{T \subset A} (-1)^{\text{card } T} f(B \cup T) \leq 0.$$

Now set $A = \{s_1, \ldots, s_n\}$ and set

$$M = \begin{cases} \Lambda_{\mathcal{m}} & \text{if } B = \emptyset \\ \zeta(t_1) \vee \ldots \vee \zeta(t_m) & \text{if } B = \{t_1, \ldots, t_m\}, \text{ where } m \geq 1. \end{cases}$$

Then

$$\sum_{T \subset A} (-1)^{\text{card } T} f(B \cup T)$$

$$= \mu(M) - \Sigma\mu(M \vee \zeta_o(s_i)) + \Sigma\mu(M \vee \zeta_o(s_i) \vee \zeta_o(s_j))$$

$$- + \ldots + (-1)^{n+1} \mu(M \vee \zeta_o(s_1) \vee \ldots \vee \zeta_o(s_n))$$

$$= \mu(M) - \mu((M \vee \zeta_o(s_1)) \wedge \ldots \wedge (M \vee \zeta_o(s_n)))$$

$$= \mu(M) - \mu(M \vee (\wedge \zeta_o(s_i)))$$

$$\leq 0.$$

////

4. Commonality Numbers

Let $\rho: \mathcal{P}(\mathcal{S}) \longrightarrow \mathcal{M}$ be a condensable allocation of probability, and let ζ be the allowment associated with ρ. In other words, $\zeta(A) = \overline{\rho(\tilde{A})}$. Then for each $s \in \mathcal{S}$, $\zeta(\{s\})$ is the total probability mass that can reach the point s. And for any non-empty finite subset $A = \{s_1, \ldots, s_n\}$ of $\mathcal{S}$, $\zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})$ is the total probability mass that can reach each and every point of A -- i.e., the total probability mass that can move completely freely within A.

Now if $A = \emptyset$, the total probability mass that can reach each and every point of A is $\mathcal{V}_{\mathcal{M}}$. Hence it is natural to define a mapping $\gamma: \mathcal{F}(\mathcal{S}) \to \mathcal{M}$ by $\gamma(\emptyset) = \mathcal{V}_{\mathcal{M}}$ and $\gamma(\{s_1, \ldots, s_n\} = \zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})$. As it turns out, the measures of the probability masses $\gamma(A)$, $A \in \mathcal{F}(\mathcal{S})$, are very important and hence deserve a name. Setting $Q = \mu \circ \gamma$, where μ is the measure on $\mathcal{M}$, I will call Q(A) the "commonality number" for A, and I will call $Q: \mathcal{F}(\mathcal{S}) \to [0, 1]$ the "commonality function" associated with ρ.

Notice that the commonality number Q(A) decreases as A is enlarged. Indeed, $Q(\emptyset) = \mu(\mathcal{V}_{\mathcal{M}}) = 1$, and $Q(\{s_1, \ldots, s_n, s_{n+1}\}) = \mu(\zeta(\{s_1\} \wedge \ldots \wedge \zeta(\{s_n\}) \wedge \quad \zeta(\{s_{n+1}\})) \leq \mu(\zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})) = Q(\{s_1, \ldots, s_n\})$. This contrasts sharply with the behavior of the upper probability $P^*(A)$ which begins at zero when $A = \emptyset$ and increases as A is enlarged.

Actually, commonality numbers and upper probabilities are related by a much more extensive duality. For while the commonality numbers give the measures of the intersections of the probability masses $\zeta(\{s_i\})$, upper probabilities give the measures of their unions:

$$Q(\{s_1, \ldots, s_n\}) = \mu(\zeta(\{s_1\}) \wedge \ldots \wedge \zeta(\{s_n\})),$$

while

$$P^*(\{s_1, \ldots, s_n\}) = \mu(\zeta(\{s_1, \ldots, s_n\})) = \mu(\zeta(\{s_1\}) \vee \ldots \vee \zeta(\{s_n\})).$$

Now we know from the theory of measure (and from Chapter 1, section 5) that the measures of finite meets can always be expressed in terms of the measures of finite joins and vice-versa:

$$\mu(M_1 \wedge \ldots \wedge M_n) = \Sigma\mu\,(M_i) - \Sigma\mu\,(M_i \vee M_j) + - \ldots + (-1)^{n+1}\mu(M_1 \vee \ldots \vee M_n)$$

and

$$\mu(M_1 \vee \ldots \vee M_n) = \Sigma\mu\,(M_i) - \Sigma\mu\,(M_i \wedge M_j) + - \ldots + (-1)^{n+1}\,\mu(M_1 \wedge \ldots \wedge M_n)$$

for all $M_1, \ldots, M_n \in \mathcal{M}$. So for all non-empty finite subsets $\{s_1, \ldots, s_n\}$ of $\mathcal{S}$,

$$Q(\{s_1, \ldots, s_n\}) = \Sigma P^*(\{s_i\}) - \Sigma P^*(\{s_i, s_j\} + - \ldots + (-1)^{n+1} P^*(\{s_1, \ldots, s_n\})$$

and

$$P^*(\{s_1, \ldots, s_n\}) = \Sigma Q(\{s_i\}) - \Sigma Q(\{s_i, s_j\}) + - \ldots + (-1)^{n+1}\, Q(\{s_1, \ldots, s_n\}).$$

It is evident from this last formula that the commonality numbers determine the upper probabilities for finite subsets and hence the entire condensable upper probability function.

So in the condensable case commonality functions are simply another form in which belief functions may be specified. It will be useful to know what properties fully characterize them.

Definition. A real function Q on the set $\mathcal{F}(\mathcal{S})$ of all finite subsets of a non-empty set $\mathcal{S}$ is called a commonality function if

(i) $Q(\emptyset) = 1$,

(ii) $\inf_{A \in \mathcal{F}(\mathcal{S})} \sum_{T \subset A} (-1)^{\text{card } T}\, Q(T) = 0$,

(iii) If $A, B \in \mathcal{F}(\mathcal{S})$, then $\sum_{T \subset B} (-1)^{\text{card } T}\, Q(A \cup T) \geq 0$.

Theorem. If the function Q on $\mathcal{F}(\mathcal{S}))$ is a commonality function, then it takes values in the interval $[0, 1]$.

Proof: Setting $B = \{s\}$ in (iii) yields $Q(A) - Q(A \cup \{s\}) \geq 0$, or $Q(A) \geq Q(A \cup \{s\})$ for all $A \in \mathcal{F}(\mathcal{S})$ and $s \in \mathcal{S}$. But $Q(\phi) = 1$. Hence $Q(A) \leq 1$ for all $A \in \mathcal{F}(\mathcal{S})$.

Setting $B = \phi$ in (iii) yields $Q(A) \geq 0$ for all $A \in \mathcal{F}(\mathcal{S})$. ▨

Lemma: Suppose f is a real function on the set $\mathcal{F}(\mathcal{S})$ of finite subsets of a set $\mathcal{S}$. And suppose $A, B \in \mathcal{F}(\mathcal{S})$. Then

$$\sum_{T \subset A \cup B} (-1)^{\text{card } T} f(T) = \sum_{R \subset A} \sum_{S \subset B} (-1)^{\text{card } R} (-1)^{\text{card } S} f(R \cup S).$$

Proof: $$\sum_{R \subset A} \sum_{S \subset B} (-1)^{\text{card } R} (-1)^{\text{card } S} f(R \cup S)$$

$$= \sum_{T \subset A \cup B} f(T) \sum \{(-1)^{\text{card } R + \text{card } S} \mid R \subset A;\ S \subset B;\ R \cup S = T\}$$

$$= \sum_{T \subset A \cup B} f(T)\ (-1)^{\text{card}(A-B) + \text{card}(B-A)} \sum \{(-1)^{\text{card } R + \text{card } S} \mid R, S \subset A \cap B;\ R \cup S = A \cap B\}.$$

But for any subset A,

$$\sum \{(-1)^{\text{card } R + \text{card } S} \mid R, S \subset A;\ R \cup S = A\} = (-1)^{\text{card } A}.$$

The lemma follows. ▨

Theorem. Suppose $P^*: \mathcal{P}(\mathcal{S}) \rightarrow [0, 1]$ is a condensable upper probability function and define the function Q on $\mathcal{F}(\mathcal{S})$ by $Q(\phi) = 1$ and

$$Q(A) = -\sum_{T \subset A} (-1)^{\text{card } T} P^*(T)$$

for non-empty $A \in \mathcal{J}(\mathcal{S})$. Then Q is a commonality function.

Proof: (i) $Q(\phi) = 1$ by definition.

(ii) If $A = \phi$, then $\sum_{T \subset A} (-1)^{\text{card } T} Q(T) = Q(\phi) = 1$.

If, on the other hand, $A \neq \phi$, then we can write $A = \{s_1, \ldots, s_n\}$ with $n \geq 1$, and

$$\sum_{T \subset A} (-1)^{\text{card } T} Q(T) = 1 - \sum_{\substack{T \subset A \\ T \neq \phi}} (-1)^{\text{card } T} \sum_{R \subset T} (-1)^{\text{card } R} P^*(R)$$

$$= 1 - \sum_{R \subset A} (-1)^{\text{card } R} P^*(R) \Big(\sum_{R \subset T \subset A} (-1)^{\text{card } T} \Big)$$

$$= 1 - P^*(A).$$

Hence $\inf_{A \in \mathcal{J}(\mathcal{S})} \sum_{T \subset A} (-1)^{\text{card } T} Q(T) = \inf_{A \in \mathcal{J}(\mathcal{S})} (1 - P^*(A)) = 1 - \sup_{A \in \mathcal{J}(\mathcal{S})} P^*(A) = 0$.

(iii) Finally, we need to show that

$$\sum_{T \subset B} (-1)^{\text{card } T} Q(A \cup T) \geq 0$$

for all $A, B \in \mathcal{J}(\mathcal{S})$. If $A = \phi$, this reduces to

$$\sum_{T \subset B} (-1)^{\text{card } T} Q(T) \geq 0,$$

and we just proved this. Hence we may assume that $A \neq \phi$, writing $A = \{s_1, \ldots, s_n\}$ and $B = \{t_1, \ldots, t_p\}$, where $n \geq 1$ and $p \geq 0$. Then

$$\sum_{T \subset B} (-1)^{\text{card } T} Q(A \cup T) = -\sum_{T \subset B} (-1)^{\text{card } T} \sum_{R \subset A \cup T} (-1)^{\text{card } R} P^*(R)$$

$$= - \sum_{T \subset B} (-1)^{\text{card } T} \sum_{R \subset A} \sum_{S \subset T} (-1)^{\text{card } R} (-1)^{\text{card } S} P^*(R \cup S)$$

$$= - \sum_{R \subset A} (-1)^{\text{card } R} \sum_{S \subset B} (-1)^{\text{card } S} P^*(R \cup S) \left(\sum_{S \subset T \subset B} (-1)^{\text{card } T} \right)$$

$$= - \sum_{R \subset A} (-1)^{\text{card } R} P^*(B \cup R).$$

But $\sum_{R \subset A} (-1)^{\text{card } R} P^*(B \cup R) \leq 0$ by the last theorem of the preceding section. ▨

<u>Theorem</u>. Suppose Q: $\mathcal{J}(\mathcal{S}) \to [0, 1]$ is a commonality function, and define the function P* on $\mathcal{P}(\mathcal{S})$ by $P^*(\phi) = 0$,

$$P^*(A) = - \sum_{\substack{T \subset A \\ T \neq \phi}} (-1)^{\text{card } T} Q(T)$$

for finite non-empty subsets A of $\mathcal{S}$, and

$$P^*(A) = \sup_{\substack{A' \subset A \\ A' \text{finite}}} P^*(A')$$

for infinite subsets $A \subset \mathcal{S}$. Then P* is a condensable upper probability function.

<u>Proof</u>: By the last theorem of the preceding section, it suffices to prove that

(i) $P^*(\phi) = 0$

(ii) $\sup_{A \in \mathcal{J}(\mathcal{S})} P^*(A) = 1$

and (iii) If A, B $\in \mathcal{J}(\mathcal{S})$ and $A \neq \phi$, then $\sum_{T \subset A} (-1)^{\text{card } T} P^*(B \cup T) \leq 0$.

But (i) is given by convention. As for (ii), for finite non-empty subsets A,

$$P^*(A) = 1 - \sum_{T \subset A} (-1)^{\text{card } T} Q(T),$$

so

$$\sup_{A \in \mathcal{F}(\mathcal{S})} P^*(A) = 1 - \inf_{A \in \mathcal{F}(\mathcal{S})} \Sigma (-1)^{\text{card } T} Q(T) = 1.$$

To prove (iii), note that

$$\sum_{T \subset A} (-1)^{\text{card } T} P^*(B \cup T) = - \sum_{T \subset A} (-1)^{\text{card } T} \sum_{\substack{R \subset B \cup T \\ R \neq \emptyset}} (-1)^{\text{card } R} Q(R)$$

$$= - \sum_{T \subset A} (-1)^{\text{card } J} \sum_{\substack{R \subset B \\ S \subset T \\ \text{either } R \text{ or } S \neq \emptyset}} (-1)^{\text{card } R} (-1)^{\text{card } S} Q(R \cup S)$$

$$= - \sum_{R \subset B} (-1)^{\text{card } R} \sum_{\substack{S \subset A \\ S \neq \emptyset \text{ if } R = \emptyset}} (-1)^{\text{card } S} Q(R \cup S) \sum_{S \subset T \subset A} (-1)^{\text{card } T}$$

$$= - \sum_{R \subset B} (-1)^{\text{card } R} Q(A \cup S).$$

But $\sum_{R \subset B} (-1)^{\text{card } R} Q(A \cup S) \geq 0$ by the definition of commonality functions. ▨

In the sequel, we will sometimes examine a real function on $\mathcal{F}(\mathcal{S}) - \{\emptyset\}$ with the question as to whether it can be "renormalized" so as to yield a commonality function. In other words, given a function Q_1 on $\mathcal{F}(\mathcal{S}) - \{\emptyset\}$, we will want to know whether there exists a constant K such that the function Q on $\mathcal{F}(\mathcal{S})$ defined by

$$Q(A) = \begin{cases} 1 & \text{if } A = \phi \\ K\,Q_1(A) & \text{if } A \neq \phi \end{cases}$$

is a commonality function. The following theorem gives the conditions under which such a constant does exist.

<u>Theorem</u>. Suppose $\mathcal{S}$ is a non-empty set and Q_1 is a real function on $\mathcal{T}(\mathcal{S}) - \{\phi\}$. And for each positive number K define a real function Q_K on $\mathcal{T}(\mathcal{S})$ by

$$Q_K(A) = \begin{cases} 1 & \text{if } A = \phi \\ K\,Q_1(A) & \text{if } A \neq \phi. \end{cases}$$

Then Q_K is a commonality function if and only if

(i) $$\sup_{A \in (\mathcal{T}(\mathcal{S}) - \{\phi\})} \sum_{\substack{T \subset A \\ T \neq \phi}} (-1)^{1 + \text{card } T} Q_1(T) = 1/K,$$

(ii) If $A, B \in (\mathcal{T}(\mathcal{S}) - \{\phi\})$, then

$$\sum_{\substack{T \subset B \\ T \neq \phi}} (-1)^{1+\text{card } T} Q_1(A \cup T) \leq 1/K.$$

This theorem follows directly from the definition of commonality functions.

The preceding discussion has been primarily concerned with the relation between Q and P*. The formulae connecting Q and Bel are in some respects simpler and worth recording:

$$Q(A) = \sum_{T \subset A} (-1)^{\text{card } T}\, \text{Bel}(\tilde{T})$$

for all $A \in \mathcal{F}(\mathcal{S})$, including ϕ; and

$$\text{Bel}(A) = \sum_{A \subset T} (-1)^{\text{card } \tilde{T}}\, Q(\tilde{T})$$

for cofinite A and

$$\text{Bel}(A) = \inf_{\substack{A \subset A' \\ A' \text{cofinite}}} \sum_{A' \subset T} (-1)^{\text{card } \tilde{T}}\, Q(\tilde{T})$$

in general. A subset A of $\mathcal{S}$ is said to be cofinite if $\tilde{A} = \mathcal{S} \sim A$ is finite. The quantity

$$\inf_{\substack{A \subset A' \\ A' \text{cofinite}}} \sum_{A' \subset T} (-1)^{\text{card } \tilde{T}}\, Q(\tilde{T})$$

can be thought of intuitively as the summation of $(-1)^{\text{card } T}\, Q(T)$ over all finite T that do not intersect A.

5. Restricting Condensable Allocations

It is not difficult to prove that a complete subalgebra of a power set is itself isomorphic to a power set. Hence, it makes sense to ask whether a condensable allocation $\rho : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{M}$ remains condensable when it is restricted to a complete subalgebra $\mathcal{A} \subset \mathcal{P}(\mathcal{S})$.

The answer is obviously yes; for, since

$$\rho(\cap B) = \bigwedge_{B \in \mathcal{B}} \rho(B)$$

holds for all $\mathcal{B} \subset \mathcal{P}(\mathcal{S})$ it will certainly hold for all $\mathcal{B} \subset \mathcal{Q}$.

CHAPTER 6. EXTENSION AND COMBINATION

In this chapter we begin to see just how flexible belief functions are. In particular, we find that belief functions on given Boolean algebras can sometimes be used to obtain belief functions on more complicated Boolean algebras.

The central concern of the chapter is a rule that enables one to combine belief functions on different Boolean algebras into a single resultant belief function on their independent sum. A quite general rule is adduced for such combination, and a much simpler rule is derived for the condensable case.

The existence of such a rule also leads to the exploration of the notion of subalgebras being "independent" with respect to a belief function. As it turns out, it is convenient to distinguish between the notions of "orthogonality" and "cognitive independence," notions which collapse into a single notion in the case of probability functions.

1. Extending Allocations of Probability

In this section we will study one of the most remarkable and fruitful features of the theory of allocations: the fact that an allocation of probability on a subalgebra of a larger algebra always has a natural extension to the larger algebra. The existence of such an extension results from the fundamental intuition that any portion of our belief that is committed to a given proposition must also be committed to any proposition that it implies -- i.e., to any more inclusive proposition.

Suppose, indeed, that we have a standard allocation of probability $\rho_o: \mathcal{Q}_o \to \mathcal{M}$, where $\mathcal{Q}_o$ is a subalgebra of a Boolean algebra of propositions $\mathcal{Q}$. And suppose further that the allocation ρ_o on the subalgebra $\mathcal{Q}_o$ exhausts our opinions about the subject matter of the propositions in $\mathcal{Q}$. Then does ρ_o endow us with positive degrees of belief for any of the propositions in $\mathcal{Q}$ that are not in $\mathcal{Q}_o$?

It may well do so. For suppose $A \in \mathcal{Q}$ and $A \notin \mathcal{Q}_o$. Then there may be an element $A_o \in \mathcal{Q}_o$ such that $A_o \leq A$; and in such a case the probability mass $\rho_o(A_o)$, being committed to A_o, will certainly be committed to A as well. In general we must commit to A all the probability masses $\rho_o(A_o)$ for all the $A_o \in \mathcal{Q}_o$ that are subelements of A. So altogether we must commit the probability mass $\vee\{\rho_o(A_o) | A_o \in \mathcal{Q}_o;\ A_o \leq A\}$ to A. So the possession of the allocation $\rho_o: \mathcal{Q}_o \to \mathcal{M}$ and the lack of any further opinions about $\mathcal{Q}$ would seem to leave us with an allocation

$$\rho: \mathcal{Q} \to \mathcal{M}: A \leadsto \vee\{\rho_o(A_o) \,|\, A_o \in \mathcal{Q}_o;\ A_o \leq A\} \tag{1}$$

on $\mathcal{Q}$. But it this an allocation?

<u>Theorem</u>. Suppose $\mathcal{Q}_o$ is a subalgebra of a Boolean algebra $\mathcal{Q}$ and $\rho_o: \mathcal{Q}_o \to \mathcal{M}$ is a standard allocation of probability. Then the mapping $\rho: \mathcal{Q} \to \mathcal{M}$ given by (1) is a standard allocation on $\mathcal{Q}$. Furthermore, $\rho | \mathcal{Q}_o = \rho_o$. And the belief functions Bel_o and Bel given by ρ_o and ρ respectively are related by the formula

$$Bel(A) = \sup_{\substack{A_o \in \mathcal{Q}_o, \\ A_o \leq A}} Bel_o(A_o) , \tag{2}$$

while the upper probability functions P^*_o and P^* are related by

$$P^*(A) = \inf_{\substack{A_o \in \mathcal{A}_o, \\ A \leq A_o}} P^*_o(A_o) \,. \tag{3}$$

<u>Proof</u>: The existence of the probability masses $\vee\{\rho_o(A_o) \mid A_o \in \mathcal{A}_o; A_o \leq A\}$ depends, of course, on the fact that $(\mathcal{M}, \mu)$ is a probability algebra, so that $\mathcal{M}$ is complete. If $A \in \mathcal{A}_o$, it is evident that $\rho(A) = \rho_o(A_o)$; hence $\rho|\mathcal{A}_o = \rho_o$. In particular, $\rho(\Lambda_{\mathcal{A}}) = \Lambda_{\mathcal{M}}$, and $\rho(V_{\mathcal{A}}) = V_{\mathcal{M}}$.

Furthermore, for all pairs $A_1, A_2 \in \mathcal{A}$,

$$\begin{aligned}
\rho(A_1)\wedge\rho(A_2) &= [\vee\{\rho_o(B_1) \mid B_1 \in \mathcal{A}_o;\ B_1 \leq A_1\}] \wedge [\vee\{\rho_o(B_2) \mid B_2 \in \mathcal{A}_o;\ B_2 \leq A_2\}] \\
&= \vee\{\rho_o(B_1) \wedge \rho_o(B_2) \mid B_1, B_2 \in \mathcal{A}_o;\ B_1 \leq A_1;\ B_2 \leq A_2\} \\
&= \vee\{\rho_o(B_1 \wedge B_2) \mid B_1, B_2 \in \mathcal{A}_o;\ B_1 \leq A_1,\ B_2 \leq A_2\} \\
&= \vee\{\rho(B) \mid B \in \mathcal{A}_o;\ B \leq A_1 \wedge A_2\} = \rho(A_1 \wedge A_2).
\end{aligned}$$

Hence ρ is an allocation. Since $(\mathcal{M}, \mu)$ is a probability algebra, ρ is standard. Finally, notice that for a given $A \in \mathcal{A}$, $\{\rho_o(A_o) \mid A_o \in \mathcal{A},\ A_o \leq A\}$ is an upward net in $\mathcal{M}$. Hence

$$\begin{aligned}
\text{Bel}(A) = \mu(\rho(A)) &= \mu(\vee\{\rho_o(A_o) \mid A_o \in \mathcal{A}_o;\ A_o \leq A\}) \\
&= \sup\{\mu(\rho_o(A_o)) \mid A_o \in \mathcal{A}_o;\ A_o \leq A\} \\
&= \sup\{\text{Bel}_o(A_o) \mid A_o \in \mathcal{A}_o,\ A_o \leq A\}.
\end{aligned}$$

And

$$P^*(A) = 1 - \mathrm{Bel}\,(\overline{A}) = 1 - \sup\,\{\mathrm{Bel}_o(A_o) \,|\, A_o \in \mathcal{A}_o,\ A_o \leq \overline{A}\}$$

$$= \inf\,\{1 - \mathrm{Bel}_o(A_o) \,|\, A_o \in \mathcal{A}_o,\ A_o \leq \overline{A}_o\}$$

$$= \sup\,\{P^*_o(A_o) \,|\, A_o \in \mathcal{A}_o,\ A \leq A_o\}. \qquad \blacksquare$$

I will call ρ, Bel and P* the natural extensions of ρ_o, Bel_o and P^*_o, respectively. It should be borne in mind that in general one's belief function on a Boolean algebra will not be the natural extension of its restriction to a given subalgebra. But it seems fair to characterize the cases where it is by saying that in those cases the restriction to the subalgebra "exhausts our opinions about the subject matter of the larger algebra." More concisely, I will say that an allocation $\rho: \mathcal{A} \to \mathcal{M}$ is supported by the subalgebra $\mathcal{A}_o$ of $\mathcal{A}$ whenever ρ is the natural extension of $\rho \,|\, \mathcal{A}_o$.

We have already seen one simple example where we wanted to adopt the natural extension of an allocation on a subalgebra -- namely, the Senate example in section 2 of Chapter 1. In that example, we obtained a belief function on a Boolean algebra corresponding to the field of all subsets of the set of twenty-two Senators. But in fact, that belief function was derived from a belief function (which happened to be a probability function) on the subalgebra corresponding to the field of all subsets of the set of eleven States. It is easily seen that the belief function we obtained on the larger Boolean algebra is the natural extension of the belief function on the subalgebra.

Let me give another example. Suppose we have a belief function concerning the possible values of an unknown quantity $\underset{\sim}{X}$ -- i.e., a belief function $\mathrm{Bel}_o: \mathcal{P}(\mathcal{S}_1) \to [0,1]$, where $\mathcal{S}_1$ is the set of all possible values of the quantity $\underset{\sim}{X}$ and $\mathrm{Bel}_o(A)$ is our degree of belief that the true value is in A. And suppose we have no opinions whatsoever about the value of a second unknown quantity $\underset{\sim}{Y}$, except the knowledge that it is in a set $\mathcal{S}_2$. And suppose we would like to define a belief function Bel: $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0,1]$ which would express our opinions about the values of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ simultaneously: we would like Bel(A) to be our degree of belief that the pair (x, y) is in A, where x is the true value of $\underset{\sim}{X}$ and y is the true value of $\underset{\sim}{Y}$. What should we do?

Well, $\mathcal{P}(\mathcal{S}_1)$ is naturally isomorphic to a subalgebra of $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. Figure 1 gives the familiar geometric picture: the horizontal axis corresponds to $\mathcal{S}_1$, the vertical axis to $\mathcal{S}_2$, the whole plane to $\mathcal{S}_1 \times \mathcal{S}_2$, and a subset A of $\mathcal{S}_1$ corresponds to a vertical "cylinder set" based on the subset A of the horizontal axis.

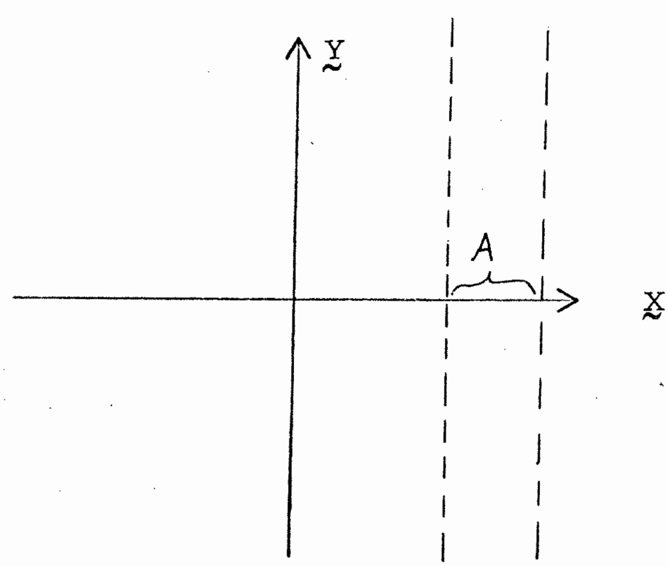

Figure 1

In symbols, the isomorphism $i: \mathcal{P}(\mathcal{S}_1) \xrightarrow{\text{into}} \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ is given by $i(A) = \{(x, y) \mid x \in A,\ y \in \mathcal{S}_2\} = A \times \mathcal{S}_2$. So we should obviously adopt as our belief function on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ the natural extension of $Bel_o \circ i^{-1}$ on the subalgebra $i(\mathcal{P}(\mathcal{S}_1))$ of $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. This will result in the belief function $Bel: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0, 1]$ defined by

$$\begin{aligned} Bel(A) &= \sup \{Bel_o \circ i^{-1}(A_o) \mid A_o \in i(\mathcal{P}(\mathcal{S}_1)),\ A_o \subset A\} \\ &= \sup \{Bel_o(A_o) \mid A_o \subset \mathcal{S}_1,\ i(A_o) \subset A\} \\ &= \sup \{Bel_o(A_o) \mid A_o \subset \mathcal{S}_1,\ A_o \times \mathcal{S}_2 \subset A\} \\ &= Bel_o(\{x \mid \{x\} \times \mathcal{S}_2 \subset A\}). \end{aligned}$$

In other words, A is awarded the degree of belief of the largest vertical cylinder set that is contained in A.

2. Restricted Allocations

In the preceding section, we saw how to obtain an allocation or belief function on a Boolean algebra of propositions $\mathcal{A}$ starting with an allocation or belief function on a subalgebra $\mathcal{A}_o$. Actually, the same sort of extension can be carried out even when the original allocation is on a subset of $\mathcal{A}$ which falls short of being a subalgebra by failing to include negations of some of its elements or disjunctions of some pairs of its elements.

Of course, our definitions for the notions of an allocation and a belief function apply only to a Boolean algebra, but they do not involve negations or disjunctions in any essential way and hence can be trivially generalized. This is done in the following definitions.

Definition. I will call a subset $\mathcal{L}$ of a Boolean algebra $\mathcal{A}$ a subtrellis of $\mathcal{A}$ if

(i) $\Lambda_{\mathcal{A}} \in \mathcal{L}$,

(ii) $V_{\mathcal{A}} \in \mathcal{L}$,

and (iii) $A_1 \wedge A_2 \in \mathcal{L}$ whenever $A_1, A_2 \in \mathcal{L}$. (This terminology is not standard.)

Definition. Suppose $\mathcal{L}$ is a subtrellis of a Boolean algebra $\mathcal{A}$. Then a function $\mathrm{Bel}: \mathcal{L} \to [0, 1]$ is a restricted belief function if

(i) $\mathrm{Bel}(\Lambda_{\mathcal{A}}) = 0$,

(ii) $\mathrm{Bel}(V_{\mathcal{A}}) = 1$,

and (iii) $\mathrm{Bel}(A) \geq \Sigma\, \mathrm{Bel}(A_i) - \Sigma\, \mathrm{Bel}(A_i \wedge A_j) +- \ldots +(-1)^{n+1}$

$\mathrm{Bel}(A_1 \wedge \ldots \wedge A_n)$

for all collections, $A, A_1, \ldots, A_n$ of elements of $\mathcal{L}$ such that $A_i \leq A$ for $i = 1, \ldots, n$.

Definition. Suppose $\mathcal{L}$ is a subtrellis of a Boolean algebra $\mathcal{A}$ and $(\mathcal{M}, \mu)$ is a measure algebra. Then a mapping $\rho: \mathcal{L} \to \mathcal{M}$ is a restricted allocation of probability if

(i) $\rho(\Lambda_{\mathcal{A}}) = \Lambda_{\mathcal{M}}$

(ii) $\rho(V_{\mathcal{A}}) = V_{\mathcal{M}}$

(iii) $\rho(A_1 \wedge A_2) = \rho(A_1) \wedge \rho(A_2)$ whenever $A_1, A_2 \in \mathcal{L}$. If $\mathcal{M}$ is a probability algebra, then ρ is called standard.

Interestingly enough, our theory for allocations and belief functions remains largely valid for the restricted variety. In particular, if $\rho: \mathcal{L} \to \mathcal{M}$ is a restricted allocation and μ is the measure on $\mathcal{M}$, then $\mathrm{Bel} = \mu \circ \rho$ will be a restricted belief function. And any restricted belief function can be represented in this way, where $(\mathcal{M}, \mu)$ is a probability algebra. These facts can be verified by noting that the proofs of Chapter 2 remain valid almost word for word for the restricted case.

It might seem desirable to cast our whole theory in a more general form by admitting restricted belief functions as belief functions. But such a generalization is unnecessary, precisely because a restricted allocation or a restricted belief function on a subtrellis $\mathcal{L}$ of a Boolean algebra $\mathcal{Q}$ can always be naturally extended to a belief function or allocation on $\mathcal{Q}$.

<u>Theorem</u>. Suppose $\mathcal{L}$ is a subtrellis of a Boolean algebra $\mathcal{Q}$ and $\rho_o: \mathcal{L} \to \mathcal{M}$ is a standard restricted allocation. Then the mapping $\rho: \mathcal{Q} \to \mathcal{M}$ given by

$$\rho(A) = \vee \{ \rho_o(L) \mid L \in \mathcal{L},\ L \leq A \}$$

is a standard allocation on $\mathcal{Q}$. Furthermore, $\rho \mid \mathcal{L} = \rho_o$. And if μ denotes the measure on $\mathcal{M}$, then the belief function $\mathrm{Bel} = \mu \circ \rho$ on $\mathcal{Q}$ and the restricted belief function $\mathrm{Bel}_o = \mu \circ \rho_o$ on $\mathcal{L}$ are related by

$$\mathrm{Bel}(A) = \sup \{ \Sigma \mathrm{Bel}_o(L_i) - \Sigma \mathrm{Bel}_o(L_i \wedge L_j) +- \ldots + (-1)^{n+1} \mathrm{Bel}_o (L_1 \wedge \ldots \wedge L_n) \mid n \geq 1; L_1, \ldots, L_n \in \mathcal{L}; \text{ and } L_i \leq A \text{ for } i = 1, \ldots, n \}$$

for all $A \in \mathcal{Q}$.

<u>Proof:</u> The proof that ρ is a standard allocation and $\rho \mid \mathcal{L} = \rho_o$ is precisely the same as the proof of the analogous assertions in the preceding section.

To verify the formula for Bel(A), notice that $\{\rho_o(L_1) \vee \ldots \vee \rho_o(L_n) \mid n \geq 1, L_1, \ldots, L_n \in \mathcal{L}; L_i \leq A \text{ for all } i\}$ is an upward net in $\mathcal{M}$. Hence, denoting by μ the measure on $\mathcal{M}$, we have

$$\mathrm{Bel}(A) = \mu(\rho(A)) = \mu(\vee \{ \rho_o(L) \mid L \in \mathcal{L}; L \leq A\})$$

$$= \mu(\vee \{\rho_o(L_1) \vee \ldots \vee \rho_o(L_n) \mid L_1, \ldots, L_n \in \mathcal{L}; L_i \leq A \text{ for all } i\})$$

$$= \sup \{ \mu(\rho_o(L_1) \vee \ldots \vee \rho_o(L_n)) \mid L_1, \ldots, L_n \in \mathcal{L}; L_i \leq A \text{ for all } i \}$$

$$= \sup \{ \Sigma \mu(\rho_o(L_i)) - \Sigma \mu(\rho_o(L_i \wedge L_j)) +- \ldots + (-1)^{n+1} \mu(\rho_o(L_1 \wedge \ldots \wedge L_n)) \mid L_1, \ldots, L_n \in \mathcal{L}; L_i \leq A \text{ for all } i \}$$

$$= \sup \{ \Sigma \mathrm{Bel}_o(L_i) - \Sigma \mathrm{Bel}_o(L_i \wedge L_j) +- \ldots + (-1)^{n+1} \mathrm{Bel}_o(L_1 \wedge \ldots \wedge L_n) \mid L_1, \ldots, L_n \in \mathcal{L}; L_i \leq A \text{ for all } i\}.$$

Of course, I will call Bel and ρ the natural extension to $\mathcal{Q}$ of Bel_o and ρ_o, respectively.

3. The Combination of Belief Functions

In section 1 I discussed an example of extension that involved two unknown quantities $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ with sets $\mathcal{S}_1$ and $\mathcal{S}_2$ of possible values, respectively. Beginning with a belief function $\mathrm{Bel}_o: \mathcal{P}(\mathcal{S}_1) \to [0,1]$ and operating on the assumption that I had no opinions about the value of $\underset{\sim}{Y}$, I obtained a belief function $\mathrm{Bel}: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0,1]$. But of course even when I have no opinions about $\underset{\sim}{Y}$ I can still claim to have a belief function Bel_2 on $\mathcal{P}(\mathcal{S}_2)$; it will be the vacuous belief function:

$$\mathrm{Bel}_2(A) = \begin{cases} 0 & \text{if } A \neq \mathcal{S}_2 \\ 1 & \text{if } A = \mathcal{S}_2. \end{cases}$$

So instead of thinking of $\mathrm{Bel}: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0,1]$ as the result of extending Bel, we can think of it as the result of combining Bel_1 on $\mathcal{P}(\mathcal{S}_1)$ with the vacuous belief function Bel_2 on $\mathcal{P}(\mathcal{S}_2)$.

This example raises the question of whether there is a natural general rule for combining belief functions on different Boolean algebras. More precisely, when Bel_1 is a belief function on the Boolean algebra $\mathcal{A}_1$, and Bel_2 is a belief function on the Boolean algebra $\mathcal{A}_2$, is there a natural way of combining the two to obtain a belief function Bel on $\mathcal{A} = \mathcal{A}_1 \oplus \mathcal{A}_2$?

Recall that $\mathcal{A}_1$ and $\mathcal{A}_2$ are thought of as independent subalgebras of $\mathcal{A}$. So one could begin to define Bel on $\mathcal{A}$ by setting $\mathrm{Bel}(A) = \mathrm{Bel}_1(A)$ when $A \in \mathcal{A}_1$ and $\mathrm{Bel}(A) = \mathrm{Bel}_2(A)$ when $A \in \mathcal{A}_2$. But many elements of $\mathcal{A}$

are in neither $\mathcal{A}_1$ nor $\mathcal{A}_2$. For example if $A_1 \epsilon \mathcal{A}_1$ and $A_2 \epsilon \mathcal{A}_2$ and neither A_1 nor A_2 is the zero or the unit, then $A_1 \wedge A_2$ will be in neither $\mathcal{A}_1$ nor in $\mathcal{A}_2$.

So suppose $A_1 \epsilon \mathcal{A}_1$, $Bel_1(A_1) = \alpha_1$, $A_2 \epsilon \mathcal{A}_2$, and $Bel_2(A_2) = \alpha_2$. Then what degree of belief should we assign to $A_1 \wedge A_2$? Well, Bel_1 directs us to commit α_1 of our belief to A_1, and Bel_2 directs us to commit α_2 of our belief to A_2. Supposing that we have already carried out Bel_1's directions, then the natural procedure is to apply Bel_2's directions not just to our probability as a whole, but to every probability mass, including the probability mass of measure α_1 that is committed to A_1. Hence we would commit α_2 of that probability mass, or a probability mass of measure $\alpha_1 \cdot \alpha_2$, to A_2 as well and hence to $A_1 \wedge A_2$. At any rate, this would be the natural procedure if Bel_1 and Bel_2 were derived from independent sources of information.

So we have a method for determining a degree of belief for each element $A \epsilon \mathcal{A}$ that can be represented in the form $A = A_1 \wedge A_2$, where $A_1 \epsilon \mathcal{A}_1$ and $A_2 \epsilon \mathcal{A}_2$: we set $Bel(A) = Bel_1(A_1) \cdot Bel_2(A_2)$. This quantity is well-defined; for if $A \neq \Lambda$, then the representation $A = A_1 \wedge A_2$ is unique; while if $A = \Lambda$, then either A_1 or A_2 is the zero and $Bel_o(A) = Bel_1(A_1) \cdot Bel_2(A_2) = 0$.

But the set $\mathcal{L} = \{A \mid A = A_1 \wedge A_2;\ A_1 \epsilon \mathcal{A}_1;\ A_2 \epsilon \mathcal{A}_2\}$ is a subtrellis of $\mathcal{A}$. Indeed, $\Lambda = \Lambda \wedge \Lambda$, $V = V \wedge V$, and $(A_1 \wedge A_2) \wedge (A_1' \wedge A_2') = (A_1 \wedge A_1') \wedge (A_2 \wedge A_2')$ is in $\mathcal{L}$ whenever $A_1, A_1' \epsilon \mathcal{A}_1$ and $A_2, A_2' \epsilon \mathcal{A}_2$. So we have a function $Bel_o: \mathcal{L} \rightarrow [0,1]: A_1 \wedge A_2 \rightsquigarrow Bel_1(A_1) \cdot Bel_2(A_2)$ on a subtrellis

$\mathcal{L}$. If this were a restricted belief function on $\mathcal{L}$ (and we have not shown that it is), then by the theory of the preceding section, we could extend it to a belief function Bel on $\mathcal{A}$ that would be given by

$$\mathrm{Bel}(A) = \sup \{\Sigma\, \mathrm{Bel}_o(L_i) - \Sigma\, \mathrm{Bel}_o(L_i \wedge L_j) + - \dots + (-1)^{n+1}\, \mathrm{Bel}_o(L_1 \wedge \dots \wedge L_n) \mid L_1, \dots, L_n \in \mathcal{L};\ L_i \leq A \text{ for all } i\}$$

$$= \sup \{\Sigma\, \mathrm{Bel}_1(A_i) \cdot \mathrm{Bel}_2(B_i) - \Sigma\, \mathrm{Bel}_1(A_i \wedge A_j) \cdot \mathrm{Bel}_2(B_i \wedge B_j) + - \dots + (-1)^{n+1}\, \mathrm{Bel}_1(A_1 \wedge \dots \wedge A_n)\, \mathrm{Bel}_2(B_1 \wedge \dots \wedge B_n) \mid A_1, \dots, A_n \in \mathcal{A}_1;\ B_1, \dots, B_n \in \mathcal{A}_2; \text{ and } A_i \wedge B_i \leq A,\ i = 1, \dots, n\}$$

But how shall we show that Bel_o is a restricted belief function on $\mathcal{L}$?

The easiest way is to turn to the theory of allocations.

<u>Theorem.</u> Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$, and suppose $\rho_1{}^o: \mathcal{A}_1 \to \mathcal{M}_1$ and $\rho_2{}^o: \mathcal{A}_2 \to \mathcal{M}_2$ are standard allocations with belief functions $\mathrm{Bel}_1 = \mu_1 \circ \rho_1{}^o$ and $\mathrm{Bel}_2 = \mu_2 \circ \rho_2{}^o$, where μ_1 and μ_2 are the measures on $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. Let $((\mathcal{M}, \mu); i_1: \mathcal{M}_1 \to \mathcal{M};\ i_2: \mathcal{M}_2 \to \mathcal{M})$ be an orthogonal sum of $(\mathcal{M}_1, \mu_1)$ and $(\mathcal{M}_2, \mu_2)$. Then $\rho_1 = i_1 \circ \rho_1{}^o$ and $\rho_2 = i_2 \circ \rho_2^o$ will be standard allocations of $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively, into $\mathcal{M}$; $\mathrm{Bel}_1 = \mu \circ \rho_1$ and $\mathrm{Bel}_2 = \mu \circ \rho_2$. Now define $\rho: \mathcal{A} \to \mathcal{M}$ by

$$\rho(A) = \vee \{\rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\} \qquad (1)$$

Then ρ is an allocation of probability. Denote Bel = $\mu \circ \rho$. Then $\text{Bel}|\mathcal{A}_1 = \text{Bel}_1$, $\text{Bel}|\mathcal{A}_2 = \text{Bel}_2$, and in general

$$\begin{aligned}\text{Bel}(A) = \sup\{&\Sigma\, \text{Bel}_1(A_i) \cdot \text{Bel}_2(B_i) - \Sigma\, \text{Bel}_1(A_i \wedge A_j) \cdot \text{Bel}_2(B_i \wedge B_j) \\ &+- \ldots + (-1)^{n+1}\, \text{Bel}_1(A_1 \wedge \ldots \wedge A_n)\, \text{Bel}_2(B_1 \wedge \ldots \wedge B_n) \,\big| \\ &n \geq 1;\ A_1, \ldots, A_n \in \mathcal{A}_1;\ B_1, \ldots, B_n \in \mathcal{A}_2;\ \text{and } A_i \wedge B_i \leq A, \\ &i = 1, \ldots, n\}.\end{aligned} \tag{2}$$

Proof: Let $\mathcal{L}$ be the subtrellis of all elements of $\mathcal{A}$ of the form $A_1 \wedge A_2$, where $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. Define $\rho_o: \mathcal{L} \to \mathcal{M}$ by $\rho_o(A) = \rho_1(A_1) \wedge \rho_2(A_2)$ whenever $A = A_1 \wedge A_2$, with $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. Since the representation $A = A_1 \wedge A_2$ is unique when $A \neq \Lambda$, ρ_o is well-defined. It is easily verified that ρ_o is a restricted allocation, and obviously $\rho_o|\mathcal{A}_1 = \rho_1$ and $\rho_o|\mathcal{A}_2 = \rho_2$. By the theorem in section 2, the formula (1) defines the natural extension of ρ_o to $\mathcal{A}$, and Bel = $\mu \circ \rho$ is given by (2). And since $\mathcal{A}_1$ and $\mathcal{A}_2$ are subsets of $\mathcal{L}$, $\text{Bel}|\mathcal{A}_i = \mu \circ \rho|\mathcal{A}_i = \mu \circ \rho_o|\mathcal{A}_i = \mu \circ \rho_i = \text{Bel}_i$ for $i = 1, 2$. ▨

From formula (2) it is evident that Bel does not depend on the choice of ρ_1 and ρ_2 or on the choice of the orthogonal sum $(\mathcal{M}, \mu)$. Hence

I will call Bel the orthogonal sum of Bel_1 and Bel_2 on $\mathcal{A}$, and sometimes I will denote it as $Bel_1 \oplus Bel_2$. Notice that $\mathcal{A}_1$ and $\mathcal{A}_2$ can be independent subalgebras of more than one Boolean algebra; hence it may be necessary to specify the algebra $\mathcal{A}$ when speaking of the orthogonal sum of Bel_1 on $\mathcal{A}_1$ and Bel_2 on $\mathcal{A}_2$. But usually this will make no practical difference, for if $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}_o$ and $\mathcal{A}_o$ is a subalgebra of $\mathcal{A}$, then the orthogonal sum of Bel_1 and Bel_2 on $\mathcal{A}$ is simply the extension of the orthogonal sum on $\mathcal{A}_o$, and both are given by (2).

In particular, given belief functions Bel_1 and Bel_2 on Boolean algebras $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively, (2) will give the orthogonal sum $Bel_1 \oplus Bel_2$ on $\mathcal{A}_1 \oplus \mathcal{A}_2$. And given belief functions Bel_1 and Bel_2 on power sets $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$, respectively, (2) will give the orthogonal sum $Bel_1 \oplus Bel_2$ on the power set $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. In this latter case, (2) becomes

$$
\begin{aligned}
Bel(A) = \sup \{ & \Sigma\, Bel_1(A_i)\, Bel_2(B_i) - \Sigma\, Bel_1(A_i \cap A_j)\, Bel_2(B_i \cap B_j) \\
& + - \ldots + (-1)^{n+1}\, Bel(A_1 \cap \ldots \cap A_n)\, Bel(B_1 \cap \ldots \cap B_n) \mid \\
& n \geq 1,\ A_1, \ldots, A_n \subset \mathcal{S}_1,\ B_1, \ldots, B_n \subset \mathcal{S}_2;\ A_i \times B_i \subset A, \\
& i = 1, \ldots, n\}
\end{aligned} \tag{2'}
$$

This brings us back to the example with which we began. In that case, Bel_2 is the vacuous belief function, and (2') becomes

$$
\begin{aligned}
Bel(A) = \sup \{ & \Sigma\, Bel_1(A_i) - \Sigma\, Bel_1(A_i \cap A_j) + - \ldots + (-1)^{n+1} \\
& Bel_1(A_1 \cap \ldots \cap A_n) \mid A_1, \ldots, A_n \subset \mathcal{S}_1;\ A_i \times \mathcal{S}_2 \subset A, \\
& i = 1, \ldots, n\}
\end{aligned}
$$

$$= \sup\{ Bel_1(A_1) | A_1 \subset \mathcal{S}_1,\ A_1 \times \mathcal{S}_2 \subset A\}$$

$$= Bel_1(\{ x \mid \{x\} \times \mathcal{S}_2 \subset A\}).$$

This does indeed agree with the method of extension.

<u>Theorem.</u> Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$, and Bel: $\mathcal{A} \to [0, 1]$ is the orthogonal sum of $Bel_1 : \mathcal{A}_1 \to [0, 1]$ and $Bel_2 : \mathcal{A}_2 \to [0, 1]$, and let P^*, P_1^* and P_2^* denote the upper probability functions corresponding to Bel, Bel_1 and Bel 2, respectively. Then for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,

$$\text{(i)}\quad Bel\,(A_1 \wedge A_2) = Bel_1(A_1) \cdot Bel_2(A_2),$$

and

$$\text{(ii)}\quad P^*(A_1 \wedge A_2) = P_1^*(A_1) \cdot P_2^*(A_2).$$

<u>Proof:</u> (i) is clear from the preceding theorem, but (ii) is more difficult. Let $(\mathcal{M}, \mu)$, ρ, ρ_1, ρ_2 be as in the preceding theorem, and let ζ, ζ_1 and ζ_2 be the allowments corresponding to ρ, ρ_1 and ρ_2, respectively. Then $P^* = \mu \circ \zeta$, $P_1^* = \mu \circ \zeta_1$, $P_2^* = \mu \circ \zeta_2$, and since $\zeta_1(\mathcal{A}_1)$ and $\zeta_2(\mathcal{A}_2)$ are in orthogonal subalgebras of $\mathcal{M}$, we can establish (ii) by showing that $\zeta(A_1 \wedge A_2) = \zeta(A_1) \wedge \zeta(A_2)$ whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. But in such a case,

$$\zeta(A_1 \wedge A_2) = \overline{\rho\ (\overline{A_1 \wedge A_2})}$$

$$= \overline{\vee \{ \rho_1(A) \wedge \rho_2(B) \mid A \in \mathcal{A}_1;\ B \in \mathcal{A}_2;\ A \wedge B \leq \overline{A_1 \wedge A_2}\}}$$

$$= \wedge \{ \overline{\rho_1(A) \wedge \rho_2(B)} \mid A \in \mathcal{A}_1;\ B \in \mathcal{A}_2:\ A \wedge B \leq \overline{A_1 \wedge A_2}\}$$

$$= \wedge \{\overline{\rho_1(\overline{A}_1)} \vee \overline{\rho_2(\overline{B})} \mid A \in \mathcal{A}_1;\ B \in \mathcal{A}_2;\ \overline{A} \wedge \overline{B} \leq \overline{A_1 \wedge A_2}\}$$

$$= \wedge \{ \zeta_1(A) \vee \zeta_2(B) \mid A \in \mathcal{O}_1;\ B \in \mathcal{O}_2;\ A_1 \wedge A_2 \leq A \vee B \}.$$

But notice that $\zeta_1(A_1) \vee \zeta_2(\Lambda) = \zeta_1(A_1)$ and $\zeta_1(\Lambda) \vee \zeta_2(A_2) = \zeta_2(A_2)$ are in this last meet. And whenever $A \in \mathcal{O}_1$, $B \in \mathcal{O}_2$ and $A_1 \wedge A_2 \leq A \vee B$, we know (by the second theorem of Chapter 3, section 9) that either $A_1 \leq A$ or $A_2 \leq B$. Hence every other probability mass in the meet will contain either $\zeta_1(A_1)$ or $\zeta_2(A_2)$ and hence, in any case, $\zeta_1(A_1) \wedge \zeta_2(A_2)$. Hence the meet is equal to $\zeta_1(A_1) \wedge \zeta_2(A_2)$. ▨

4. A Combinatorial Lemma

Lemma. Suppose m and n are positive integers, $I \subset \{1, \ldots, n\}$, $J \subset \{1, \ldots, m\}$, and I and J are non-empty. Set

$$\mathcal{K} = \{K \mid \emptyset \neq K \subset \{1, \ldots, n\} \times \{1, \ldots, m\};\ I = \{i \mid (i,j) \in K \text{ for some } j\};\ J = \{j \mid (i,j) \in K \text{ for some } i\}\}.$$

Then

$$\sum_{K \in \mathcal{K}} (-1)^{1 + \text{card } K} = (-1)^{\text{card } I + \text{card } J}.$$

Proof: Set card I = i and Card J = j, and denote $L = \{1, \ldots, i\} \times \{1, \ldots, j\}$, and think of L as an i x j matrix. I will call a subset A of L a covering of L if A contains at least one entry in every row and every column of L. I will call such a covering even or odd according as it contains an even or odd number of entries. I will prove the following assertion: The number of odd coverings of L is one greater than the number of even coverings if $i + j$ is even, and one less if $i + j$ is odd. In symbols: #(odd coverings) - # (even coverings) = $(-1)^{i+j}$.

The proof will be by induction on $i+j$. Since I and J are non-empty, $i + j \geq 2$; and if $i + j = 2$, the assertion is trivially true. Indeed, it is trivially true whenever $i = 1$ or $j = 1$ So suppose that $i + j = k$, that the assertion is true whenever $i + j < k$, and that $i > 1$ and $j > 1$. Let L_o be the $(i - 1) \times (j - 1)$ matrix obtained by omitting the first row and column of L. Let R and C be the subsets of L indicated in Figure 2. Then by our inductive hypothesis, our assertion holds for the $(i-1) \times (j-1)$ matrix L_o, the $i \times (j-1)$ matrix $R \cup L_o$ and the $(i-j) \times j$ matrix $C \cup L_o$.

~~Let us~~ classify the coverings of L according as they (i) intersect both R and C, (ii) intersect R but not C, (iii) intersect C but not R, or (iv) intersect neither R nor C.

Consider category (i). Some of the coverings in this category contain (1, 1) but they will remain coverings if (1, 1) is omitted. Hence

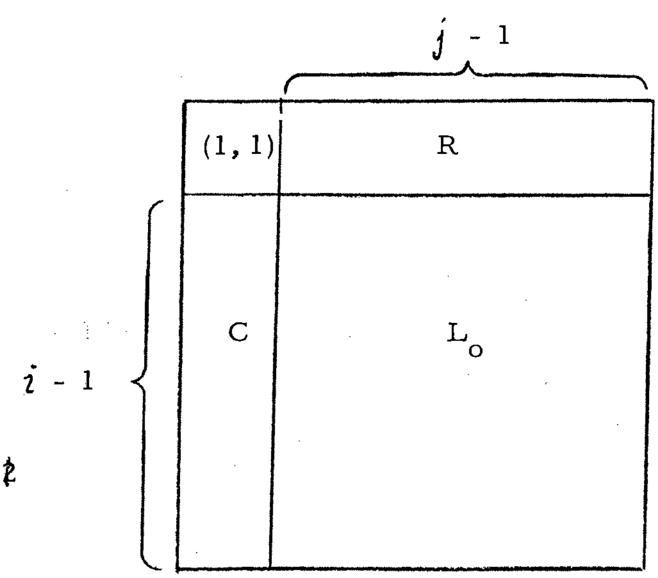

Figure 2

the coverings in this category can be arranged in pairs, the two members of which differ only in that one contains (1, 1) and the other does not. Hence there are the same number of even as odd coverings in this category.

Consider category (ii). Each covering in this category must contain (1, 1). As a matter of fact, each one is obtained from a covering of $R \cup L_o$ by adding (1, 1). Hence for this category #(odd coverings) - #(even coverings) = #(even coverings of $R \cup L_o$) - #(odd coverings of $R \cup L_o$) = $-(-1)^{i+(j-1)} = (-1)^{i+j}$.

It can be shown quite analogously for category (iii) that #(odd coverings) - #(even coverings) = $-(-1)^{i-1)+j} = (-1)^{i+j}$.

Finally, consider category (iv). Each covering in this category must contain L_o and must also be a covering of L_o. As a matter of fact, the elements of this category are obtained by taking coverings of L_o and adding (1, 1). Hence for this category #(odd coverings) - #(even coverings) = #(even coverings for L_o) - #(odd coverings for L_o) = $-(-1)^{(i-1)+(j-1)} = (-1)^{i+j-1}$.

Adding the results for all four categories, we find that ~~overall~~ #(odd coverings) - #(even coverings) = $(-1)^{i+j} + (-1)^{i+j} + (-1)^{i+j-1}$ $= (-1)^{i+j}$.

The lemma follows immediately ~~from~~ this result.

<u>Corollary.</u> Suppose $(\mathcal{M}, \mu)$ is a probability algebra, $\mathcal{E}_o$ and $\mathcal{F}_o$ are subtrellises of $\mathcal{M}$, and

$$\mu(E \wedge F) = \mu(E) \cdot \mu(F) \tag{3}$$

for all $E \epsilon \mathcal{E}_o$ and $F \epsilon \mathcal{F}_o$. Denote by $\mathcal{E}$ and $\mathcal{F}$ the subalgebras of $\mathcal{M}$ generated by $\mathcal{E}_o$ and $\mathcal{F}_o$, respectively. Then (1) holds for all $E \in \mathcal{E}$ and $F \in \mathcal{F}$.

Proof: Consider first elements E and F of $\mathcal{M}$ of the form

$$E = E_o \wedge \overline{E}_1 \wedge \ldots \wedge \overline{E}_K = (E_o \vee E_1 \vee \ldots \vee E_K) - (E_1 \vee \ldots \vee E_K)$$

and

$$F = F_o \wedge \overline{F_1} \wedge \ldots \wedge \overline{F}_\ell = (F_o \vee F_1 \vee \ldots \vee F_\ell) - (F_1 \vee \ldots \vee F_\ell),$$

where $E_o, E_1, \ldots, E_K \in \mathcal{E}_o$ and $F_o, F_1, \ldots, F_\ell \in \mathcal{F}_o$. We have

$$E \wedge F = E_o \wedge F_o \wedge \overline{E}_1 \wedge \ldots \wedge \overline{E}_K \wedge \overline{F_1} \wedge \ldots \wedge \overline{F}_\ell$$

$$= (E_o \wedge F_o) \wedge \overline{E_1 \vee \ldots \vee E_K \vee F_1 \vee \ldots \vee F_\ell}$$

$$= [(E_o \wedge F_o) \vee E_1 \vee \ldots \vee E_K \vee F_1 \vee \ldots \vee F_\ell] -$$

$$[(E_1 \vee \ldots \vee E_K \vee F_1 \vee \ldots \vee F_\ell],$$

and

$$\mu(E \wedge F) = \sum_{I \subset \{1,\ldots,k\}} \sum_{J \subset \{1,\ldots,\ell\}} (-1)^{\text{card } I + \text{card } J} \cdot$$

$$\mu\left(E_o \wedge F_o \wedge \left(\bigwedge_{i \in I} E_i\right) \wedge \left(\bigwedge_{j \in J} F_j\right)\right)$$

$$= \sum_{I \subset \{1,\ldots,k\}} (-1)^{\text{card } I} \mu\left(E_o \wedge \left(\bigwedge_{i \in I} E_i\right)\right)$$

$$\times \sum_{J \subset \{1,\ldots,\ell\}} (-1)^{\text{card } J} \mu\left(F_o \wedge \left(\bigwedge_{j \in J} F_j\right)\right)$$

$$= \mu(E) \cdot \mu(F).$$

Now by section 7 of Chapter 3, any element $E \in \mathcal{E}$ can be written in the form

$$E = E_1 \vee \ldots \vee E_m,$$

where for each i, i = 1, . . . , m,

$$E_i = E_{io} \wedge \overline{E_{i1}} \wedge \ldots \wedge \overline{E_{ik_i}}$$

for some elements E_{io}, E_{i1}, . . . , E_{ik_i} of $\mathcal{E}_o$. Similarly, any element $F \varepsilon \mathcal{F}$ can be written in the form

$$F = F_1 \vee \ldots \vee F_n,$$

where for each i, i = 1, . . . , n,

$$F_i = F_{io} \wedge \overline{F_{i1}} \wedge \ldots \wedge \overline{F_{i\ell_i}}$$

for some elements F_{io}, F_{i1}, . . . , $F_{i\ell_i}$ of $\mathcal{F}_o$. If E and F are expressed in this way, then

$$E \wedge F = (E_1 \vee \ldots \vee E_m) \wedge (F_1 \vee \ldots \vee F_n)$$

$$= \bigvee_{i=1}^{m} \bigvee_{j=1}^{n} (E_i \wedge F_j).$$

And by the lemma,

$$\mu(E \wedge F) = \mu\Big(\bigvee_{i=1}^{m} \bigvee_{j=1}^{n} (E_i \wedge E_j)\Big)$$

$$= \sum_{\substack{K \subset \{1, \ldots, m\} \times \{1, \ldots, n\} \\ K \neq \phi}} (-1)^{1 + \text{card } K} \mu\Big(\bigwedge_{(i,j) \varepsilon K} (E_i \wedge F_j)\Big)$$

$$= \sum_{\substack{I \subset \{1, \ldots, m\} \\ I \neq \phi}} (-1)^{\text{card } I} \sum_{J \subset \{1, \ldots, n\}} (-1)^{\text{card } J}$$

$$\mu\Big(\big(\bigwedge_{i \varepsilon I} E_i\big) \wedge \big(\bigwedge_{j \varepsilon J} F_j\big)\Big).$$

But

$$\bigwedge_{i\in I} E_i = (\bigwedge_{i\in I} E_{io}) \wedge (\bigwedge_{i\in I} (\overline{E}_{i1} \wedge \ldots \wedge \overline{E}_{ik_i})),$$

where E_{io}, E_{i1}, ..., E_{ik_i} are all in $\mathcal{E}_o$ for all i; and $\bigwedge_{j\in J} F_j$ is of a similar form. Hence by the first paragraph

$\mu((\bigwedge_{i\in I} E_i) \wedge (\bigwedge_{j\in J} F_j)) = \mu(\bigwedge_{i\in I} E_i)\, \mu(\bigwedge_{j\in J} F_j)$. So

$$\mu(E \wedge F) = \sum_{\substack{I \subset \{1, \ldots, m\} \\ I \neq \emptyset}} (-1)^{1+\text{card } I} \mu(\bigwedge_{i\in I} E_i)$$

$$\times \sum_{\substack{J \subset \{1, \ldots, n\} \\ J \neq \emptyset}} (-1)^{1+\text{card } J} \mu(\bigwedge_{j\in J} F_j)$$

$$= \mu(E) \cdot \mu(F).$$

5. Orthogonality and Independence

As we have just seen, our rule of combination obeys a multiplicative rule for both Bel and P*. In this section, I will explore the implications of these two rules.

Definition. Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$, and suppose Bel: $\mathcal{A} \to [0, 1]$ is a belief function. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel if

$$\text{Bel}(A_1 \wedge A_2) = \text{Bel}(A_1) \cdot \text{Bel}(A_2) \quad (1)$$

whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. And $\mathcal{A}_1$ and $\mathcal{A}_2$ are cognitively independent with respect to Bel if

$$P^*(A_1 \wedge A_2) = P^*(A_1) \cdot P^*(A_2) \tag{2}$$

whenever $A_1 \epsilon \mathcal{Q}_1$ and $A_2 \epsilon \mathcal{Q}_2$. ($P^*$ is, of course, the upper probability function corresponding to Bel.)

A justification for the term "cognitively independent" will be offered in the next chapter. The term "orthogonal," on the other hand, can be justified immediately.

Theorem. Suppose $\mathcal{Q}$ is a Boolean algebra, $(\mathcal{M}, \mu)$ is a probability algebra, and $\rho: \mathcal{Q} \to \mathcal{M}$ is a standard representation for the belief function Bel on $\mathcal{Q}$. Then two independent subalgebras $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are orthogonal with respect to Bel if and only if the subalgebras of $\mathcal{M}$ generated by $\rho(\mathcal{Q}_1)$ and $\rho(\mathcal{Q}_2)$ are orthogonal with respect to μ.

Proof: Denote by $\mathcal{M}_1$ and $\mathcal{M}_2$ the subalgebras of $\mathcal{M}$ generated by $\rho(\mathcal{Q}_1)$ and $\rho(\mathcal{Q}_2)$, respectively. Clearly, if $A_1 \epsilon \mathcal{Q}_1$ and $A_2 \epsilon \mathcal{Q}_2$, then $\rho(A_1) \epsilon \mathcal{M}_1$ and $\rho(A_2) \epsilon \mathcal{M}_2$, so that the orthogonality of $\mathcal{M}_1$ and $\mathcal{M}_2$ will imply (1).

Suppose, on the other hand, that (1) holds for all $A_1 \epsilon \mathcal{Q}_1$ and all $A_2 \epsilon \mathcal{Q}_2$. Then since $\rho(\mathcal{Q}_1)$ and $\rho(\mathcal{Q}_2)$ are subtrellises, it follows by the corollary in the preceding section that

$$\mu(M_1 \wedge M_2) = \mu(M_1) \cdot \mu(M_2)$$

for all $M_1 \epsilon \mathcal{M}_1$ and $M_2 \epsilon \mathcal{M}_2$. Since μ is positive it follows that $\mathcal{M}_1$ and $\mathcal{M}_2$ are independent subalgebras and hence orthogonal.

▨

It is not obvious at first glance that orthogonality and cognitive independence are distinct conditions, and hence it is worthwhile to provide examples showing that each of the conditions can hold without the other holding. To this end, set $\mathcal{A} = \mathcal{P}(\mathcal{S})$, where $\mathcal{S} = \{a, b, c, d\}$ as shown in Figure 3. And set $\mathcal{A}_1 = \{\emptyset, \{a,b\}, \{c,d\}, \mathcal{S}\}$ and $\mathcal{A}_2 = \{\emptyset\ \{a, c\}, \{b, d\}, \mathcal{S}\}$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$. Let us define two belief functions Bel_1 and Bel_2 on $\mathcal{A}$ as follows: Bel_1 is given by the basic probability numbers $\{m_A\}_{A \in \mathcal{A}}$, where

$$m_{\mathcal{S}\{a,b,c\}} = 1/4,$$
$$m_{\{a,b\}} = 1/4,$$
$$m_{\{a,c\}} = 1/4,$$
$$m_{\{a\}} = 1/4,$$

and $m_A = 0$ for all other $A \in \mathcal{A}$. And Bel_2 is given by the basic probability numbers $\{m'_A\}_{A \in \mathcal{A}}$, where

$$m'_{\mathcal{S}} = 1/4,$$
$$m'_{\{a,b,c\}} = 1/4,$$
$$m'_{\{a\}} = 1/2,$$

and $m'_A = 0$ for all other $A \in \mathcal{A}$. Then it can be verified that $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal but not cognitively independent with respect to Bel_1 and cognitively independent but not orthogonal with respect to Bel_2.

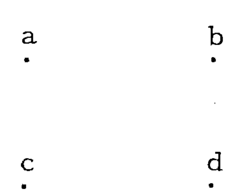

Figure 3

As we saw in section 3, when a belief function on $\mathcal{A}$ is the orthogonal sum of belief functions on independent subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$, those subalgebras are both orthogonal and cognitively independent with respect to that belief function. In fact a converse of this theorem is also true; if two independent subalgebras are both orthogonal and cognitively independent with respect to a belief function, then on the subalgebra generated by the union of the two subalgebras that belief function will agree with the orthogonal sum of its restrictions to the two subalgebras. This assertion follows from the following theorem.

<u>Theorem.</u> Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of a Boolean algebra $\mathcal{A}$, and suppose $\mathcal{A}$ is the subalgebra generated by $\mathcal{A}_1 \cup \mathcal{A}_2$. Suppose Bel: $\mathcal{A} \to [0,1]$ is a belief function and $\rho: \mathcal{A} \to \mathcal{M}$ is a standard representation for Bel. Then the following conditions are all equivalent:

(1) $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal and cognitively independent with respect to Bel.

(2) $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel and $\rho(A \vee B) = \rho(A) \vee \rho(B)$ whenever $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$.

(3) $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal with respect to Bel and

$\rho(A) = \vee\{\rho(A_1 \wedge A_2)\ \ A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\}$ for all $A \in \mathcal{A}$.

(4) For all $A \in \mathcal{A}$,

$$\begin{aligned} \mathrm{Bel}(A) = \sup\{&\Sigma\,\mathrm{Bel}(A_i) \cdot \mathrm{Bel}(B_i) - \Sigma\,\mathrm{Bel}(A_i \wedge A_j) \cdot \mathrm{Bel}(B_i \wedge B_j) \\ &+- \dots + (-1)^{n+1}\,\mathrm{Bel}(A_1 \wedge \dots \wedge A_n)\,\mathrm{Bel}(B_1 \wedge \dots \wedge B_n) \,| \\ &n \geq 1;\ A_1, \dots, A_n \in \mathcal{A}_1;\ B_1, \dots, B_m \in \mathcal{A}_2;\ \text{and } A_i \wedge B_i \leq A, \\ &i = 1, \dots, n\}. \end{aligned}$$

<u>Proof</u>: (1) $\Rightarrow$ (2). Suppose $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$. Then by orthogonality and cognitive independence,

$$1 - Bel(A \vee B) = P^*(\overline{A} \wedge \overline{B}) = P^*(\overline{A}) \cdot P^*(\overline{B})$$

$$= (1 - Bel(A))\,(1 - Bel(B))$$

$$= 1 - Bel(A) - Bel(B) + Bel(A \wedge B),$$

Hence

$$Bel(A \vee B) = Bel(A) + Bel(B) - Bel(A \wedge B),$$

or

$$\mu(\rho\,(A \vee B)) = \mu\,(\rho\,(A) \vee \rho\,(B)).$$

Since μ is positive, it follows that

$$\rho\,(A \vee B) = \rho\,(A) \vee \rho\,(B).$$

(2) $\Rightarrow$ (3). Since $\mathcal{A}$ is the subalgebra generated by $\mathcal{A}_1 \cup \mathcal{A}_2$, every element $A \in \mathcal{A}$ must be of the form

$$A = (A_1 \wedge B_1) \vee \ldots \vee (A_n \wedge B_n),$$

where the A_i are all in $\mathcal{A}_1$ and the B_i are all in $\mathcal{A}_2$. Hence, by (2),

$$\rho(A) = \rho\,(A_1 \wedge B_1) \vee \ldots \vee \rho\,(A_n \wedge B_n),$$

and (3) follows.

(3) $\Rightarrow$ (4). For any $A \in \mathcal{A}$,

$$Bel(A) = \mu\,(\rho\,(A))$$

$$= \mu(\vee\,\{\rho\,(A_1 \wedge A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\})$$

$$= \mu\,(\vee\{\rho\,(A_1 \wedge B_1) \vee \ldots \vee \rho\,(A_n \wedge B_n) \mid A_i \in \mathcal{A}_1,\ B_i \in \mathcal{A}_2$$

$$\text{and } A_i \wedge B_i \leq A \text{ for all } i\}\,)$$

$$= \sup \{ \mu (\rho(A_1 \wedge B_1) \vee \ldots \vee \rho(A_n \wedge B_n)) \mid A_i \in \mathcal{Q}_1,\ B_i \in \mathcal{Q}_2 \text{ and } A_i \wedge B_i \leq A \text{ for all } i \}$$

$$= \sup \{ \Sigma \mathrm{Bel}(A_i) \cdot \mathrm{Bel}(B_i) - \Sigma \mathrm{Bel}(A_i \wedge A_j) \cdot \mathrm{Bel}(B_i \wedge B_j) +- \ldots + (-1)^{n+1} \mathrm{Bel}(A_1 \wedge \ldots \wedge A_n) \mathrm{Bel}(B_1 \wedge \ldots \wedge B_n) \mid n \geq 1;\ A_1, \ldots, A_n \in \mathcal{Q}_1;\ B_1, \ldots, B_n \in \mathcal{Q}_2;\ A_i \wedge B_i \leq A,\ i = 1, \ldots, n \}.$$

(4) $\Longrightarrow$ (1). This is merely a restatement of the last theorem of section 3. ▨

Finally, it is useful to note that the formulae in (2) and (3) can also be stated in terms of the allowment ζ. In terms of ζ, (2) becomes

(2') $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are orthogonal with respect to Bel and $\bar{\zeta}(A \wedge B) = \zeta(A) \wedge \zeta(B)$ whenever $A \in \mathcal{Q}_1$ and $B \in \mathcal{Q}_2$;

and (3) becomes

(3') $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are orthogonal with respect to Bel and

$$\zeta(A) = \wedge \{ \zeta(A_1 \vee A_2) \mid A_1 \in \mathcal{Q}_1;\ A_2 \in \mathcal{Q}_2;\ A_1 \vee A_2 \geq A \}.$$

6. The Finite Case

Recall that a belief function Bel on a finite Boolean algebra $\mathcal{Q}$ is completely determined by the basic probability numbers m_A for $A \in \mathcal{Q}$. These numbers are non-negative, $m_{\Lambda} = 0$, and Bel is given by

$$\mathrm{Bel}(A) = \sum_{A' \leq A} m_{A'}.$$

Intuitively, the basic probability number m_A measures the total probability mass that is constrained to A but not to any proper sub-element of A. In other words, if $\rho: \mathcal{A} \to \mathcal{M}$ is an allocation representing Bel, then

$$m_A = \mu(\rho(A) - \vee\{\rho(A') \mid A' < A\}),$$

where μ is the measure on $\mathcal{M}$. It is worth noting how these basic probability numbers behave under combination.

<u>Theorem</u>. Suppose $\rho: \mathcal{A} \to \mathcal{M}$ is a standard allocation on the finite Boolean algebra $\mathcal{A}$, and suppose the independent subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\mathcal{A}$ are orthogonal and cognitively independent with respect to ρ. Denote by $\{m_A\}_{A \in \mathcal{A}}$ the basic probability numbers for ρ, by $\{n_{A_1}\}_{A_1 \in \mathcal{A}_1}$ the basic probability numbers for $\rho | \mathcal{A}_1$ and by $\{p_{A_2}\}_{A_2 \in \mathcal{A}_2}$ the basic probability numbers for $\rho | \mathcal{A}_2$. Then

$$m_A = \begin{cases} n_{A_1} \cdot p_{A_2} & \text{whenever } A = A_1 \wedge A_2 \text{ with } A_1 \in \mathcal{A}_1 \text{ and } A_2 \in \mathcal{A}_2 \\ 0 & \text{if } A \neq A_1 \wedge A_2 \text{ for any } A_1 \in \mathcal{A}_1 \text{ and } A_2 \in \mathcal{A}_2 . \end{cases}$$

<u>Proof</u>: First consider the case where $A \neq A_1 \wedge A_2$ for any $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. In that case, $A_1 \wedge A_2 < A$ whenever $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$ and $A_1 \wedge A_2 \leq A$. Hence

$$\rho(A) = \vee\{\rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\}$$

$$= \vee\{\rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 < A\}$$

$$= \vee\{\rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A' \in \mathcal{A};\ A_1 \wedge A_2 \leq A' < A\}$$

$$= \vee \{ \vee \{ \rho(A_1) \wedge \rho(A_2) \mid A_1 \in \mathcal{Q}_1;\ A_2 \in \mathcal{Q}_2;\ A_1 \wedge A_2 \leq A' \} \mid A' < A \}$$

$$= \vee \{ \rho(A') \mid A' < A \},$$

so

$$m_A = \mu(\rho(A) - \vee \{ \rho(A') \mid A' < A \}) = \mu(\Lambda) = 0.$$

Now consider the case where $A = A_1 \wedge A_2$ with $A_1 \in \mathcal{Q}_1$ and $A_2 \in \mathcal{Q}_2$. In that case,

$$\begin{aligned}
&(\rho(A_1) - \vee \{ \rho(A_1') \mid A_1' \in \mathcal{Q}_1;\ A_1' < A_1 \}) \\
&\wedge (\rho(A_2) - \vee \{ \rho(A_2') \mid A_2' \in \mathcal{Q}_2;\ A_2' < A_2 \})
\end{aligned}$$

$$\begin{aligned}
&= \rho(A_1) \wedge \rho(A_2) - \vee \{ \rho(A_1') \mid A_1' \in \mathcal{Q}_1;\ A_1' < A_1 \} \\
&\quad - \vee \{ \rho(A_2') \mid A_2' \in \mathcal{Q}_2;\ A_2' < A_2 \}
\end{aligned}$$

$$\begin{aligned}
&= \rho(A_1) \wedge \rho(A_2) - \vee \{ \rho(A_1') \vee \rho(A_2') \mid A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2; \\
&\quad A_1' < A_1;\ A_2' < A_2 \}
\end{aligned}$$

$$\begin{aligned}
&= \rho(A_1) \wedge \rho(A_2) - \vee \{ [\rho(A_1') \vee \rho(A_2')] \wedge \rho(A_1) \wedge \rho(A_2) \mid \\
&\quad A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2;\ A_1' < A_1;\ A_2' < A_2 \}
\end{aligned}$$

$$\begin{aligned}
&= \rho(A_1) \wedge \rho(A_2) - \vee \{ [\rho(A_1') \wedge \rho(A_2)] \vee [\rho(A_1) \wedge \rho(A_2')] \mid \\
&\quad A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2;\ A_1' < A_1;\ A_2' < A_2 \}
\end{aligned}$$

$$\begin{aligned}
&= \rho(A_1) \wedge \rho(A_2) - \vee \{ \rho(A_1') \wedge \rho(A_2') \mid A_1' \in \mathcal{Q}_1;\ A_2' \in \mathcal{Q}_2; \\
&\quad A_1' \leq A_1;\ A_2' \leq A_2;\ \text{either } A_1' < A_1 \text{ or } A_2' < A_2 \}
\end{aligned}$$

$$= \rho(A_1 \wedge A_2) - \vee\{\rho(A_1' \wedge A_2') \mid A_1' \epsilon \mathcal{A}_1;\ A_2' \epsilon \mathcal{A}_2;$$
$$A_1' \leq A_1;\ A_2' \leq A_2;\ \text{either } A_1' < A_1 \text{ or } A_2' < A_2\}$$

$$= \rho(A) - \vee\{\rho(A') \mid A' < A\}$$

The last few equalities depend on the thorem of Chapter 3, section 9.

Since $\rho(\mathcal{A}_1)$ and $\rho(\mathcal{A}_2)$ are in orthogonal subalgebras of $\mathcal{M}$, the measure of $\rho(A) - \vee\{\rho(A') \mid A' < A\}$ must equal the product of the measures of

$$\rho(A_1) - \vee\{\rho(A_1') \mid A_1' \epsilon \mathcal{A}_1;\ A_1' < A_1\}$$

and

$$\rho(A_2) - \vee\{\rho(A_2') \mid A_2' \epsilon \mathcal{A}_2,\ A_2' < A_2\}.$$

In other words, $m_A = n_{A_1} \cdot p_{A_2}$. ▨

7. The Condensable Case

In this section, we will see how the orthogonal sum of two condensable belief functions can be described in terms of the commonality numbers.

When we are dealing with two condensable belief functions, say one on $\mathcal{P}(\mathcal{S}_1)$ and one on $\mathcal{P}(\mathcal{S}_2)$, it is most natural to consider their orthogonal sum on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$. This orthogonal sum will itself be condensable, as we see from the following theorem.

Theorem. Suppose Bel: $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0,1]$ is a belief function, $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$ are orthogonal and cognitively independent with respect to Bel, and Bel $|\mathcal{P}(\mathcal{S}_1)$ and Bel$|\mathcal{P}(\mathcal{S}_2)$ are condensable. Then Bel is condensable.

Proof: Let $\zeta: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to \mathcal{M}$ be a standard allowment for Bel, and recall that ζ is condensable if and only if

$$\zeta(A) = \bigvee_{s \in A} \zeta(\{s\})$$

for all $A \subset \mathcal{S}_1 \times \mathcal{S}_2$. Now by orthogonality and cognitive independence

$$\zeta(A) = \wedge \{ \zeta(A_1 \times \mathcal{S}_2) \vee \zeta(\mathcal{S}_1 \times A_2) \,|\, A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ (A_1 \times \mathcal{S}_2) \cup (\mathcal{S}_1 \times A_2) \supset A \}$$

for all $A \subset \mathcal{S}_1 \times \mathcal{S}_2$. Since the restrictions to $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$ are condensable, this becomes

$$\zeta(A) = \wedge \{ (\bigvee_{s_1 \in A_1} \zeta(\{s_1\} \times \mathcal{S}_2)) \vee (\bigvee_{s_2 \in A_2} \zeta(\mathcal{S}_1 \times \{s_2\})) \,|\, A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ (A_1 \times \mathcal{S}_2) \cup (\mathcal{S}_1 \times A_2) \supset A \}.$$

$$= \vee \{ \zeta(\{s_1\} \times \mathcal{S}_2) \vee \zeta(\mathcal{S}_1 \times \{s_2\}) \,|\, (s_1, s_2) \in A \}$$

$$= \bigvee_{(s_1, s_2) \in A} (\wedge \{ \zeta(A_1 \times \mathcal{S}_2) \vee \zeta(\mathcal{S}_1 \times A_2) \,|\, A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ (A_1 \times \mathcal{S}_2) \cup (\mathcal{S}_1 \times A_2) \supset \{(s_1, s_2)\} \})$$

$$= \bigvee_{(s_1, s_2) \in A} \zeta(\{(s_1, s_2)\}).$$

And furthermore, the commonality numbers for Bel are obtained from those for Bel $|\mathcal{P}(\mathcal{S}_1)$ and Bel $|\mathcal{P}(\mathcal{S}_2)$ by a simple multiplicative rule.

Theorem. Suppose Bel: $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ is condensable and $\mathcal{P}(\mathcal{S})_1$ and $\mathcal{P}(\mathcal{S}_2)$ are orthogonal and cognitively independent with respect to Bel. Let

$$Q_1 : \mathcal{T}(\mathcal{S}_1) \rightarrow [0, 1],$$

$$Q_2 : \mathcal{T}(\mathcal{S}_2) \rightarrow [0, 1],$$

and $\quad Q : \mathcal{T}(\mathcal{S}_1 \times \mathcal{S}_2) \rightarrow [0, 1]$

be the commonality functions for Bel $|\mathcal{P}(\mathcal{S}_1)$, Bel $|\mathcal{P}(\mathcal{S}_2)$ and Bel, respectively. Then

$$Q(\{(a_1, b_1), \ldots, (a_n, b_n)\}) = Q_1(\{a_1, \ldots, a_n\} Q_2(\{b_1, \ldots, b_n\})$$

for all $\{(a_1, b_1), \ldots, (a_n, b_n)\} \subset \mathcal{S}_1 \times \mathcal{S}_2$.

Proof: Letting ς be the allowment and μ the measure on the probability algebra, we have

$$Q(\{(a_1, b_1), \ldots, (a_n, b_n)\}) = \mu(\varsigma(\{(a_1, b_1)\}) \wedge \ldots \wedge \varsigma(\{(a_n, b_n)\})),$$

$$Q_1(\{a_1, \ldots, a_n\}) = \mu(\varsigma(\{a_1\} \times \mathcal{S}_2) \wedge \ldots \wedge \varsigma(\{a_n\} \times \mathcal{S}_2)),$$

and $\quad Q_2(\{b_1, \ldots, b_n\}) = \mu(\varsigma(\mathcal{S}_1 \times \{b_1\}) \wedge \ldots \wedge \varsigma(\mathcal{S}_1 \times \{b_n\})).$

But by (2') from section 5, we know that

$$\varsigma(\{(a_i, b_i)\}) = \varsigma(\{a_i\} \times \mathcal{S}_2) \wedge \varsigma(\mathcal{S}_1 \times \{b_i\})$$

for all i. Hence

$$\varsigma(\{(a_1, b_1)\} \wedge \ldots \wedge \varsigma(\{(a_n\ b_n)\})$$

$$= (\varsigma(\{a_1\} \times \mathcal{S}_2) \wedge \ldots \wedge \varsigma(\{a_n\} \times \mathcal{S}_2)) \wedge (\varsigma(\mathcal{S}_1 \times \{b_1\}) \wedge \ldots \wedge \varsigma(\mathcal{S}_1 \times \{b_n\}));$$

and the theorem follows by orthogonality. $\blacksquare$

8. An Example of Combination

In this section I will illustrate the rule of combination with a simple example.

Suppose Mr. and Mrs. Jones are discussing over their breakfast coffee whether they should attend a ballet in the evening. Mr. Jones has no opinions about how enjoyable the ballet may prove to be, yet has opinions about whether it will rain, while Mrs. Jones has no inkling as to whether it will rain yet has definite ideas about the quality of the ballet. Assuming that they trust each other's judgments in their respective areas of competency, how might Mr. and Mrs. Jones combine their opinions in order to obtain, as it were, a joint opinion about the possibility of attending an enjoyable ballet without getting wet?

Let us be more concrete. Suppose Mr. Jones has a belief function Bel_1 on $\mathcal{P}(\mathcal{S}_1)$, where $\mathcal{S}_1$ = {rain, no rain}, and Mrs. Jones has a belief function Bel_2 on $\mathcal{P}(\mathcal{S}_2)$, where $\mathcal{S}_2$ = {enjoyable ballet, unenjoyable ballet}. And suppose Bel_1 and Bel_2 are given by

$Bel_1(\emptyset) = 0$	$Bel_2(\emptyset) = 0$
$Bel_1(\{\text{rain}\}) = 1/2$	$Bel_2(\{\text{enjoyable ballet}\} = 1/2$
$Bel_1(\{\text{no rain}\}) = 0$	$Bel_2(\{\text{unenjoyable ballet}\} = 1/3$
$Bel_1(\mathcal{S}_1) = 1$	$Bel_2(\mathcal{S}_2) = 1.$

These two belief functions can also be described by saying that Bel_1 is given by the basic probability numbers $\{n_A\}_{A \subset \mathcal{S}_1}$ and Bel_2 is given by the basic probability numbers $\{p_A\}_{A \subset \mathcal{S}_2}$, where

$n_{\emptyset} = 0$ $\qquad$ $p_{\emptyset} = 0$

$n_{\{\text{rain}\}} = 1/2$ $\qquad$ $p_{\{\text{enjoyable ballet}\}} = 1/2$

$n_{\{\text{no rain}\}} = 0$ $\qquad$ $p_{\{\text{unenjoyable ballet}\}} = 1/3$

$n_{\mathcal{S}_1} = 1/2$ $\qquad$ $p_{\mathcal{S}_2} = 1/6.$

In other words, Mr. Jones puts half of his probability on the occurrence of rain and does not commit the other half, while Mrs. Jones puts half of her probability on an enjoyable ballet and a third of it on an unenjoyable one. If we represent each person's probability by a mass that is uniformly distributed over a line segment, then we can depict this situation as in Figure 4.

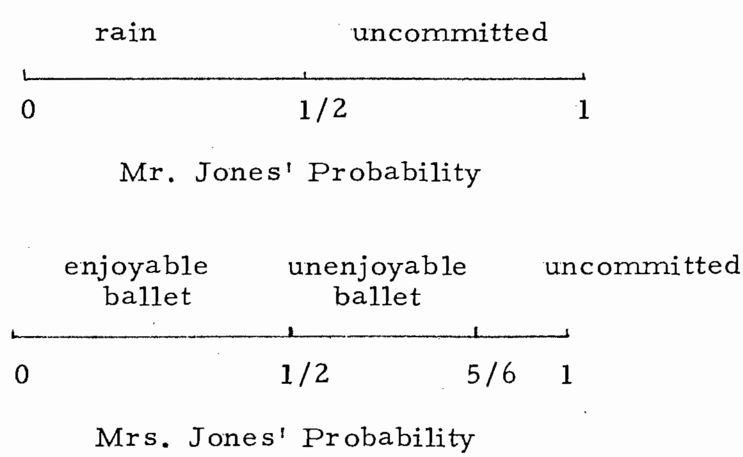

Figure 4

We require a combined belief function Bel on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$; and in particular we require a degree of belief and an upper probability for the subset $\{\text{no rain}\} \times \{\text{enjoyable ballet}\}$ of $\mathcal{S}_1 \times \mathcal{S}_2$.

Let us consider the matter from Mrs. Jones' point of view. Her belief function Bel_2 can be described by the three basic probability masses shown in Figure 4. Now she is confronted with Mr. Jones' opinions about the weather and decides to adopt them as her own. What does this mean? Well, the message from Mr. Jones can be stated simply: Put half your probability on rain. The natural thing for Mrs. Jones to do is to carry out this recommendation for each of her three basic probability masses: she should commit half of each of them to rain.

The result can be depicted geometrically if we use a square instead of a line segment to represent Mrs. Jones' probability. In the first panel of Figure 5, Mrs. Jones' three basic probability masses are depicted, each labelled with its "region of mobility". The second panel shows the situation after she has committed half of each of her probability masses to rain but left the other halves uncommitted between rain and no rain.

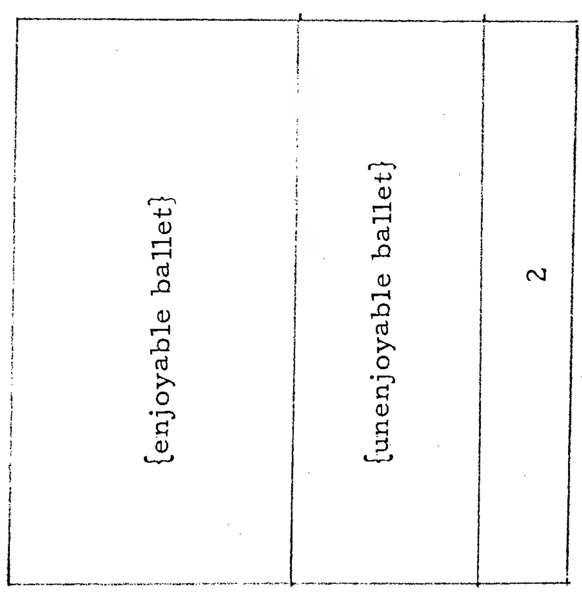

Figure 5a. Before

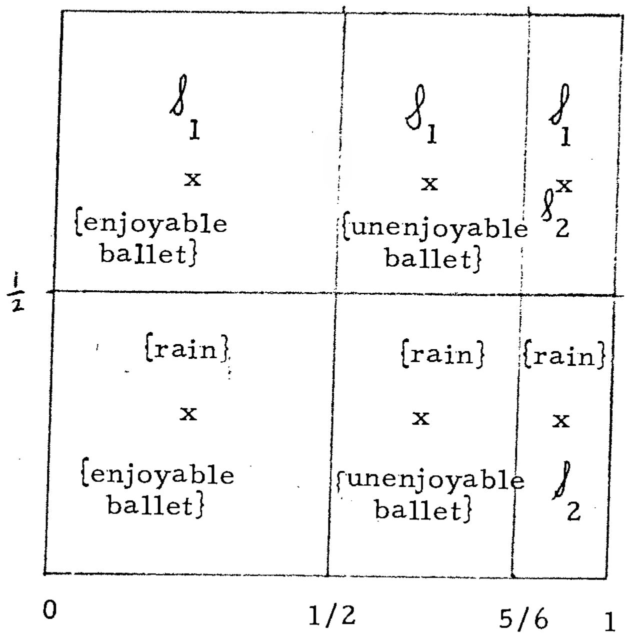

Figure 5b. After

So we obtain six basic probability masses, with the following corresponding basic probability numbers:

$$m_{\{\text{rain}\} \times \{\text{enjoyable ballet}\}} = 1/4$$

$$m_{\{\text{rain}\} \times \{\text{unenjoyable ballet}\}} = 1/6$$

$$m_{\{\text{rain}\} \times \mathcal{S}_2} = 1/12$$

$$m_{\mathcal{S}_1 \times \{\text{enjoyable ballet}\}} = 1/4$$

$$m_{\mathcal{S}_1 \times \{\text{unenjoyable ballet}\}} = 1/6$$

$$m_{\mathcal{S}_1 \times \mathcal{S}_2} = 1/12.$$

The basic probability numbers m_A for other $A \subset \mathcal{S}_1 \times \mathcal{S}_2$ are, of course, zero.

The belief function Bel on $\mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2)$ can be easily computed from this table of basic probability numbers. For example, we find that

$$\text{Bel}(\{\text{no rain}\} \times \{\text{enjoyable ballet}\}) = 0$$

and

$$P^*(\{\text{no rain}\} \times \{\text{enjoyable ballet}\}) = 1/3.$$

CHAPTER 7. DEMPSTER'S RULES OF CONDITIONING AND COMBINATION

In this chapter I adduce Demster's rules for modifying a belief function on the basis of new evidence or opinion. Dempster's rule of conditioning tells us how to modify a belief function Bel: $\mathcal{C} \rightarrow [0, 1]$ when we learn that $A \in \mathcal{C}$ is true. His more general rule of combination tells us how to modify Bel when the evidence underlying it is pooled with independent evidence underlying a second belief function Bel': $\mathcal{C} \rightarrow [0, 1]$.

In section 7, we will see how the rule of combination provides a justification for the term "cognitively independent," which was introduced in the preceding chapter.

1. Dempster's Rule of Conditioning

The central feature of the theory of subjective probability is its rule of conditioning. The rule is open to criticism but it has a tremendous intuitive appeal and has always been accepted by students of subjective probability. In this section, I will describe the rule from an intuitive point of view and introduce the analogous rule for belief functions.

Suppose we are dealing with a set $\mathcal{S}$ which is the set of all possible values of some quantity $\underset{\sim}{s}$ whose true value is unknown, and suppose we have a probability function

$$P: \mathcal{P}(\mathcal{S}) \rightarrow [0, 1],$$

P(S) being our degree of belief (or subjective probability) that the true value of $\underset{\sim}{s}$ is in S. Then we can describe our situation intuitively by saying that our probability is distributed over the set $\mathcal{S}$. Now suppose we learn, from new evidence, that the true value of $\underset{\sim}{s}$ is really in a proper subset $\mathcal{S}_o$ of $\mathcal{S}$. Then if $P(\mathcal{S}_o) < 1$ our probability function P will evidently require modification, for we will now wish to assert a degree of belief 1 in $\mathcal{S}_o$. How should P be modified?

The obvious thing to do is to "throw away" that portion of our probability that was distributed over $\mathcal{S}_o$; it was committed to something that is now seen as impossible, so it seems that the only thing that can be done is to discard it. This will leave us, of course, with a total amount of probability that has measure $P(\mathcal{S}_o)$, which may be less than one. Hence we will want to "renormalize" the measure of all our remaining probability, multiplying all the measures by $1/P(\mathcal{S}_o)$ so as to bring the measure of the total back up to one again.

This procedure will result in a new probability function P' over $\mathcal{S}$, one that now gives $P'(\mathcal{S}_o) = 1$. In order to describe this probability function explicitly, let us refer to Figure 1 and calculate the value of P' for each of the sets S_1, S_2 and S_3 shown there. First of all, all the probability that was committed to S_1 has been thrown away; hence we now have

$$P'(S_1) = 0. \tag{1}$$

As for S_2, none of the probability associated with it has been thrown away, but its measure has been renormalized, so we have

$$P'(S_2) = P(S_2) \;/\; P(\mathcal{S}_o). \tag{2}$$

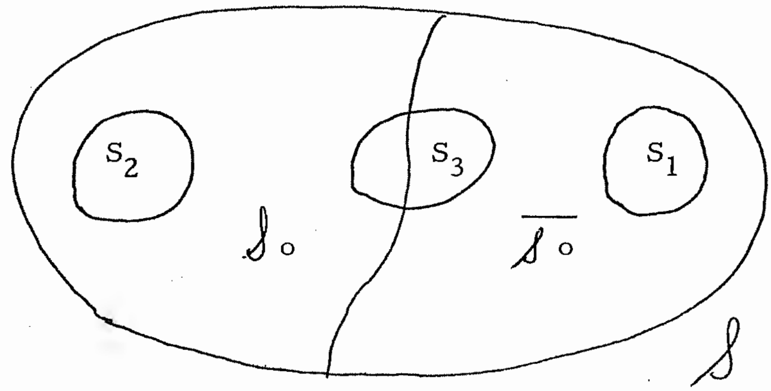

Figure 1.

Finally, consider S_3. Some of the probability that was distributed over S_3, namely the portion which was distributed over $S_3 \cap \overline{\mathcal{S}_o}$, has been eliminated. Hence the portion remaining is what was distributed over $S_3 \cap \mathcal{S}_o$, which did have measure $P(S_3 \cap \mathcal{S}_o)$ and now has measure

$$P'(S_3) = P(S_3 \cap \mathcal{S}_o)/P(\mathcal{S}_o). \tag{3}$$

An examination of (1), (2) and (3) shows that (1) and (2) are actually special cases of (3) , which is thus the general rule for conditioning P on $\mathcal{S}_o$.

The fact that P' is conditional on $\mathcal{S}_o$ is often indicated by denoting it by $P_{\mathcal{S}_o}$ or $P(\cdot \mid \mathcal{S}_o)$. In these notations, our rule becomes

$$P_{\mathcal{S}_o}(S) = P(S \cap \mathcal{S}_o)/P(\mathcal{S}_o)$$

or

$$P(S \mid \mathcal{S}_o) = P(S \cap \mathcal{S}_o) \,/\, P(\mathcal{S}_o) \tag{4}$$

for all $S \subset \mathcal{S}$. This is the classical rule for conditional probability; it is easily verified directly that $P(\cdot \mid \mathcal{S}_o)$ does indeed satisfy the axioms for probability functions, provided only that $P(\mathcal{S}_o) > 0$. Of course, if

$P(\mathcal{S}_o) = 0$ then our new knowledge that the true value of $\underset{\sim}{s}$ is in $\mathcal{S}_o$ is in direct contradiction with P, and the conditioning cannot be carried out.

An analogous rule applies, of course, to a probability function P on any Boolean algebra $\mathcal{Q}$. If $P(A) > 0$, then conditioning P on A yields a probability function $P(\cdot \mid A)$ on $\mathcal{Q}$ given by

$$P(B \mid A) = P(B \wedge A) \;/\; P(A) \tag{5}$$

for all $B \epsilon \mathcal{Q}$.

The intuition behind this classical rule generalizes directly to the case of belief functions. For suppose we begin with a belief function

$$\text{Bel}: \mathcal{P}(\mathcal{S}) \to [0, 1]$$

and then learn that the true value of $\underset{\sim}{s}$ is actually in $\mathcal{S}_o \subset \mathcal{S}$. What do we do? Well, we eliminate the probability that is committed to $\overline{\mathcal{S}_o}$ and renormalize the rest; the measure of the probability eliminated is Bel $(\overline{\mathcal{S}_o})$, so the measure of the remainder will be $1 - \text{Bel}(\overline{\mathcal{S}_o})$ and the constant of renormalization will be $(1 - \text{Bel}(\overline{\mathcal{S}_o}))^{-1}$. There is only one new idea that must be introduced: since our probability is allocated in a semi-mobile way over $\mathcal{S}$ rather than being distributed in a fixed way, we must recognize that the restriction to $\mathcal{S}_o$ may further restrict the mobility of some of our probability without eliminating it entirely. This means that some of our probability that was not committed to a set $S \subset \mathcal{S}$ may become committed to S by the restriction to $\mathcal{S}_o$. In fact, any probability that was committed to $S \cup \overline{\mathcal{S}_o}$ before will now be committed to S, unless it was committed to $\overline{\mathcal{S}_o}$ and hence must be eliminated. In general, then, the amount of probability committed to S after conditioning will be the measure of the probability previously committed to $S \cup \overline{\mathcal{S}_o}$

less the measure of the probability eliminated, or

$$\mathrm{Bel}(S \cup \overline{\mathcal{S}_0}) - \mathrm{Bel}(\overline{\mathcal{S}_0}).$$

But this must be renormalized, so we obtain

$$\mathrm{Bel}(S \mid \mathcal{S}_0) = \frac{\mathrm{Bel}(S \cup \overline{\mathcal{S}_0}) - \mathrm{Bel}(\overline{\mathcal{S}_0})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_0})} \tag{6}$$

as our degree of belief in S conditional on $\mathcal{S}_0$.

As it turns out, this rule is stated more easily in terms of the upper probability functions. Indeed,

$$P^*(S \mid \mathcal{S}_0) = 1 - \mathrm{Bel}(\overline{S} \mid \mathcal{S}_0)$$

$$= 1 - \frac{\mathrm{Bel}(\overline{S} \cup \overline{\mathcal{S}_0}) - \mathrm{Bel}(\overline{\mathcal{S}_0})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_0})}$$

$$= \frac{1 - \mathrm{Bel}(\overline{S} \cup \overline{\mathcal{S}_0})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_0})}$$

$$= \frac{1 - \mathrm{Bel}(\overline{S \cap \mathcal{S}_0})}{1 - \mathrm{Bel}(\overline{\mathcal{S}_0})},$$

or

$$P^*(S \mid \mathcal{S}_0) = \frac{P^*(S \cap \mathcal{S}_0)}{P^*(\mathcal{S}_0)} \quad . \tag{7}$$

This is <u>Dempster's rule of conditioning.</u> It is easily verified that $P^*(\cdot \mid \mathcal{S}_0)$ does indeed satisfy the rules for upper probability functions, provided only that $P^*(\mathcal{S}_0) > 0$. If $P^*(\mathcal{S}_0) = 0$, then our new knowledge that the true value of $\underset{\sim}{s}$ is in $\mathcal{S}_0$ is in direct contradiction with P*, and the conditioning cannot be carried out.

Dempster's rule of conditioning need not, of course, be restricted to upper probability functions on power sets; it can be applied to the conditioning of any upper probability function

$$P^*: \mathcal{Q} \to [0,1]$$

on any proposition $A \in \mathcal{Q}$ such that $P^*(A) > 0$. The resulting conditional upper probability function $P^*(\cdot \mid A)$ is given by

$$P^*(B \mid A) = \frac{P^*(B \wedge A)}{P^*(A)} \tag{8}$$

for all $B \in \mathcal{Q}$. If P^* is actually a probability function, this rule reduces to (5), the classical rule of conditional probability.

There is a difficulty with the application of the classical rule, and the generalization (8) might seem to suffer from the same difficulty. The difficulty is that we sometimes feel that $P(A) = 0$ does not really mean that A is impossible. In the case of a "continuous" distribution of probability P over a set $\mathcal{S}$, for example, $P(\{s\}) = 0$ for every $s \in \mathcal{S}$; yet this is not taken to mean that it is impossible for the true value of $\underset{\sim}{s}$ to be s for every $s \in \mathcal{S}$. Hence in general it may be impossible to carry out the conditioning even in cases where we would like to do so. Interestingly enough, though, <u>condensable</u> belief functions are exempt from this difficulty. Indeed, when an upper probability function $P^*: \mathcal{P}(\mathcal{S}) \to [0,1]$ is condensable we are entitled to interpret $P^*(S) = 0$ as meaning that P^* holds it to be impossible for the true value of $\underset{\sim}{s}$ to be in S. (See the end of section 1 of Chapter 5.) Hence our inability to condition a condensable upper probability on a set of upper probability zero need never be embarrassing, and the rule of conditioning appears to be most adapted to the condensable case.

It is easily verified that if $P^*: \mathcal{P}(\mathcal{S}) \to [0, 1]$ is condensable and $P^*(A) > 0$, then $P^*(\cdot|A)$ will also be condensable. And the commonality numbers are affected by conditioning in a very simple way. The commonality function Q for P^* is given, of course, by

$$Q(B) = -\sum_{T \subset B} (-1)^{\text{card } T} P^*(T)$$

for finite non-empty subsets B of $\mathcal{S}$. And the commonality function $Q(\cdot|A)$ for $P^*(\cdot|A)$ will be given by

$$Q(B|A) = -\sum_{T \subset B} (-1)^{\text{card } T} P^*(T|A)$$

$$= -\sum_{T \subset B} (-1)^{\text{card } T} \frac{P^*(T \cap A)}{P^*(A)}$$

$$= \frac{-1}{P^*(A)} \Big(\sum_{R \subset B \cap A} \sum_{S \subset B \cap \bar{A}} (-1)^{\text{card } R} (-1)^{\text{card } S} P^*((R \cup S) \cap A) \Big)$$

$$= \frac{-1}{P^*(A)} \Big(\sum_{R \subset B \cap A} (-1)^{\text{card } R} P^*(R) \Big) \Big(\sum_{S \subset B \cap \bar{A}} (-1)^{\text{card } S} \Big).$$

Now if $B \subset A$, then the last factor is equal to one; otherwise it is equal to zero. Hence

$$Q(B|A) = \begin{cases} 1 & \text{if } B = \phi \\ \dfrac{Q(B)}{P^*(A)} & \text{if } \phi \neq B \subset A, \\ o & \text{otherwise.} \end{cases}$$

So conditioning a condensable allocation can be carried out by renormalizing the relevant commonality numbers.

In the case of a belief function on a finite Boolean algebra $\mathcal{Q}$, it is also possible to describe the conditioning process in terms of the basic probability numbers. Suppose indeed that Bel: $\mathcal{Q} \to [0, 1]$ is given by the basic probability numbers $\{m_A\}_{A \in \mathcal{Q}}$. Then upon conditioning on A, the basic probability mass that was associated with $A' \in \mathcal{Q}$ will be constrained to $A' \wedge A$. Hence there will come to be associated with $B \in \mathcal{Q}$ a total basic probability mass of measure

$$\Sigma \{m_{A'} | A' \wedge A = B\}.$$

In particular a basic probability mass of measure

$$\Sigma \{m_{A'} | A' \wedge A = \Lambda\} = \mathrm{Bel}(\overline{A})$$

will come to be associated with Λ. This latter probability mass must of course be eliminated, and we must renormalize by the factor $(P^*(A))^{-1}$, thus obtaining the new basic probability numbers $\{m'_B\}_{B \in \mathcal{Q}}$ given by

$$m'_B = \frac{\Sigma \{ m_{A'} | A' \wedge A = B \}}{P^* (A)}$$

for all $B \neq \Lambda_{\mathcal{Q}}$ and, of course, $m_\Lambda = 0$.

2. The Conditional Allocation

Dempster's rule of conditioning is most simply described intuitively in terms of mobile probability masses: in order to condition Bel: $\mathcal{Q} \to [0,1]$ on $A \in \mathcal{Q}$, we add to the constraints on all our probability masses by constraining each one to A, and hence to $A \wedge A'$ for all $A' \in \mathcal{Q}$ to which it was previously constrained; we then eliminate all the probability that is constrained to Λ by this process. In order to represent this process mathematically, we must use the formal procedure that we learned in section 4 of Chapter 4 for "discarding" a probability mass from a probability algebra.

Theorem. Let $\rho: \mathcal{Q} \to \mathcal{M}$ be an allocation into the probability algebra $(\mathcal{M}, \mu)$. Suppose $A \in \mathcal{Q}$ and $\rho(\overline{A}) \neq \mathcal{V}$. Let I be the ideal in generated by $\rho(\overline{A})$, and let $(\mathcal{M}/I, \nu)$ be as in section 4 of Chapter 4. Let $f: \mathcal{M} \to \mathcal{M}/I$ be the canonical homomorphism. Then

$$\rho_A: \mathcal{Q} \to \mathcal{M}/I : A' \leadsto f(\rho(A' \vee \overline{A}))$$

is an allocation, and $\mathrm{Bel}_A = \nu \circ \rho_A$ is given by

$$\mathrm{Bel}_A(A') = \frac{\mathrm{Bel}(A' \vee \overline{A}) - \mathrm{Bel}(\overline{A})}{1 - \mathrm{Bel}(\overline{A})}$$

for all $A' \in \mathcal{Q}$.

Proof: It is easy to verify that ρ_A is an allocation:

(i) $\rho_A(\Lambda) = f(\rho(\overline{A})) = \Lambda$,

(ii) $\rho_A(\mathcal{T}) = f(\rho(\bar{A} \vee \mathcal{T})) = f(\mathcal{T}) = \mathcal{T}$,

(iii) $\rho_A(A_1 \wedge A_2) = f(\rho((A_1 \wedge A_2) \vee \overline{A})) = f(\rho((A_1 \vee \overline{A}) \wedge (A_2 \vee \overline{A})))$

$$= f(\rho(A_1 \vee \overline{A})) \wedge f(\rho(A_2 \vee \overline{A}))$$

$$= \rho_A(A_1) \wedge \rho_A(A_2).$$

And

$$\mathrm{Bel}_A(A') = \nu(f(\rho(A' \vee \overline{A}))) = \frac{1}{1 - \mu(\rho(\overline{A}))} \mu(\rho(A' \vee \overline{A}) - \rho(\overline{A}))$$

$$= \frac{\mathrm{Bel}(A' \vee \overline{A}) - \mathrm{Bel}(\overline{A})}{1 - \mathrm{Bel}(\overline{A})}$$

by the formula in section 4 of Chapter 4. ▨

The allocation ρ_A is called, of course, the conditional allocation obtained from ρ by conditioning on A.

3. Two Examples of Conditioning

In this section I will illustrate Dempster's rule of conditioning with two simple examples.

A. The Senate Example

First let us reconsider the example from Chapter 1 that involved an allocation of probability over the set of twenty-two Senators. That set is pictured again in Figure 2. Recall that our allocation of probability involved eleven basic probability masses, one corresponding to each

Langdon	(D)	Wingate	(D)
Few	(D)	Gunn	(D)
Lee	(D)	Grayson	(D)
Izard	(D)	Butler	(D)
Johnson	(D)	Ellsworth	(F)
Maclay	(D)	Morris	(F)
Strong	(F)	Dalton	(F)
Paterson	(F)	Elmer	(F)
Bassett	(F)	Read	(F)
Carroll	(F)	Henry	(F)
King	(F)	Schuyler	(F)

Figure 2.

State, and that each of these is free to move back and forth between the two Senators from the State to which it corresponds. We concluded that the degree of belief and the upper probability for the proposition A = "A Democratic-Republican will be chosen" were given by Bel(A) = 4/11 and P*(A) = 6/11.

Now Senator Maclay of Pennsylvania was particularly well known as a staunch anti-Federalist. Let us suppose that we begin with the allocation of probability just described but that we then learn -- say from a friend galloping past who pauses only to mention the fact with a sigh of relief -- that Maclay was <u>not</u> chosen. After the receipt of this information, what degree of belief and upper probability ought we to accord to the proposition A?

Well, we must condition our allocation of probability to the set $\overline{\{\text{Maclay}\}}$, i.e., to the set of the twenty-one Senators other than Maclay. This conditioning will not eliminate any of our probability, and it will change the region of mobility of only one of the eleven basic probability masses. The basic probability mass corresponding to the State of Pennsylvania, instead of moving freely between Senators Maclay and Morris, will now be constrained to Senator Morris. Hence there will still be only four basic probability masses constrained to Democratic-Republican Senators, but six of the seven remaining ones will be constrained to Federalist Senators. So conditionally we will have a degre of belief of 4/11 for A but an upper probability of only 5/11.

B. Conditioning on the Diagonal

In section 1 of Chapter 6 we considered an example in which we began with a belief function

$$\text{Bel}_o : \mathcal{P}(\mathcal{S}_1) \to [0, 1],$$

which expressed our degrees of belief about the true value of an unknown quantity $\underset{\sim}{X}$, $\mathcal{S}_1$ being the set of possible values of $\underset{\sim}{X}$. We also considered a second unknown quantity $\underset{\sim}{Y}$, about the true value of which we had no opinions save that it was in $\mathcal{S}_2$; and we used Bel_o to obtain a belief function

$$\text{Bel}: \mathcal{P}(\mathcal{S}_1 \times \mathcal{S}_2) \to [0, 1],$$

which expressed our degrees of belief in joint propositions about the true values of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$. Bel was given by

$$\mathrm{Bel}(A) = \mathrm{Bel}_o(\{x \mid \{x\} \times \mathcal{S}_2 \subset A\}),$$

Bel(A) being our degree of belief that the pair consisting of the true value of $\underset{\sim}{X}$ and the true value of $\underset{\sim}{Y}$ was in A.

Now let us suppose that $\mathcal{S}_1$ and $\mathcal{S}_2$ are actually the same set: $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$; so that our belief function Bel is actually on $\mathcal{P}(\mathcal{S} \times \mathcal{S})$. Now suppose that we suddenly learn that the quantities $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are identical -- that they have the same value. Then how should be modify Bel?

Evidently, we should condition Bel on the "diagonal" -- on the set

$$D = \{(s,s) \mid s \in \mathcal{S}\}.$$

This does not result in the elimination of any probability, for

$$\mathrm{Bel}(\overline{D}) = \mathrm{Bel}_o(\{x \mid \{x\} \times \mathcal{S}_2 \subset \overline{\{(s,s) \quad s \in \mathcal{S}\}}\})$$

$$= \mathrm{Bel}_o(\emptyset) = 0.$$

So the conditional belief function Bel_D is given simply by

$$\begin{aligned}\mathrm{Bel}_D(A) &= \mathrm{Bel}(A \vee \overline{D}) \\ &= \mathrm{Bel}_o(\{x \mid \{x\} \times \mathcal{S}_2 \subset A \vee \overline{D}\}) \\ &= \mathrm{Bel}_o(\{x \mid (x,x) \in A\})\end{aligned}$$

We might be interested in particular in $\mathrm{Bel}_D \big| \mathcal{P}(\mathcal{S}_2)$, which would give our conditional degrees of belief that the true value of $\underset{\sim}{Y}$ is in various subsets of $\mathcal{S}_2 = \mathcal{S}$. Denoting this belief function by

$$\mathrm{Bel}' : \mathcal{P}(\mathcal{S}_2) \to [0,1],$$

we would have

$$\mathrm{Bel}'(A) = \mathrm{Bel}_D(\mathcal{S}_1 \times A) = \mathrm{Bel}_o(\{x \mid (x,x) \in \mathcal{S}_1 \times A\})$$

$$= \mathrm{Bel}_o(A).$$

Hence our conditioning has resulted in the same degrees of belief for $\underset{\sim}{Y}$ as we formerly had for $\underset{\sim}{X}$. Nothing could be more reasonable.

4. Dempster's Rule of Combination: Finite Case

Suppose we have two belief functions Bel_1 and Bel_2 on the same Boolean algebra $\mathcal{A}$, and suppose the two are based on independent sources of evidence. Then it would be pleasant if we could combine them in some orthogonal way so as to produce a single resulting belief function on $\mathcal{A}$; this would correspond to pooling the evidence from which the two belief functions arose. How might we carry out such a combination?

This question can be approached most easily in the case where $\mathcal{A}$ is finite. In that case, it should be recalled, a belief function Bel on $\mathcal{A}$ can be described by "basic probability numbers" $\{m_A\}_{A \in \mathcal{A}}$. The intuitive understanding is that the basic probability number m_A represents the measure of a "basic probability mass" which is constrained to A but not to any proper subelement of A. Suppose we have two belief functions Bel_1 and Bel_2 on $\mathcal{A}$, with basic probability numbers $\{n_A\}_{A \in \mathcal{A}}$ and $\{p_A\}_{A \in \mathcal{A}}$, respectively. In order to think about combining Bel_1 and Bel_2, let us think of Bel_1 as our own original belief function, while Bel_2 is the belief function of a second person whose opinions we wish to combine orthogonally with our own. How can be use Bel_2 to modify our original beliefs?

Well, let us consider each of the other person's basic probability masses separately. The basic probability mass which he associates with A is committed to A but to no proper subelement of A. As far as that probability mass is concerned, the natural thing seems to be to condition Bel_1 on A. In other words, we should restrict each of the basic probability masses for Bel_1 to A, thus obtaining a basic probability mass for each $B \in \mathcal{Q}$ of measure

$$\Sigma \{n_{A'} \mid A' \wedge A = B\}.$$

But this should apply only for Bel_2's basic probability mass for A, which has measure p_A. Doing to the same for each $A \in \mathcal{Q}$, we would obtain the total

$$\sum_{A \in \mathcal{Q}} p_A \; \Sigma \{ n_{A'} \mid A' \wedge A = B \}. \tag{1}$$

as the measure of the new basic probability mass associated with B.

The difficulty with (1) is, of course, that it may be positive for $B = \Lambda$; there may be some probability that is constrained to Λ as a result of this rule. Hence we must discard that portion of our probability and renormalize the measure of the remainder. This results in a new belief function Bel with basic probability numbers $\{m_B\}_{B \in \mathcal{Q}}$, where

$$m_B = \frac{\sum_{A \in \mathcal{Q}} p_A \; \Sigma \{n_{A'} \mid A' \wedge A = B \}}{1 - \sum_{A \in \mathcal{Q}} p_A \; \Sigma \{n_{A'} \mid A' \wedge A = \Lambda \}} \tag{2}$$

for $B \neq \Lambda$, and $m_\Lambda = 0$.

The numbers $\{m_B\}_{B \in \mathcal{Q}}$ defined by (2) are evidently non-negative, so in order to show that they determine a belief function it suffices to show that they add to one, and this is easily verified. The only difficulty that might arise is that we might have

$$\sum_{A \in} p_A \, \Sigma \{n_{A'} \mid A' \wedge A = \Lambda\} = 1; \tag{3}$$

in such a case the denominator in (2) would be zero and the combination could not be carried out. But since $\sum_{A \in} p_A = 1$, (3) would imply that

$$\Sigma \{n_{A'} \mid A' \wedge A = \Lambda\} = 1$$

for all A for which $p_A > 0$. Denoting the A for which $p_A > 0$ by $A_1, \ldots, A_k$, we find that

$$\mathrm{Bel}_1(\overline{A_i}) = \Sigma \{ n_{A'} \mid A' \leq \overline{A_i}\} = \Sigma \{n_{A'} \mid A' \wedge A_i = \Lambda\} = 1$$

for each i, $i = 1, \ldots, k$. Setting $C = A_1 \vee \ldots \vee A_k$, this implies that

$$\mathrm{Bel}_1(\overline{C}) = \mathrm{Bel}_1(\overline{A_1} \wedge \ldots \wedge \overline{A}_k) = 1,$$

while

$$\mathrm{Bel}_2(C) = \sum_{A \leq C} p_A = 1.$$

So the combination of Bel_1 and Bel_2 is impossible only when there exists $C \in \mathcal{Q}$ such that $P_1{*}(C) = 0$ but $\mathrm{Bel}_2(C) = 1$; i.e., when the two belief functions contradict each other.

5. Dempster's Rule of Combination: General Case

There are several approaches that we might take to adduce Dempster's rule of combination for the infinite case. One approach would be to develop the theory of integration for probability algebras and use it to adduce integrals analogous to the sums in formula (1) of the preceding chapter. An approach that we are better equipped to pursue is to draw an analogy with the "orthogonal combination" of Chapter 6, modifying that method by adding the element of renormalization. This is the approach of the following theorem.

Theorem. Suppose $Bel_1 : \mathcal{A} \to [0,1]$ and $Bel_2 : \mathcal{A} \to [0,1]$ are both belief functions, with standard representations $\rho_1^{\Delta} : \mathcal{A} \to \mathcal{M}_1$ and $\rho_2^{\Delta} : \mathcal{A} \to \mathcal{M}_2$. Let $((\mathcal{M}, \mu); i_1 : \mathcal{M}_1 \to \mathcal{M}\ ;\ i_2 : \mathcal{M}_2 \to \mathcal{M})$ be an orthogonal sum of $(\mathcal{M}_1, \mu_1)$ and $(\mathcal{M}_2, \mu_2)$. Denote $\rho_1' = i_1 \circ \rho_1^{\Delta}$ and $\rho_2' = i_2 \circ \rho_2^{\Delta}$. And suppose that

$$M = \bigvee_{A \in \mathcal{A}} (\rho_1'(A) \wedge \rho_2'(\bar{A})) \neq V_{\mathcal{M}}.$$

Denote by I the principal ideal of $\mathcal{M}$ generated by M, and let $(\mathcal{M}/I, \nu)$ and $f : \mathcal{M} \to \mathcal{M}/I$ be as in section 4 of Chapter 4. Then

$$\rho' : \mathcal{A} \to \mathcal{M}/I : A \mapsto f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{A};$$
$$A_1 \wedge A_2 \leq A \})$$

is a standard allocation of probability on $\mathcal{A}$. And the belief function $Bel = \nu \circ \rho'$ is given by

$$Bel(A) = \frac{k(A) - k}{1 - k} ,$$

where

$$k(A) = \sup \{ \Sigma \mathrm{Bel}_1(A_i)\, \mathrm{Bel}_2(B_i) - \Sigma \mathrm{Bel}_1(A_i \wedge A_j)\, \mathrm{Bel}_2(B_i \wedge B_j) +- \dots +(-1)^{n+1} \mathrm{Bel}_1 (A_1 \wedge \dots \wedge A_n)\, \mathrm{Bel}_2(B_1 \wedge \dots \wedge B_n) \mid n \geq 1;\ A_i,\ B_i \in \mathcal{Q};\ A_i \wedge B_i \leq A \} \tag{1}$$

and $k = k(\Lambda) = \mu(M)$

$$= \sup \{ \Sigma \mathrm{Bel}_1(A_i)\, \mathrm{Bel}_2(\overline{A_i}) - \Sigma \mathrm{Bel}_1(A_i \wedge A_j)\, \mathrm{Bel}_2(\overline{A_i} \wedge \overline{A_j}) +- \dots + (-1)^{n+1} \mathrm{Bel}_1(A_1 \wedge \dots \wedge A_n)\, \mathrm{Bel}_2(\overline{A_1} \wedge \dots \wedge \overline{A_n}) \mid n \geq 1;\ A_1, \dots, A_n \in \mathcal{Q} \}. \tag{2}$$

<u>Proof:</u> To show that ρ' is an allocation, notice that

(i) $\rho'(\Lambda) = f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{Q};\ A_1 \wedge A_2 \leq \Lambda \})$

$= f(M) = \Lambda,$

(ii) $\rho'(\Upsilon) = f(\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{Q};\ A_1 \wedge A_2 \leq \mathcal{V} \})$

$= f(\mathcal{V}) = \Upsilon,$

and (iii) $\rho'(A) \wedge \rho'(B) = f(\vee \{ \rho_1'(R) \wedge \rho_2'(S) \mid R \wedge S \leq A \}) \wedge f(\vee \{ \rho_1'(T) \wedge \rho_2'(U) \mid T \wedge U \leq B \})$

$= f(\vee \{ \rho_1'(R) \wedge \rho_2'(S) \wedge \rho_1'(T) \wedge \rho_2'(U) \mid R \wedge S \leq A;\ T \wedge U \leq B \})$

$= f(\vee \{ \rho_1'(R \wedge T) \wedge \rho_2'(S \wedge U) \mid R \wedge S \leq A;\ T \wedge U \leq B \})$

$= f (\vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq A \wedge B \})$

$= \rho' (A \wedge B).$

Now by the formula in section 4 of Chapter 4,

$$\mathrm{Bel}(A) = \nu(f(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq A\}))$$

$$= \frac{1}{1-\mu(\mathcal{M})}\ \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq A\} - M)$$

$$= \frac{1}{1-\mu(\mathcal{M})}\Big[\mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq A\})$$

$$- \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq \Lambda\})\Big]$$

$$= \frac{1}{1-k}(k(A) - k),$$

where

$$k(A) = \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq A\})$$

$$= \mu(\vee\{[\rho_1'(A_1) \wedge \rho_2'(B_1)] \vee [\rho_1'(A_2) \wedge \rho_2'(B_2)] \vee \ldots \vee$$

$$[\rho_1'(A_n) \wedge \rho_2'(B_n)] \mid A_i \wedge B_i \leq A \text{ for each } i\})$$

$$= \sup\{\mu([\rho_1'(A_1) \wedge \rho_2'(B_1)] \vee \ldots \vee [\rho_1'(A_n) \wedge \rho_2'(B_n)]) \mid$$

$$A_i \wedge B_i \leq A \text{ for each } i\}$$

$$= \sup\{\Sigma\, \mathrm{Bel}_1(A_i)\, \mathrm{Bel}_2(B_i) - \Sigma\, \mathrm{Bel}_1(A_i \wedge A_j)\, \mathrm{Bel}_2$$

$$(B_i \wedge B_j) +- \ldots + (-1)^{n+1}\, \mathrm{Bel}_1(A_1 \wedge \ldots \wedge A_n)$$

$$\mathrm{Bel}_2(B_1 \wedge \ldots \wedge B_n) \mid n \geqq 1;\ A_i, B_i \in \mathcal{Q};\ A_i \wedge B_i \leq A\},$$

and $\quad k = k(\Lambda)$

$$= \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq \Lambda\})$$

$$= \mu(\bigvee_{A \in \mathcal{Q}}\{\rho_1'(A) \wedge \rho_2'(\bar{A})\})$$

$$= \sup \{\Sigma \mathrm{Bel}_1(A_i)\ \mathrm{Bel}_2(\bar{A}_i) - \Sigma \mathrm{Bel}_1(A_i \wedge A_j)\ \mathrm{Bel}_2(\bar{A}_i \wedge \bar{A}_j)$$

$$+ - \dots + (-1)^{n+1}\ \mathrm{Bel}_1(A_1 \wedge \dots \wedge A_n) \cdot \mathrm{Bel}_2(\bar{A}_1 \wedge \dots \wedge \bar{A}_n) \mid$$

$$n \geqq 1;\ A_1, \dots, A_n \in \mathcal{O}.\}.$$

Definition. Suppose Bel_1 and Bel_2 are two belief functions on a Boolean algebra $\mathcal{O}$. If k, as given by (2) above, obeys $k < 1$, then the belief function Bel defined in the above theorem is called the orthogonal sum of Bel_1 and Bel_2 and is denoted $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$. If $k = 1$, then the orthogonal sum of Bel_1 and Bel_2 is said not to exist.

Notice that the formulae giving the orthogonal sum do not depend on the particular representations ρ_1', ρ_2' and ρ'.

The preceding is a definition of "orthogonal sum" in the case of two belief functions on the same Boolean algebra. But in the preceding chapter we defined the notion of an orthogonal sum of two belief functions on different independent subalgebras of a Boolean algebra. The following theorem shows in what sense the present definition is a generalization of the previous definition.

Theorem. Suppose $\mathcal{O}_1$ and $\mathcal{O}_2$ are independent subalgebras of a Boolean algebra $\mathcal{O}$. And suppose $\mathrm{Bel}_1 : \mathcal{O}_1 \to [0, 1]$ and $\mathrm{Bel}_2 : \mathcal{O}_2 \to [0, 1]$ are belief functions. Denote by Bel_1' and Bel_2' the natural extensions of Bel_1 and Bel_2, respectively to $\mathcal{O}$. Let $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ be the orthogonal sum of Bel_1 and Bel_2 on $\mathcal{O}$, as defined in the preceding chapter. And let $\mathrm{Bel}_1' \oplus \mathrm{Bel}_2'$ be the orthogonal sum of Bel_1' and Bel_2', as defined above. Then $Bel_1 \oplus Bel_2 = Bel_1' \oplus Bel_2'$.

Proof: Let $\rho_1^{o}: \mathcal{A}_1 \to \mathcal{M}_1$ and $\rho_2^{o}: \mathcal{A}_2 \to \mathcal{M}_2$ be as in the first theorem of section 3 of Chapter 3. Let $((\mathcal{M}, \mu); i_1: \mathcal{M}_1 \to \mathcal{M}; i_2: \mathcal{M}_2 \to \mathcal{M})$, ρ_1 and ρ_2 and ρ be as in that theorem as well. Then $Bel_1 \oplus Bel_2 = \mu \circ \rho$, where

$$\rho = \vee\{ (\rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A \} .$$

But $Bel_1': \mathcal{A} \to [0,1]$ and $Bel_2': \mathcal{A} \to [0,1]$ are given by the allocations

$$\rho_1^{\Delta}: \mathcal{A} \to \mathcal{M}_1 : A \rightsquigarrow \vee\{ \rho_1^{o}(A_1) \mid A_1 \in \mathcal{A}_1;\ A_1 \leq A\}$$

and

$$\rho_2^{\Delta}: \mathcal{A} \to \mathcal{M}_2 : A \rightsquigarrow \vee\{ \rho_2^{o}(A_2) \mid A_2 \in \mathcal{A}_2;\ A_2 \leq A\}.$$

So $Bel_1' \oplus Bel_2'$ is given by

$$\rho': \mathcal{A} \to \mathcal{M}/I : A \rightsquigarrow f(\vee\{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\}),$$

where $\rho_1' = i_1 \circ \rho_1^{\Delta}$, $\rho_2' = i_2 \circ \rho_2^{\Delta}$ and

$$\begin{aligned}
I &= \bigvee_{A \in \mathcal{A}} (\rho_1'(A) \wedge \rho_2'(\bar{A})) \\
&= \bigvee_{A \in \mathcal{A}} (i_1(\vee\{ \rho_1^{\Delta}(A_1) \mid A_1 \in \mathcal{A}_1;\ A_1 \leq A \}) \wedge i_2(\vee\{ \rho_2^{\Delta}(A_2) \mid \\
&\qquad A_2 \in \mathcal{A}_2;\ A_2 \leq A\})) \\
&= \bigvee_{A \in \mathcal{A}} (\vee\{ i_1(\rho_1^{\Delta}(A_1)) \wedge i_2(\rho_2^{\Delta}(A_2)) \mid A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2; \\
&\qquad A_1 \leq A;\ A_2 \leq \bar{A} \}) \\
&= \perp ,
\end{aligned}$$

since $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent. Hence $(\mathcal{M}/I, \nu) = (\mathcal{M}, \mu)$ and f is the identity mapping. And ρ' is given by

$$\rho'(A) = \vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\}$$

$$= \vee\{i_1(\vee\{\rho_1^{\Delta}(A_1') \mid A_1 \in \mathcal{A}_1;\ A_1' \leq A_1\}) \wedge i_2(\vee\{\rho_2^{\Delta}(A_2') \mid A_2' \in \mathcal{A}_2;\ A_2' \leq A_2\}) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\}$$

$$= \vee\{(\vee\{\rho_1(A_1') \mid A_1 \in \mathcal{A}_1;\ A_1' \leq A_1\}) \wedge (\vee\{\rho_2(A_2') \mid A_2 \in \mathcal{A}_2;\ A_2' \leq A_2\}) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\}$$

$$= \vee\{\vee\{\rho_1(A_1') \wedge \rho_2(A_2') \mid A_1 \in \mathcal{A}_1;\ A_1' \leq A_1;\ A_2' \leq A_2\} \mid A_1, A_2 \in \mathcal{A},\ A_1 \wedge A_2 \leq A\}$$

$$= \vee\{\rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \in \mathcal{A}_1;\ A_2 \in \mathcal{A}_2;\ A_1 \wedge A_2 \leq A\}$$

$$= \rho(A).$$

So $\rho' = \rho$, and hence $Bel_1' \oplus Bel_2' = Bel_1 \oplus Bel_2$. ▨

So our present notion of combination is quite general. Of course, one can combine more than two belief functions at a time; the more general definition should be obvious. The operation of combination is commutative whenever it can be carried out, and it has a unit -- the vacuous belief function -- which when combined with any belief function always yields that belief function again. The operation of conditioning is also a special case of combination, as the following theorem shows:

Theorem. Suppose $Bel: \mathcal{A} \to [0, 1]$ is a belief function, $A \in \mathcal{A}$ and $P^*(A) > 0$. Let $Bel_2: \mathcal{A} \to [0, 1]$ be the belief function defined by

$$Bel_2(A') = \begin{cases} 1 & \text{if } A \leq A' \\ 0 & \text{otherwise.} \end{cases}$$

Then $Bel_A = Bel \oplus Bel_2$.

Proof: If $A' \in \mathcal{A}$, then

$$Bel_A(A') = \frac{Bel(A' \vee \bar{A}) - Bel(\bar{A})}{1 - Bel(\bar{A})}$$

Now if we let ρ_1', ρ_2' and $(\mathcal{M}, \mu)$ be as in the first theorem of this section, we have $Bel = \mu \circ \rho_1'$, and ρ_2' is given by

$$\rho_2'(A') = \begin{cases} \Upsilon & \text{if } A \leq A' \\ \Lambda & \text{otherwise.} \end{cases}$$

Hence

$$Bel \oplus Bel_1(A') = \frac{k(A') - k}{1 - k},$$

where

$$\begin{aligned} k &= \mu\Big(\bigvee_{A' \in \mathcal{A}} (\rho_1'(A') \wedge \rho_2'(\bar{A}'))\Big) \\ &= \mu(\vee\{\rho_1'(A') \mid A \leq \bar{A}'\}) \\ &= \mu(\rho_1'(\bar{A})) = Bel(\bar{A}), \end{aligned}$$

and

$$\begin{aligned} k(A') &= \mu(\vee\{\rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1 \wedge A_2 \leq A'\}) \\ &= \mu(\vee\{\rho_1'(A_1) \mid A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A';\ A \leq A_2\}) \\ &= \mu(\rho_1'(A' \vee \bar{A})) = Bel(A' \vee \bar{A}). \end{aligned}$$

6. The Condensable Case

In the previous section we saw that Dempster's rule of combination could be adduced for belief functions in general. But in fact, this rule, like the rule of conditioning, is most adapted to the condensable case. For in that case the rule can be stated quite simply in terms of the commonality numbers, and it will fail only when the belief functions contradict each other.

<u>Theorem</u>. Suppose Bel_1 and Bel_2 are condensable belief functions on $\mathcal{P}(\mathcal{S})$. Then $Bel_1 \oplus Bel_2$ fails to exist if and only if there exists $S \subset \mathcal{S}$ such that $Bel_1(S) = Bel_2(\bar{S}) = 1$. And in the case where $Bel_1 \oplus Bel_2$ does exist, it is condensable, and its commonality function Q is given by

$$Q(S) = \frac{1}{1 - k} \; Q_1(S) \, Q_2(S) \qquad (1)$$

for all finite non-empty subsets $S \subset \mathcal{S}$, where Q_1 and Q_2 are the commonality functions for Bel_1 and Bel_2, respectively, and k is the constant given in the first theorem of section 5.

<u>Proof</u>: This theorem is most easily established by comparing the construction in section 5 with the construction in section 3 of Chapter 6.

Think of Bel_1 and Bel_2 as belief functions on $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$, respectively, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are distinct copies of $\mathcal{S}$. Let $\rho_1^{\Delta}: \mathcal{P}(\mathcal{S}) \to \mathcal{M}_1$, $\rho_2^{\Delta}: \mathcal{P}(\mathcal{S}) \to \mathcal{M}_2$, $\mathcal{M}$, i_1, i_2, ρ_1', ρ_2', M and ρ' be as in the theorem of section 5. And let ρ_1^{o} and ρ_2^{o} be identical with ρ_1^{Δ} and ρ_2^{Δ} respectively, except that they are thought of as being on the copies $\mathcal{P}(\mathcal{S}_1)$ and $\mathcal{P}(\mathcal{S}_2)$, respectively. Let ρ_1, ρ_2 and ρ be the allocations based on ρ_1° and ρ_2° according to the formulae in section 3 of Chapter 6.

Let

$$D = \{ (s, s) \quad s \in \mathcal{S} \} \subset \mathcal{S}_1 \times \mathcal{S}_2.$$

Then

$$\begin{aligned}
\rho(\bar{D}) &= \vee \{ \rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ A_1 \times A_2 \subset \bar{D} \}. \\
&= \vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1, A_2 \subset \mathcal{S};\ A_1 \cap A_2 = \phi \} \\
&= \vee \{ \rho_1'(A) \wedge \rho_2'(\bar{A}) \mid A \subset \mathcal{S} \} \\
&= M,
\end{aligned}$$

and in general, for all $S \subset \mathcal{S}$, if $S' = \{ (s, s) \mid s \in S \} \subset \mathcal{S}_1 \times \mathcal{S}_2$, then

$$\begin{aligned}
\rho(S' \cup \bar{D}) &= \vee \{ \rho_1(A_1) \wedge \rho_2(A_2) \mid A_1 \subset \mathcal{S}_1;\ A_2 \subset \mathcal{S}_2;\ A_1 \times A_2 \subset S' \cup \bar{D} \} \\
&= \vee \{ \rho_1'(A_1) \wedge \rho_2'(A_2) \mid A_1,\ A_2 \subset \mathcal{S};\ A_1 \cap A_2 \subset S \}.
\end{aligned}$$

By comparing the theorem in section 2 with the first theorem in section 5, we see that ρ' is obtained from ρ by conditioning on D and then identifying D with $\mathcal{S}$ by the mapping $(s, s) \rightsquigarrow s$.

Hence our formula (1) becomes transparent; the multiplication follows from the similar rule in section 7 of Chapter 6, while the constant $1/(1 - k)$ results from the conditioning. ▨

7. Cognitive Independence

In the preceding chapter I suggested that two subalgebras $\mathcal{Q}_1$ and $\mathcal{Q}_2$ of a Boolean algebra $\mathcal{Q}$ deserved to be called cognitively independent with respect to a belief function Bel on $\mathcal{Q}$ if

$$P^*(A_1 \wedge A_2) = P^*(A_1) \cdot P^*(A_2) \tag{1}$$

for all $A_1 \in \mathcal{Q}_1$ and $A_2 \in \mathcal{Q}_2$. We are now in a position to examine the basis of that suggestion.

What ought we to mean when we say that two subalgebras are cognitively independent with respect to our opinions? Intuitively, we ought to mean that the assimilation of new evidence or opinion about the propositions in one of them would not change our degrees of belief in the propositions in the other. But Dempster's rules of conditioning and combination provide us with a mathematical representation of how new evidence or opinion can be assimilated, and hence we can make this intuitive understanding mathematically precise.

Indeed, if our new evidence about $\mathcal{Q}_1$ comes down to the knowledge that $A_1 \in \mathcal{Q}_1$ is true, then we would modify Bel by conditioning it on A_1. And, more generally, if our new evidence induced a belief function Bel_1 on $\mathcal{Q}_1$, then we would modify Bel by replacing it with $Bel_1' \oplus Bel$, where Bel_1' is the natural extension of Bel_1 to $\mathcal{Q}$. And as the following theorems show, these sorts of modifications in Bel will always fail to modify the degrees of belief in elements of $\mathcal{Q}_2$ if and only if (1) holds.

Theorem. Suppose Bel: $\mathcal{Q} \to [0,1]$ is a belief function and $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are independent subalgebras of $\mathcal{Q}$. Then $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are cognitively independent with respect to Bel if and only if

$Bel_{A_1} | \mathcal{A}_2 = Bel | \mathcal{A}_2$ whenever $A_1 \in \mathcal{A}_1$ and $P^*(A_1) > 0$.

Proof: $Bel_{A_1} | \mathcal{A}_2 = Bel | \mathcal{A}_2$ whenever $A_1 \in \mathcal{A}_1$ and $P^*(A_1) > 0$ if and only if

$$P^*(A_2) = \frac{P^*(A_2 \wedge A_1)}{P^*(A_1)}$$

for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ such that $P^*(A_1) > 0$. But this equation holds for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ such that $P^*(A_1) > 0$ if and only if

$$P^*(A_1 \wedge A_2) = P^*(A_1) \cdot P^*(A_2)$$

for all $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. ▨

Theorem. Suppose $Bel: \mathcal{A} \to [0, 1]$ is a belief function and $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent subalgebras of $\mathcal{A}$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are cognitively independent with respect to Bel if and only if $Bel_1' \oplus Bel | \mathcal{A}_2 = Bel | \mathcal{A}_2$ whenever Bel_1' is the natural extension to of a belief function Bel_1 on $\mathcal{A}_1$ and $Bel_1' \oplus Bel$ exists.

Proof. In view of the preceding theorem, it suffices to show that (1) and the existence of $Bel_1' \oplus Bel$ implies that

$$(Bel_1' \oplus Bel)(A) = Bel(A)$$

for all $A \in \mathcal{A}_2$. But by the formulae of section 5, we find that

$$(Bel_1' \oplus Bel)(A) = \Big(\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A_2) | A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq A\}) - \mu(\vee\{\rho_1(A_1) \wedge \rho_2(A_2) | A_1, A_2 \in \mathcal{A};\ A_1 \wedge A_2 \leq \Lambda\})\Big) / \Big(1 - \mu(\vee\{\rho_1(A_1) \wedge \rho_2(A_2) | A_1, A_2 \in \mathcal{A};\ A \wedge A_2 \leq \Lambda\})\Big)$$

where $\rho_1: \mathcal{A} \to \mathcal{M}$ and $\rho_2: \mathcal{A} \to \mathcal{M}$ are allocations which represent Bel_1' and Bel, respectively, and which map $\mathcal{A}$ into orthogonal subalgebras of $\mathcal{M}$. Now Bel_1' is supported by $\mathcal{A}_1$; hence

$$\rho_1(A_1) = \vee \{ \rho_1(A') \mid A' \in \mathcal{A}_1;\ A' \leq A_1 \}$$

for all $A_1 \in \mathcal{A}$, and it follows that

$$\begin{aligned}(Bel_1' \oplus Bel)(A) = \Big(\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A \vee \bar{A}_1) \mid A_1 \in \mathcal{A}_1\}) - \mu(\vee\{\rho_1 \\ (A_1) \wedge \rho_2(\bar{A}_1) \mid A_1 \in \mathcal{A}_1\})\Big) \Big/ \Big(1 - \mu(\vee\{\rho_1(A_1) \\ \wedge \rho_2(\bar{A}_1) \mid A_1 \in \mathcal{A}_1\})\Big).\end{aligned} \tag{2}$$

But

$$\begin{aligned}&\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A \vee \bar{A}_1) \mid A_1 \in \mathcal{A}_1\}) \\ &= \sup_{A_1, \ldots, A_n \in \mathcal{A}_1} \mu((\rho_1(A_1) \wedge \rho_2(A \vee \bar{A}_1)) \vee \ldots \vee \\ &\quad (\rho_1(A_n) \wedge \rho_2(A \vee \bar{A}_n))) \\ &= \sup_{A_1, \ldots, A_n \in \mathcal{A}_1} \Big[\sum_i \mu(\rho_1(A_i) \wedge \rho_2(A \vee \bar{A}_i)) - \sum_{i<j} \mu(\rho_1(A_i \wedge A_j) \\ &\quad \wedge \rho_2(A \vee (\bar{A}_i \wedge \bar{A}_j))) + - \ldots\Big] \\ &= \sup_{A_1, \ldots, A_n \in \mathcal{A}_1} \Big[\sum_i Bel_1(A_i)\, Bel(A \vee \bar{A}_i) - \sum_{i<j} Bel_1(A_i \wedge A_j) \\ &\quad Bel(A \vee (\bar{A}_i \wedge \bar{A}_j)) + - \ldots\Big].\end{aligned}$$

Now since $\mathcal{A}_1$ and $\mathcal{A}_2$ are cognitively independent with respect to Bel, we have

$$P^*(A \wedge B) = P^*(A) \cdot P^*(B),$$

or

$$\mathrm{Bel}(A \vee B) = \mathrm{Bel}(A) + \mathrm{Bel}(B) - \mathrm{Bel}(A)\,\mathrm{Bel}(B)$$

whenever $A \in \mathcal{Q}_2$ and $B \in \mathcal{Q}_1$. We are indeed assuming that $A \in \mathcal{Q}_2$, so our preceding formula becomes

$$\mu(\vee\{\rho_1(A_1) \wedge \rho_2(A \vee \overline{A}_1) \mid A_1 \in \mathcal{Q}_1\})$$

$$= \sup_{A_1,\ldots,A_n \in \mathcal{Q}_1} \Big[\sum_i \mathrm{Bel}_1(A_i)\big((\mathrm{Bel}(A) + \mathrm{Bel}(\overline{A}_i) - \mathrm{Bel}(A)\,\mathrm{Bel}(\overline{A}_i)\big)$$
$$- \sum_{i<j} \mathrm{Bel}_1(A_i \wedge A_j)\big((\mathrm{Bel}(A) + \mathrm{Bel}(\overline{A}_i \wedge \overline{A}_j) -$$
$$- \mathrm{Bel}(A)\,\mathrm{Bel}(\overline{A}_i \wedge A_j)\big) + - \ldots \Big]$$

$$= \sup_{A_1,\ldots,A_n \in \mathcal{Q}_1} \Big[\mathrm{Bel}(A)\,\big(\Sigma\,\mathrm{Bel}_1(A_i) - \Sigma\,\mathrm{Bel}_1(A_i \wedge A_j) + - \ldots\big)$$
$$+ (1 - \mathrm{Bel}(A))\,(\Sigma\,\mathrm{Bel}_1(A_i)\,\mathrm{Bel}(\overline{A}_i) -$$
$$- \Sigma\,\mathrm{Bel}_1(A_i \wedge A_j)$$
$$\mathrm{Bel}(\overline{A}_i \wedge \overline{A}_j) + - \ldots)\Big]$$

$$= \sup_{A_1,\ldots,A_n \in \mathcal{Q}_1} \Big[\mathrm{Bel}(A)\,(\mu(\rho_1(A_1) \vee \ldots \vee \rho_1(A_n)))$$
$$+ (1 - \mathrm{Bel}(A))\,(\mu((\rho_1(A_1) \wedge \rho_2(\overline{A}_1)) \vee \ldots \vee$$
$$(\rho_1(A_n) \wedge \rho_2(\overline{A}_n))))\Big]$$

$$= \mathrm{Bel}(A) \sup_{A_1,\ldots,A_n \in \mathcal{Q}_1} \mu(\rho_1(A_1) \vee \ldots \vee \rho_1(A_n))$$
$$+ (1 - \mathrm{Bel}(A)) \sup_{A_1,\ldots,A_n \in \mathcal{Q}_1} \mu((\rho_1(A_1) \wedge \rho_2(\overline{A}_1)) \vee \ldots \vee$$
$$(\rho_1(A_n) \wedge \rho_2(\overline{A}_n)))$$

$$= \mathrm{Bel}(A) + (1 - \mathrm{Bel}(A))\,\mu(\vee\{\rho_1(A_1) \wedge \rho_2(\overline{A}_1) \mid A_1 \in \mathcal{Q}_1\}).$$

So, setting

$$k = \mu (\vee \{ \rho_1(A_1) \quad \wedge \rho_2(\bar{A}_1) \mid A_1 \in \mathcal{A}_1 \}),$$

(2) becomes

$$(\mathrm{Bel}_1{}' \oplus \mathrm{Bel})(A) = \frac{\mathrm{Bel}(A) + (1 - \mathrm{Bel}(A))\,k - k}{1 - k}$$

$$= \mathrm{Bel}(A). \qquad \blacksquare$$

8. Conclusion

It is evident that Dempster's rule of combination will play a central role in any application of the theory of belief functions, for we always encounter the need to combine evidence. In view of this importance, the rule deserves a much closer scrutiny -- we need to examine a good many examples of its application so as to understand its behavior clearly. I cannot undertake such an examination here, but I have made some efforts to examine its behavior in the paper entitled "A Theory of Statistical Support."

I have not developed the formulae for combining more than one belief function at a time, but it should be evident that such combination is possible. Furthermore, it can be carried out stepwise, and the order will not matter: the operation of combination is commutative. This is particularly obvious in the condensable case, for aside from an appropriate renormalization, the combination of condensable belief functions is affected merely by multiplying the commonality functions.

It should be noted that this operation of combination is not idempotent. In other words, Bel ⊕ Bel is not, in general, equal to Bel. This fact is best explicated if we think in terms of the evidence underlying Bel. Since the operation of combination corresponds to the pooling of evidence, Bel ⊕ Bel will be appropriate for the situation where all the evidence is twice as strong as that underlying Bel.

It is not so easy, of course, to go back and forth from the commonality functions, which are easy to manipulate, to the belief functions and upper probability functions, which are of greater immediate interest; the formulae for doing so that were adduced in Chapter 5 are hardly of practical use. Hence any application of this theory will involve the rather difficult task of developing effective computational methods for combination. This difficulty is central in the theory of Dempsterian inference, for which the present essay is meant as a foundation.

BIBLIOGRAPHY

Bartle, Robert G. (1966). *The Elements of Integration*. Wiley, New York.

Bernoulli, Daniel. (1777). "The most probable choice between several discrepant observations and the formation therefrom of the most likely induction." *Acta Acad. Petrop.* pp. 3-33. Reprinted as pp. 157-167 of Pearson and Kendall's *Studies in the History of Statistics and Probability*, Griffin, 1970.

Bernoulli, James. (1713). *Artis Conjectandi*. Basel. Especially pp. 217-223. See also pp. 22-34 of the translation by Bing Sung, issued as Technical Report No. 2 of the Department of Statistics, Harvard University, February 12, 1966.

Birkhoff, Garrett. (1967). *Lattice Theory*, Third Edition. American Mathematical Society.

Choquet, Gustave. (1953-4). "Theory of Capacities." *Annales de l'Institut Fourier, Université de Grenoble*, V. Pp. 131-296.

Dempster, A.P. (1966). "New methods for reasoning towards posterior distributions based on sample data." *Ann. Math. Statist.* 37, pp. 355-374.

Dempster, A.P. (1967). "Upper and lower probabilities induced by a multi-valued mapping." *Ann. Math. Statist.* 38, pp. 325-339.

Dempster, A.P. (1967). "Upper and lower probability inferences based on a sample from a finite univariate population." *Biometrika* 54, pp. 515-528.

Dempster, A.P. (1968). "Upper and lower probabilities generated by a random closed interval." *Ann. Math. Statist.* 39, pp. 957-966.

Dempster, A.P. (1968). "A generalization of Bayesian inference (with discussion)." *J. Roy. Statist. Soc. Ser. B.* 30, pp. 205-247.

Dempster, A.P. (1968). *The Theory of Statistical Inference: A Critical Analysis. Chapter 1.* Research Report S-2, Department of Statistics, Harvard University, August 26, 1968.

Dempster, A.P. (1968). *The Theory of Statistical Inference: A Critical* Analysis. Chapter 2. Research Report S-3, Department of Statistics, Harvard University, September 27, 1968.

Dempster, A.P. (1969). "Upper and lower probability inferences for families of hypotheses with monotone density ratios." *Ann. Math. Statist.* 40, pp. 953-969.

DePauw, Linda Grant, Ed. (1972). Documentary History of the First Federal Congress of the United States of America. Volume I: Senate Legislative Journal. The John Hopkins University Press. (Library of Congress Catalogue Number 73-155164.)

Diderot, Dénis. (1785). "Probabilité." Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers. Vol. 27, pp. 443-454.

Halmos, Paul R. (1963). Lectures on Boolean Algebras. Van Nostrand, New York.

Halmos, Paul R. (1950). Measure Theory. Van Nostrand, New York.

Halmos, Paul R. (1960). Naive Set Theory. Van Nostrand, New York.

Kappos, Demetrios A. (1969). Probability Algebras and Stochastic Spaces. Academic Press, New York.

Kolmogorov, A.N. and S.V. Fomin. (1961). Elements of the Theory of Functions and Functional Analysis. Volume 2. Graylock, Baltimore.

Lambert, Johann Heinrich. (1764). Neues Organon, Zweiter Band. Pp. 318-421. Reprinted in 1965 as Vol. II of Lambert's Philosophische Schriften by Georg Olms Verlagsbuchhandlung, Hildesheim.

Prevost and Lhuilier. (1797). "Mémoire sur l'application du Calcul des probabilités à la valeur du témoignage." Mémoires de l'Académie Royale de Berlin. Pp 120-152.

Shafer, Glenn. (1972). The Problem of Statistical Support. Department of Statistics, Princeton University. (This seminar paper is Part I of the larger work of which the present essay is Part II.)

Shafer, Glenn. (1973). A Theory of Statistical Support. An essay presented at a symposium on the foundations of probability and statistics at the University of Western Ontario, May 13, 1973.

Sikorski, Roman. (1969). Boolean Algebras. Springer-Verlag, New York.

Todhunter, Issac. (1865). A History of the Mathematical Theory of Probability. Reprinted by the Chelsea Publishing Co., New York, 1949.